

GOOD COMPUTING: A VIRTUE APPROACH TO COMPUTER ETHICS

A book under contract with
Jones & Bartlett Publishers
40 Tall Pine Drive
Sudbury, MA 01776
<http://www.jbpub.com/>

**DRAFT FOR
June Puerto Rico writing retreat**

Chapter 4

Testing

Version 1: 03/31/05 by Chuck Huff

Version 1.1: 04/11/05 by Chuck Huff (minor revision)

(to do: add Bandura's moral distancing checklist)

Based on writings in www.computingcases.org prepared by Chuck Huff and Bill Frey and class documents from University of Mayaguez, PR, FILO 3185 prepared by Bill Frey

(All rights reserved. For classroom use only. Please do not cite without permission.)

©Charles Huff, William J. Frey, & José Cruz-Cruz

In this chapter, you will learn a series of four ethics tests that will help you to test proposed solutions to ethical difficulties. Most of your ethical decisions in computing will not require the use of these tests. Most ethical decisions are done in the process of doing something like (like building a software product) and are guided by your habitual way of approaching things. This, in fact, is what virtues are about, they guide our actions.

But in some contexts, or with some problems, you will find that you have several possible solutions or choices of action. If the choice among these is an important one (say, whether to blow the whistle on illegal behavior in your company) you will want to think carefully about your responsibilities before committing yourself to a decision. These tests will help you to structure the ethical decision making process in these cases. The better you get at using these tests, the more they will become second nature as you make other decisions, and the more they will become a part of the skill set that supports your habitual approach, your virtues, in computing.

Introduction:

To begin, take a look at the following scenario. Then think about the questions and meta-questions that follow. (Meta-questions are questions that step back from the practical matters to reflect on background issues. For example, a meta-question invites us to think about the definition of a key term or concept used in a scenario.)

- *You are an employee of company X and use your PC at work to send email to friends and relatives.*
1. **Is this an ethical issue?** (Meta-question: What makes something an ethical issue?)
 2. **Does your action harm anyone?** For example: (a) Does it harm the company you work for, your supervisor, or even you? (b) What kind of harm? (Physical? Mental? Financial?)
 3. **Does your action benefit anyone?** Does it benefit your friends and family? Does it benefit you? How? Think about the benefits of communicating with friends and family. Is sending email messages during work time the best way to bring about these benefits? Do these benefits come only attached to certain costs? Are there viable alternatives that avoid these costs, i.e., is this the only way to stay in touch with friends and family?
 4. **If you traded places with your supervisor, would you want her to send emails during company time using company equipment?** Reverse roles: what would *you* do if she were under your supervision, and you found she did this?
 5. **What does this say about you as an employee?** What kind of worker does this sort of thing? A responsible worker? Suppose that when someone approaches, you minimize the email window and maximize the report you were supposed to be working on. Why would you do this?
 6. **If you were a government employee in Puerto Rico, your action would be illegal.** If it's illegal, would you still do it? What if the penalty is minimal, and

there's only a small chance of getting caught? (Meta-question: Does the fact that something is illegal make it unethical?)

Doing ethical reasoning in the everyday world consists of asking and answering questions like these. Many assume that the email scenario doesn't raise any ethical issues. So ask yourself this question: Are ethical problems a part of day-to-day life or do they only arise in exceptional or unusual circumstances? For instance, the following scenario doesn't seem to raise any ethical issues:

You are a computer engineer and have been subcontracted to local your government to purchase new portable computers for high school teachers. Your job includes...

- *selecting the kind of computer to be used*
- *identifying vendors who will sell you the computers*
- *overseeing the distribution of computers to high school teachers*
- *developing and implementing a training program to help teachers learn to use computers*
- *designing a technical support hotline to help teachers work through the problems that arise*

Distributing computers to high school teachers seems simple enough. You select the computers, buy them, and give them to the teachers. They use them. Yet only a slight change in circumstances can bring into the open latent or potential ethical issues:

1. **How should you go about selecting the computers for the teachers?** What should you do to identify their needs? Would you find the computers and the software you selected acceptable if you were in their place? How do you successfully put yourself in the place of those who will be affected by your actions and decisions? (For example, you are comfortable using computers. But how do you understand and respond to those who find computers strange, exotic, or intimidating?)
2. **Who stands to benefit from or be harmed by your actions?** Does your decision disproportionately benefit (or harm) anyone? Do some reap the benefits of your actions while others bear the harms?
3. **Latent ethical issues often arise when events take a certain line of development.** Consider these four slight changes in circumstances:
 - a. Someone you knew well—say, your cousin—submits a bid. What ethical issues does this turn of events give rise to?
 - b. The contract to provide computers is awarded to your cousin, and he provides reliable computers at a reasonable price. Then, a few weeks later, you read the following headline in the newspaper: **“More Government Corruption—Computer Tsar’s Cousin Reaps Millions in Cozy Computer Deal.”** What would people think about you now? What would you think about yourself? What should you do?
 - c. A group of angry high school teachers holds a press conference in which they accuse the government of forcing them to use computing technology

in their classes. How should you respond? They say you are violating their academic freedom. Are you?

- d. Someone in the government suggests placing a program in each computer that allows law enforcement officials to monitor the computers and track user behavior. How would you feel if your computer use were being monitored without your knowledge or consent? Are there circumstances under which monitoring could bring about any social benefits? What are the likely harms? Do the benefits outweigh the harms? Suppose you go along with this and read the following headline in the morning newspaper: **“Government Snoops Bug High School Computers!”** What would people think of you for going along with this? What would you think of yourself?

Watching Our Language

Ethical theory is embedded in our language and reveals itself through the words we use and the reasons we provide when discussing ethical issues. For example, consider the following definitions of ethics formulated by students at the beginning of a computer ethics class:

- Ethics is doing the right thing.
- Ethics is a series of rules that govern right decision making.
- A combination of virtues that allow us to do what is best for our community, as a whole, in a moral manner.
- Ethics is following laws and respecting the rights of others.
- Ethics is a way of considering what is correct and what is not in a certain issue or situation.
- The proper way to behave and interact with others.
- The feeling of responsibility a person has toward various things
- At the moment of decision, to be able to identify what is good and what is bad
- Ethics is what society accepts as right.

Each of these definitions assumes one or more ethical theories. For example, saying that ethics is a combination of virtues that allows us to do what is best for our community assumes the theory of virtue ethics. To say that ethics is what a society accepts as right presupposes a view of cultural relativism. So in our everyday language we are already operating from a variety of theoretical perspectives that are embedded in our language. These theoretical perspectives can be teased out of our ordinary language by paying careful attention to word use, to implications and assumptions, and even to what is not said.

But ordinary language often obscures the actual structure of the ethical theories we assume when we talk. One of the tasks of this chapter is to help you become aware of the structure of the ethical approaches you use when you talk about these issues. This is important because different ethical approaches can lead you to emphasize different aspects of a situation, and can even lead to different conclusions.

To see a more forceful instance of this, consider the dilemma constructed by Bernard Williams (ref), *Jim and the Jungle*. These philosophical dilemmas are a bit stilted because they are designed to abstract out the messiness of real decision-making so we can focus on the underlying structure of ethical theories.

You are in a remote mountain village. A group of terrorists has lined up 20 people from the village; they plan on shooting them for collaborating with the enemy. Since you are not from the village, you will not be killed. Taking advantage of your position, you plead with the terrorists not to carry out their plan. Finally, you convince the leader that it is not necessary to kill all 20. He takes a gun, empties it of all its bullets except one, and then hands it to you. He has decided to kill only one villager to set an example to the rest. As an honored guest and outsider, you will decide who will be killed, and you will carry out the execution. The terrorists conclude with a warning; if you refuse to kill the villager, then they will revert back to the original plan of killing all 20. And if you try any “funny business,” they will kill you and then the 20 villagers. What should you do?

The mountain terrorist scenario is formulated as a classical dilemma. There appear to be only two choices: (1) take the gun, select the victim, and execute him or her; then the terrorists will allow the remaining 19 to walk away unharmed; or (2) walk away from the whole situation, in which case the terrorists will kill all 20 villagers. (Funny business would seem to represent a third set of options. But the terrorists, says the philosopher,, block this option with their warning.)

This scenario clearly distinguishes three important ethical approaches. For example, those who opt for taking the gun and killing one of the villagers offer the following arguments:

- It's not murder; it's killing one person in order to save 19.
- I would ask for a volunteer. Then I would shoot that person.
- I would choose someone old or someone sick and about to die; then I would be doing less harm than if I killed someone young
- I would ask the villagers to give me the name of a collaborator. This would prevent me from killing an innocent person.
- If I walked away and the terrorists killed all 20 villagers, then I would be responsible for their deaths (since I could have prevented it).

These responses make use of different forms of *consequentialism*, including *utilitarianism*. They are based on the idea that we evaluate our actions by the results they bring about. Killing 1 villager is justified by the consequences: saving the remaining 19. Then the decision maker may attempt to minimize or mitigate the harmful consequences by asking for a volunteer (so the victim is a willing participant) or by choosing someone who has already lived a full life and is about to die. This is an attempt to produce a positive balance of good vs. bad consequences. (For more on utilitarianism, see chapter 12 and the textbox in this chapter.)

On the other hand, those who argue against killing the villager offer reasons like the

following:

- Who am I to choose who will live and who will die?
- I have no right to take another person's life.
- Killing a villager violates God's law.
- How do you know that the terrorist will keep their word and refrain from killing the others after you have done your part?
- I couldn't live with myself if I did such a thing.

These responses turn the focus away from the consequences toward a consideration of the moral nature of the act itself. There is something wrong with killing an innocent human being (killing not done in self-defense). It violates another's right to life, or it goes against nature or God's law. Many who argue from this perspective dismiss the relevance of consequences by questioning whether the terrorists will keep their word and not kill anyone after you have done your part. Given the fallibility of human predictive knowledge, wouldn't it be better if we looked at the formal characteristics of the action itself (in terms of rights, duties etc.), independently of the consequences the action may or may not produce? (For more on this see chapter 12 and the text box on Kantian Formalism, called the *deontological* approach.)

Finally, some may argue that they couldn't carry through with such an act because they wouldn't be able to live with themselves afterwards. This argument seems based on the idea that some actions destroy one's personal or role based integrity. The appeal here is to a cluster of approaches oriented around the virtues one exemplifies in one's actions, called *virtue ethics*. (See the textbox on virtue ethics.)

When you are discussing an ethical issue, look for key words and phrases that show how certain theoretical approaches to ethics are embedded in their language. Try out the Mountain Terrorist Scenario with others to see the line of argument they adopt. If they are willing to play the game, do they argue for killing the villager and justify this choice in terms of its likely results? Then they may be partial to the utilitarian approach. Do they dismiss consequences and argue from the wrongness of the act of killing? Then they may be partial to the deontological approach. Often people try a mix of these two approaches. And of course there are numerous other approaches and many variations on these two approaches (again, chapter 12 covers some of this territory).

When discussing these scenarios (or any others) take the time to step back and reflect. When do you raise ethical issues? What kinds of positions do you take when discussing ethical problems? How do you argue for your positions? What positions do others take, and how do they frame their arguments? If we watch our language we find that we often employ such ethical considerations to make decisions, to justify them to others, and to anticipate and resolve possible problems. In short, ethics is embedded in our everyday life and in our language.

So watching your language is the beginning of systematic thought about what is ethical. As a practical matter, it is a good thing to do. Thoughtful computer professionals have

these skills, and they owe it to their clients to exercise them carefully. Can you identify the three different kinds of appeal we have just made for being careful in your ethical conversation?

Decision Making Tests:

Three of the ethics tests we propose here (reversibility, harm/beneficence, and public identification) conveniently encapsulate three widely recognized ethical approaches (deontology, utilitarianism, and virtue ethics). A fourth test, the code of ethics test, presents the special obligations that professionals have because of their expertise and because of the relations into which they enter. We will cover a fifth test, a feasibility test, in chapter 5.

Decision Making Tests—A Quick Summary

The Ethics Tests:

- **Harm/Beneficence:** Does it do less harm and more good than the alternatives? This test is about the alternatives for all stakeholders.
- **Public Identification:** Would I want to be publicly identified with this action? Would I want others to evaluate me as a person on the basis of this action? This test probes what the decision would show about your character.
- **Reversibility:** Would I think this a good choice if I were among those affected by it? This choice is about whether stakeholders are being treated with respect.

Professional Obligations:

- **Code of Ethics:** How does this choice relate to the ethical standards of my profession? This test is about the decision-maker's duties in the role of a professional.

Will it work?

- **Feasibility:** Can this solution be implemented given the constraints posed by such situational factors as time, technical limitations, manufacturability, economic parameters, the legal/regulatory climate, and the social/political environment? This test is about practical issues.

Once again, let's examine a short case, *Treasure from Troy*:

You are studying frantically for your exam in a computer engineering course. It will be very difficult. But your roommate, who is also taking the course and has the exam tomorrow, seems unconcerned. When you ask why, he tells you that he has a copy of the exam. Apparently, a group of students in the class found out how to hack into the professor's computer and download the exam. (They install a Trojan horse called Sub-Seven into the professor's computer which allows unauthorized access; then they search through the professor's files, find the exam

and download it.) Your roommate has the exam in his hand and asks you if you would like to look at it. What should you do?

Imagine that at this moment you could take a philosopher's time out to consider your options. The task in the next several pages is to help you do so by setting up your analysis, identifying your options, and using the tests to evaluate them.

The Set Up:

The first thing to do is to set up your analysis by...

- (1) *identifying the agent* (the person who is going to perform the action),
- (2) *describing the action* (that which the agent is going to perform),
- (3) *uncovering the stakeholders* (those individuals or groups who are going to be affected by the action performed by the agent), and...
- (4) *identifying and describing the consequences* (the likely results brought about by your action)

Treasure from Troy sets up as follows:

- **Agent:** You, the student studying for the computer engineering exam.
- **Action:** Taking the computer engineering exam from your roommate.
- **Stakeholders:** You, your roommate, the hacker classmates, the rest of your classmates, your teacher, the university
- **Consequences:**
 - you'll know the test questions in advance
 - you'll have a competitive advantage over those who haven't seen the test and do better if the exam is curved
 - you might get caught and punished
 - those who haven't seen the test will be treated unfairly
 - the teacher will be deceived

Let's look at this in more systematic detail.

Agent: *You, the student to whom the exam has been offered.*

How to Identify the Agent. The agent is the person or group of persons about to make a decision and act on it. So there you are, sitting in your dorm room with your roommate. He produces the test. He asks you if you would like to look at it. You have to decide what to do. There may be multiple agents, and there may be many other people attempting to influence the agent. But doing this analysis at the beginning level requires settling on one. When we get to considering more complex decisions in more complex socio-technical systems, it will be helpful to use some creativity in changing the system, getting agents to work together, etc. But for now, stick to one agent.

Why is this important? If you shift from one agent to another, you will begin shifting perspectives on the problem. This can lead you to wander off into speculation about others' motives or character and your whole analysis will become unglued. For the skilled ethical analyst, this sort of wandering can be informed by intuition and experience to reduce the time it takes to find a satisfactory solution. But for the beginner (and

occasionally for the expert) it can become an exercise in avoiding blame or difficult choices by focusing on others. Part of the skill of thinking ethically is having the perseverance to think through the difficult or threatening parts of the problem even when there are tempting detours. In sum, for now sorting out all the issues relevant to one agent will give you more than enough to do. Don't dilute your analysis by shifting from one agent to another.

Action: *Taking the computer engineering exam from your roommate.*

How to identify the action. This description above certainly gets things started, but it needs elaboration. If you decide to look at the exam, what does this entail? Do you look at the exam *in order to* study the questions *in order to* find out what the correct answers are? (Maybe you intend just to verify that it is the exam but intend to look no further.) That phrase, "in order to," can be used to group different things under a single description. Joel Feinberg (ref) calls this the *accordion effect* of action description; we use phrases like "in order to" to draw actions together into causal chains or to link actions to motives and goals. An action description functions like an accordion; it can be expanded or contracted as needed. Since our purpose is to evaluate the action alternative ethically, we should adjust our description to respond to questions posed by different ethical frameworks. Consider these accordion positions:

1. *Include a goal:* you are going to look at the exam *in order to* find out what the questions on the exam will be.
2. *Include consequences.* Reading the questions in advance will likely raise your grade.
3. *Classify the action under a moral concept.* Could this action be considered cheating? Are you looking at the exam questions *in order to* cheat? You could test this by reflecting on how the teacher would respond if she found out. Would *she* treat this as an instance of cheating? What would she do if you told her you had already seen the test questions? Perhaps another concept is relevant here. Since your teacher doesn't know that you have seen the test in advance, she will assume that you are seeing it for the first time when you take it. She will grade accordingly. Could you describe your action as a form of deception, i.e., leading the teacher to grade your exam under one impression (that you haven't seen the questions before), when in fact a different set of impressions is the case (you saw the questions to the exam last night). (Maybe deceiving the teacher is not your primary intention; what you really want is to get the best grade possible on the exam. But it may be a necessary—albeit regrettable—secondary consequence. Are you still responsible in this case, since you didn't intend it?)

Why is this important? Simple descriptions of actions can be expanded by bringing in the goals, motives, consequences, and relevant moral concepts. Expanding the action description helps to construct it in a way that makes one or more ethical issues relevant. This makes it easier to apply the ethics tests. It gives us an initial intuition about the ethical status of the action (or at least the dimension on which to evaluate its ethical status). We can now evaluate the action by examining its harms and benefits, its reversibility, and its impact on the agent's integrity.

Stakeholders: *You, your roommate, the hacker classmates, the rest of your classmates, your teacher, the university*

How to identify stakeholders. Strictly speaking, a stakeholder is any group or individual that has a vital or essential interest in what is about to happen. Thinking of a gambling stake helps here. You go into a casino and make a bet on the roulette wheel. Your stake is the money you put on the table, your bet. If you're lucky you'll win. If you're unlucky then you lose your money. A stake is something important to you that could be gained, kept, or lost in the situation at hand. It all depends on how things turn out.

In "Treasure from Troy" *you* are a stakeholder. Your grade is at risk. Looking at the test might help you to do better on the exam tomorrow. But this is not a risk-free benefit. If you get caught, then you will likely be punished. These (and other) considerations make you a stakeholder; your stakes are the benefits and harms you stand to gain or lose, depending on the outcome.

Your roommate is also a stakeholder. (What are his stakes?) Other *students in the class* are stakeholders. This includes both those who have prior access to the exam and those who do not. (What is at stake for these groups?) *Your professor* is a stakeholder. (What is her stake?) Can you think of any other stakeholders? (The university? Other professors? A Board of Academic Integrity?)

Looking through the socio-technical system is one good way to identify stakeholders. Not everyone in a socio-technical system is a stakeholder, but this approach may help you identify others who might not immediately come to mind. Socio-technical analysis encourages you to look at higher levels of organization (e.g. the school, clients of graduates, accrediting bodies, etc) that aren't immediate stakeholders but still may have a stake when the wider repercussions of the decision are considered.

Why is this important? For one, it allows you to set up a table of benefits and harms to do one of the ethics tests. But it also helps you identify people to whom you might have a duty, and people whose opinion about your character might matter. In more complex cases it can help you identify other people you can recruit to help in problem solving.

Consequences

How to identify consequences. These are the results likely to be produced by the action alternative you are considering. Here are some likely candidates:

- You'll know the test questions in advance.
- You'll have a competitive advantage over those who haven't seen the test and do better if the exam is curved.
- You might get caught and punished.
- Those who haven't seen the test will be placed at a competitive disadvantage.
- The teacher (if she doesn't find out) will be deceived.

This, of course, is a preliminary list. Brainstorming will expand it and make it more comprehensive. Then the list needs refining. You will want to sort out your consequences into harms and benefits and allot them to their respective stakeholders. You should also look at the *magnitude* (the degree and range of impact), the *likelihood* (the probability that the harms and benefits will, in fact, occur), and the *distribution* (who will benefit, who will be harmed). As you do this, of course, other agents, actions, and stakeholders may occur to you. It now is likely getting complex enough that you will need to keep notes.

Why is this important? Identifying consequences is in many ways where the rubber hits the road in the setup of the analysis. You may find that some consequences sound suspiciously like things that are relevant to each of the three tests. For instance, one consequence might be that if you refuse, your roommate will think differently of you. This can be phrased in terms of reputation in a community (the community of students). But it can also be phrased in terms of obligations you might have to your roommate. It is important to be very clear when specifying these things. Are we talking about a potential harm, a right or duty, or a reputation? Keep asking yourself precisely what it is, and if it is more than one thing, consider them all, but consider them separately.

Describing the agent, action, stakeholder, and consequences will set up your ethical analysis. This, in turn, will keep you focused throughout your ethical analysis. Now it is time to bring in the central analysis engine—the ethics and decision-making tests.

Harm/Beneficence Test

The Harm/Beneficence Test highlights an essential component of the utilitarian ethical approach, the goal of minimizing harm and maximizing benefit. In contemplating an action, we do our best to envision its consequences, especially those likely to take place and those whose occurrence would produce severe harm. We then compare different courses of action in terms of the benefit-to-harm ratio they are likely to produce. We attempt to maximize this ratio.

Steps in Applying the Harm/Beneficence Test

1. Set up your analysis (identify the agent, describe the action, identify those who will be affected by your action (the stakeholders) and list the impacts your action will have on these people).
2. Sort out the consequences into harms and benefits
 - a. What harms would accompany this action? Would it produce physical or mental suffering, impose financial or non-financial costs, or deprive others of important or essential goods?
 - b. What benefits would this action bring about? Would it increase safety, quality of life, health, security, enhance goods (financial, non-financial, moral)?
3. What is the magnitude of each these consequences? What is their likelihood? Magnitude refers to the severity of the impact of the consequence and the range of people affected. Likelihood is the probability that the consequence will, in fact, occur.

4. Identify one or two other viable alternatives and repeat these steps for them. Some of these may be modifications of the basic action that attempt to minimize some of the likely harms. These alternatives will establish a basis for assessing your alternative by comparing it with others.
5. Answer this question: Which alternative produces the best ratio of benefit to harm?
6. As a final step, check for inequities in the distribution of harms and benefits. Do all the harms fall on one individual (or group)? Do all of the benefits fall on another? If harms and benefits are inequitably distributed, can they be redistributed? Would such redistribution involve inflicting further harm?

Let's see what we can learn about *Treasure from Troy* from using the harm/beneficence test. Repeating the set up presented above we have...

- **Agent:** You, the student studying for the computer engineering exam.
- **Action:** Taking the computer engineering exam from your roommate and looking at the questions in order to do better on the exam tomorrow.
- **Stakeholders:** You, your roommate, the hacker classmates, the rest of your classmates, your teacher, the university
- **Consequences:**
 - You'll know the test questions in advance.
 - You'll have a competitive advantage over those who haven't seen the test and do better if the exam is curved.
 - You might get caught and punished.
 - Those who haven't seen the test will be placed at a competitive disadvantage.
 - The teacher (if she doesn't find out) will be deceived.

Carrying out the harm/beneficence test requires that we first expand on the above consequence list. Brainstorming possible consequences might produce a list like the following:

- You would do well on the test.
- You would effectively retaliate against the professor for giving such unfair exams.
- You would enjoy solidarity with the other students who have advance copies of the exam.
- You might graduate sooner and augment your grade point average.
- Graduating sooner would allow you to practice your skills sooner, give people the benefit from these skills, and allow you to get an advance start on earning money.
- The professor might find out that you (and others) had unauthorized access to the exam.
- The students who didn't have access to the exam before the test would not do as well. This would be unfair and lead to strained relations between these students and others who had access to the exam.
- This kind of activity, in the long run, would undermine the evaluation system of the class and even have harmful impact on that of the university.

- Looking at the exam ahead of time might undermine your moral integrity. It might be habit-forming.
- You might feel ashamed of your action and feel that your grade would not be deserved.
- This kind of activity might lead to students passing the exam without having the required knowledge. Graduating such individuals makes it more likely that innocent individuals would be harmed by incompetent and negligent professional practice.
- Should the teacher discover that her students had gained unauthorized access to her exam, she might adopt more aggressive preventive measures that could lead to falsely accusing and punishing innocent students.
- Knowledge of widespread cheating (or irregularities such as the one in this case) may lead teachers to impose more intrusive surveillance methods upon all students. This adds the injustice of increased surveillance of innocent students, i.e., those who did not do anything irregular in this situation.

This rather formidable list of consequences needs to be refined. The following questions help carry this out:

1. *Is this list comprehensive?* (Is it complete or does it leave out significant consequences?)
2. *Which consequences can be classified as benefits? Which as harms?*
3. *What is the magnitude of each consequence?* (Magnitude refers to the degree and range of impact. Are the harms severe, for example, life-threatening? Are the benefits merely trivial? Is a single person affected? A few? A large multitude? A harm of great magnitude would be an activity that threatens the lives of thousands of people, say, a meltdown in a nuclear reactor.)
4. *How likely is its occurrence?* (This refers to the probability that the harm or benefit designated will, in fact, occur.)
5. *How are the consequences (the benefits and the harms) distributed among the stakeholders?* Equally? How fair is this distribution? (Do some stakeholders stand to bear all the harms/risks without any of the benefits? Do others stand to receive all of the benefits without any of the harms/risks?)

A complete harm/benefit analysis would produce a list of consequences refined by these five considerations. Then it would identify other, competing alternatives of action, list their consequences, and refine the list. (It would be best to restrict yourself to only those that are serious competitors, i.e., the best available alternatives.) With two or more refined lists, it should be possible to compare them and choose the alternative that best maximizes benefits and minimizes harms.

Apply this to the *Treasure from Troy* case. What are serious competing alternatives to taking the exam? What are the consequences of these alternatives? How can they be refined by sorting them into harms and benefits, determining their magnitude, their likelihood and their distribution? You could use a matrix to help structure this work:

Consequence	Harm	Benefit	Magnitude (Degree and Range of Impact)	Likelihood	Distribution Among Stakeholders
You do well on the test		X	High impact on several people	Middle to high probability	Benefits distributed only to those who saw exam in advance
You (and the others) are caught by the teacher	X		High impact on several people	Fairly low probability	Harm falls on those who saw exam in advance

This is just one suggestion of a matrix for keeping track. But you will need to keep notes in some way, since the analysis has already gotten complex.

Classical Utilitarianism

The Harm/Beneficence test comes from the ethical theory of Utilitarianism which was first set forth by David Hume and fully developed by Jeremy Bentham and John Stuart Mill (see chapter 12 for our summary). Its major parts can be summarized in the following points:

1. *Utilitarianism is a consequentialist ethical approach; it evaluates actions and policies in terms of the consequences they are likely to produce.*
2. *The results produced by the action or policy are compared and ranked in terms of a theory of value based on the idea of intrinsic value (something valuable in itself and not merely as a means to something else). There are three kinds of utilitarianism based on three different approaches to value:*
 - a. **Hedonism:** Here, happiness (the Greek word for happiness is *hedone*) is the only intrinsic value; everything else is valued as a means to happiness (such as money) or as a part of happiness (such as virtue).
 - b. **Pluralism:** Pluralism differs from Hedonism by holding that there are several things of intrinsic value. Examples include pleasure, friendship, and aesthetic enjoyment. (See Rachels, 104)
 - c. **Individual Preference:** Individual Preference Utilitarianism abandons intrinsic value. Instead, they ground value in the satisfaction of individual preferences, whatever they are. The advantage here is that preferences can be determined through market interactions: market value or price measures can be thought of as tracking both preferences and their intensity.
3. Utilitarianism synthesizes its consequentialism and its value approach in its fundamental principle, the Principle of Utility. This principle enjoins us to *choose that action (or rule) that produces the greatest good for the greatest number.*
 - a. This formula (greatest good for greatest number) is based on a maximization/minimization strategy; one seeks to maximize the positive (whether it be happiness, pleasure, intrinsic value, or individual preferences) while minimizing the negative (whether it be unhappiness, pain, disvalue, or frustrated individual preferences).
4. In its initial formulation, utilitarianism provides a means of comparing and ranking individual action alternatives. However, two criticisms have been addressed to act

utilitarianism, that is, to that form of utilitarianism that applies the principle of utility directly to each individual action:

- a. Working through all of the consequences of all of the possible courses of action offered by a situation is practically impossible—there simply isn't enough time. We had to take a philosopher's time out to do it in the example here.
 - b. Act utilitarianism can be used to justify actions that our intuition tells us are immoral. Under act utilitarianism, accusing and punishing an innocent person for murder would be justified if the consequences maximized utility—say they avoided a destructive civil war. Yet punishing innocent individuals violates our sense of justice.
5. To avoid this problem, some utilitarians have embraced rule utilitarianism. Here the Principle of Utility is directly applied only to general rules of actions. For example, we ask whether utility would be maximized if we adopted truth-telling as a general rule. Should the rule satisfy the utility test, then we would adopt the rule and apply it (not the principle of utility) to individual actions.
6. Utilitarianism has been subjected to other criticisms. One, already discussed above in the context of the harm/beneficence test, is that utilitarianism is indifferent to the justness of the distribution of harms and benefits in a given situation. On this basis, many argue that utilitarianism fails as an ethical approach because it is not comprehensive. This is a good a reason to not adopt it as the only approach in our analysis.

Problems, pitfalls, and temptations with Harm Test

Too Little Moral Imagination: One may be tempted to stop too soon in drawing out the consequences of an action. Too little inquiry could stem from a lack of moral imagination. For example, in *Treasure from Troy*, one might consider only those consequences that have an immediate impact on oneself, leaving out the impact one's actions may have on others or the long term impacts they may have on oneself.

Remedy: The key here is to develop tricks to activate moral imagination. Here's one we use. Return to your stakeholder analysis (from the set up) and think about the different impacts your action will have on these stakeholders and their stakes. Will it enhance or diminish their wellbeing? Will it promote, maintain, or diminish their stakes? Here's another idea: try inventing scenarios and working out their possible results. Suppose in *Treasure from Troy* that the teacher becomes suspicious when a large number of students do extremely well on what is usually a difficult exam. She suspects something illicit has taken place but lacks proof. How would she react? Would she reschedule the exam? What impact would this have on you and the other students? Would she take her concerns to the university's Committee on Academic Integrity? Would they carry out an investigation that might lead to the discovery of the hackers? Maybe this committee would instigate a general policy of cracking down on cheating. What impact would this have on you and your fellow students? With these speculations in mind, can you rework their consequences and enter them into your harm/beneficence calculation as possibilities

or risks? Playing with different possibilities—even improbable or unlikely scenarios—can help uncover hidden consequences.

Paralysis of Analysis: The opposite problem of too little is too much. Trying to uncover all possible consequences can produce a “paralysis of analysis.” You list all the harms and benefits you can think of. There are so many already, and these are only the obvious ones. What about the unexpected? As you consider more and more possibilities, your resolve becomes unglued. Every slight modification of the primary action promises a bewildering array of good and bad consequences. Why not just flip a coin?

Remedy: A good stakeholder analysis can also help here by prompting you to distinguish between those who have something substantial or vital at risk in the situation and those who have only a minor or incidental interest at risk. By using a stakeholder analysis, you can limit the range of consequences you consider to those that stem from the impacts of your action on the key stakeholders.

Another approach to solving the “paralysis of analysis” is to bring in information from the other ethics tests. The public identification test would expose the impact of the action on your integrity. The reversibility test, because it focuses on your relationships with key stakeholders, would help you to uncover the harms and benefits your action would bring to these stakeholders (as well as the impact on their stakes). Bringing the other ethics tests to bear on this situation will highlight the key consequences on which you should focus your harm/beneficence analysis. This will, in turn, help you to limit the range of consequences you review to a list that is comprehensive, relevant, and manageable.

Lack of Justice Consideration: The harm/beneficence test also falls prey to a general problem with utilitarianism and other consequence-based ethical approaches: their tendency to neglect the justness or fairness of the circumstances created by the action. For example in the *Treasure from Troy* case, those who did not view the exam in advance must compete with those who did. This is unfair, even when the disadvantaged group does as well on the exam as the advantaged group; that is, it is unfair independently of the consequences. The unfairness lies in a competitive situation (created, in part, by your taking and viewing the exam in advance) in which one group has an unearned and unjustified advantage. The harm/beneficence test takes into account the total amount of harm and benefit produced. But it neglects this issue of justice or merit. In the *Treasure from Troy* case, it misses the unfairness that some people take the exam under different circumstances than others.

Remedy: Any use of the harm/beneficence test must take this into account. In general, the distribution of harms and benefits must be equal unless there is a strong, overriding reason for departing from the “equality default.” For example, greater benefits may be given to those who merit them or to those who have a greater need. (See the textbox on justice for more on this issue.)

An unfair or unjust distribution of benefits and harms or an unfair situation can be factored into the Harm/Beneficence test as another harm. In essence we count the harm

twice, once as a harm and then as an injustice. The other ethics tests can help here, since they are frequently more sensitive to unfair or inequitable situations and distributions. This is one of the many ways in which the decision-making tests compliment one another.

So, as you learn to think more clearly about justice in the distribution of harms and benefits (see the Primer on Justice textbox), you can sort your way through the socio-technical system, see which stakeholders have more or less power, and ask about the justice of the distribution of harms and benefits among them.

A Primer on Justice

A weakness of the Harm/Beneficence test is its tendency to gloss over the distribution of harms and benefits. The ethical concept that deals with this distribution and ethically evaluates this distribution is justice, in particular distributive justice. In this textbox, we consider briefly some recent discussions of justice.

There are three different senses of justice, and each involves a different sort of logic:

1. *Distributive Justice* examines how to divide and allot fairly the benefits and harms that result from social cooperation.
2. *Retributive Justice* concerns itself with the fair and impartial administration of punishment to wrongdoers.
3. *Compensatory Justice* scrutinizes how we fairly compensate those who have been wrongfully harmed by others.

All these have in common the idea of fairness or balance. But in the first case, it is the fairness of distribution among a set of people, while in the second two it involves repayment for harm, either with returned harm or returned good.

Distributive justice is most relevant to our concern here of fairly distributing harm and benefit. We need a robust approach to distributive justice to help us sort out these issues. Philosophers are good at giving robust interpretations of basic intuitions and the philosopher John Rawls is no exception. In his 1971 book, *Theory of Justice*, John Rawls constructed a thought experiment to find the basic principles of distributive justice.

1. He begins with the central problem of distributive justice: the goods, harms, and risks that accompany social cooperation must be fairly and justly distributed. Three methods of distribution present themselves as leading candidates: *equality*, *merit*, or *need*.
 - a. *Equality*: the benefits, harms, and risks of social cooperation are distributed equally among members of the social group.
 - b. *Merit*: the greatest share goes to those who deserve it; merit can be defined in terms of knowledge, skill, productivity or even moral virtue.
 - c. *Need*: the greatest share goes to those who have the greatest need.
2. We often choose among these as it pleases us in an argument to obtain our best advantage. But Rawls constructs a thought experiment designed to solve the problem of choosing among these different rules. Imagine a situation where a group of rationally self-interested individuals choose principles of distribution

- under a *veil of ignorance*.
- a. Rational self-interest leads us to acquire as many primary goods as possible. These include (a) rights and liberties, (b) opportunities and powers, and (c) income and wealth.
 - b. Under the *veil of ignorance*, we pretend to know nothing of our situation. As Rawls puts it, under the veil of ignorance, “no one knows his place in society, his class position, social status, race, age, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength and the like.”
3. The veil of ignorance channels rational self-interest toward an impartial and fair system of distribution.
- a. Without the veil of ignorance, those who are rich would gravitate toward a scheme of distribution that maintained and even enhanced their wealth. Those who were poor would opt for a scheme that redistributed the wealth of others to themselves. The scheme could also be shifted towards one’s natural talents: if one were strong, one would choose a system of distribution biased toward strength; if one were intelligent, one would choose a system of distribution that rewarded intelligence; if one were male, one would choose a system that favored men. Rational self interest without the veil of ignorance would bias the principles of justice chosen.
 - b. But the veil of ignorance pushes rational self-interest toward impartiality because the rationally self-interested individual must choose to protect all possibilities, not knowing in advance which one will apply.
4. With this in mind, Rawls’ basic position can be summarized in the following manner:
- a. Rational Self-Interest + Veil of Ignorance \Rightarrow Theory of Distributive Justice.
 - b. Distributive Justice, in turn, is captured by two principles: the *Equal Liberties Principle* and the *Difference Principle*.
 - c. *Equal Liberties Principle*: “First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others.” “The basic liberties of citizens are, roughly speaking, political liberty (the right to vote and to be eligible for public office), together with freedom of speech and assembly; liberty of conscience and freedom of thought; freedom of the person along with the right to hold (personal) property; and freedom from arbitrary arrest and seizure as defined by the concept of the rule of law.
 - d. *Difference Principle*: “Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone’s advantage [most especially to those most disadvantaged] and, (b) attached to positions and offices open to all....”
5. The Equal Liberties Principle has priority over the Difference Principle so that equality becomes the default pattern of distribution. Thus, any departure from an equal pattern of distribution must have a strong, overriding justification. Moreover, the equal distribution of political liberties is, for Rawls, absolute and cannot be overridden. (This is the basis of Rawls’ dislike of utilitarianism where

we can override basic rights and liberties to bring about the greatest good for the greatest number.) But, under the Difference Principle, a departure from equality can be justified in the economic sphere if all stand to benefit, most especially the disadvantaged. In this way, Rawls attempts a synthesis that captures the strengths of equal, merit-based, and need-based patterns of distribution.

6. Rawls' theory of justice has been intensely debated and scrutinized. Some interesting criticisms have emerged.
 - a. From the libertarian standpoint, Nozick criticizes Rawls for developing a system of justice that sacrifices liberty for equality. Nozick argues that a patterned system of justice (like Rawls') must continually interfere with a distribution voluntarily reached to maintain a privileged pattern of distribution. Nozick's example: we voluntarily transfer our money to Michael Jordan to see him play. But this transfer results in Jordan's having a disproportionate share of wealth, as judged by our ideal pattern of distribution. So to restore justice, we take back some of Jordan's money—through taxation—and redistribute it to those who gave it to him in the first place. Overriding the initial, voluntary transfer by a second involuntary transfer doesn't make sense to Nozick. Moreover, he finds it wrong because it sacrifices liberty to equality (or some other privileged pattern of distribution).
 - b. From a communitarian standpoint, Michael Sandel argues that Rawls starts with an overly abstract conception of the self and winds up with an overly abstract and inapplicable concept of justice. Rawls errs, according to Sandel, when he starts out with rationally self-interested individuals (one abstraction) and then adds the veil of ignorance (which further abstracts from real individuals in real communities). These two abstractions then produce a third: a theory of justice that cannot be brought into touch with the real world.
 - c. Sandel draws some conclusions from this criticism of Rawls. He argues that Rawls' methodology leads to a theory of justice that tries to be neutral in all moral disputes. As an example, he cites the disagreement between Lincoln and Douglas on slavery. Lincoln opposed slavery on moral grounds; during the debates he offered substantive moral reasons for rejecting it. Douglas, on the other hand, tried to remain morally neutral by advocating a procedural solution to the problem—each state or territory should make its own choice. The content of the choice, its moral substance, wasn't important to Douglas. What was important was that each state chose its own stance toward slavery; the process (free choice) and not substantive arguments (the moral character of slavery) should determine the issue. Sandel generalizes from this to criticize attempts to solve substantive moral issues by appealing to a just process or procedure. They fail because they don't make the substantive moral problems disappear; instead they import substantive moral positions in through the backdoor. Again, consider the Lincoln-Douglas debates. Douglas claimed neutrality; but allowing each state to decide the issue for itself by means of a democratic procedure, ignores the fact that slavery can only be

advocated on the basis of bad moral reasons. Procedural resolutions place bad substantive moral arguments on an equal footing with good substantive moral arguments. Procedural justice does not resolve substantive moral issues; it only makes it more likely that these issues will be resolved on faulty moral grounds or on non-moral grounds.

Public Identification Test:

The public identification test begins with the assumption that our identity is revealed through our actions. Moreover, what we do partially constitutes the person we become by creating the habits and character traits that together form our identity. Aristotle puts it nicely: we become virtuous by performing virtuous actions, and we become vicious by performing vicious actions.

If what you do reveals—or even partially constitutes—who you are, then your actions serve as a window through which your identity is laid bare. This provides the basis for the test: when you contemplate an action, you should ask whether you would want to be known as the kind of person who would perform that action. Would you want other people to judge you as a person in terms of that action? If the action is cowardly, would you want to be known as a coward? If the action is irresponsible, would you want to be known as irresponsible?

Built into the public identification test is a model of morally exemplary behavior. We observe and emulate moral exemplars; they provide us with models that concretely depict the kind of person we want to be. We compare ourselves to them, feel proud when we stand up to this comparison and feel shame when we fall short. In this way, the public identification test exhorts us to strive for a higher, ideal level of behavior.

Consequently, a crucial component of this test is the imaginative projection of a moral exemplar into the role of agent. (We'll call this moral model the Morally Exemplary Computing Professional or MECP for short.) Projecting the MECP into the agent's role, we ask whether the action under consideration is consistent with this agent's character. Furthermore, to provide content to this imaginary person, we add a list of virtues. Michael Pritchard (ref) has identified many of these by interviewing exemplary engineers and asking them to list the characteristics that make a good engineer. Among those identified were *responsibility, honesty, justness (fairness), courage, loyalty, articulateness, perseverance, creative imagination, habit of documenting work, civic-mindedness, openness to correction, and commitment to quality*. An additional virtue mentioned, *integrity*, refers to the overall strength of character exhibited by the moral exemplar; the MECP not only exhibits the virtues listed above, but can remain committed to these modes of behavior even when there is great pressure to do otherwise. The following textbox presents a summary of the steps in the Public Identification Test:

Steps in Applying the Public Identification Test

1. Set up your analysis (identify the agent, describe the action, identify those who will be affected by your action (the stakeholders) and list the impacts your action will have on these people).

2. Project a MECP into the role of the agent.
3. Would this MECP find the action consistent with her (or his) character?
4. To gain a better picture of the character of a morally exemplary computing professional, consider the following list of virtues:
 - *Responsibility, Honesty, Justness (fairness), Articulateness, Perseverance, Loyalty, Cooperativeness, Creative Imagination, Habit of Documenting Work, Civic-Mindedness, Courage, Openness to Correction, Commitment to Quality, and Integrity.*
5. Which virtues, if any, are manifested in this action? Which vices, if any, are manifested in this action?
6. As a check, project yourself into the role of agent and then associate the action with your character. Would you be proud of being publicly identified with this action? Ashamed? What does this action say about you as a person?

Let's look at *Treasure from Troy* in terms of the public identification test.

Set Up: (The same as above)

- **Agent:** You, the student studying for the computer engineering exam.
- **Action:** Taking the computer engineering exam from your roommate and looking at the questions in order to do better on the exam tomorrow.
- **Stakeholders:** You, your roommate, the hacker classmates, the rest of your classmates, your teacher, the university
- **Consequences:**
 - You'll know the test questions in advance.
 - You'll have a competitive advantage over those who haven't seen the test and do better if the exam is curved.
 - You might get caught and punished.
 - Those who haven't seen the test will be placed at a competitive disadvantage.
 - The teacher (if she doesn't find out) will be deceived.

Two characteristics of this action stand out: its deceptiveness and its unfairness. We have discussed both in the context of the harm/beneficence test. If you look at the exam the night before, you are not likely to go to the teacher and tell her of this. So by not informing the teacher, you—and the others involved—are deceiving her. With this in mind, project the MECP into the role of the agent. Because this action involves a deception (dishonesty), while honesty is the virtue exhibited by the MECP, it fails the public identification test; the MECP would not take the exam because she (or he) is honest. Second, the action is unfair to those who did not look at the exam ahead of time. So the action, reflected back on the character of the agent, reveals the vice of unjustness. This contradicts the virtue of justice present in the MECP. The MECP would do something other than take the exam, something consistent with her (or his) character. Since you are modeling your behavior on this exemplar, you also would look for another alternative.

One might argue that this action (taking the exam) manifests the virtue of loyalty. Loyalty to your friends, on this view, is indicated by your willingness to join in with the crowd for solidarity's sake. But loyalty has two vices; too little loyalty and too much. Following your roommate and others into a morally questionable action may exhibit loyalty, but loyalty of a peculiar kind, one blind to the wrongness of the act. (It, perhaps, represents too much loyalty, a willingness to go along with your friends no matter what.) Certainly, it is possible to express a different kind of loyalty. For example, isn't a loyal friend one who would point out to you the wrongness of what you are about to do? Aren't loyal friends supposed to help us to keep our integrity rather than encourage us to pursue projects that undermine it? Looking at the exam your roommate offers you manifests a shallow loyalty. A deeper and more fundamental loyalty is revealed by explaining to him that what he is doing is plainly and simply wrong? This use of the public identification test now casts your action as one of loyalty to your friend rather than the simple action of cheating on the test.

William May presents an interesting variation on the public identification test.

“One test of character and virtue is what a person does when no one else is watching. A society that rests on expertise needs more people who can pass that test.” (William F. May, “Professional Virtue and Self-regulation,” in *Ethical Issues in Professional Life*, ed. Joan C. Callahan, Oxford: Oxford University Press, 1988. p. 408)

The ethical professional acts virtuously even when—or especially when—nobody is looking. May also presents a list of what he sees as the virtues of professionalism, many of which are included in Pritchard's list: perseverance, public-spiritedness, integrity, veracity, fidelity, benevolence, and humility. May's test connects nicely with our public identification test. Those who act when nobody is watching as they would act under public scrutiny indicate their willingness to be publicly identified with their actions. Their actions provide no occasion for shame. May's test provides a good validation approach to the public identification test, one that can be used to cross-check the results of the procedure we set forth above.

Problems, pitfalls, and temptations of Public Identification Test

Drifting off the Virtue target: You may be tempted to reduce the public identification test to the harm/beneficence test by considering only the consequences of making the action public. You wouldn't take the exam your roommate offers because if you were found out (if the action were publicized) the consequences would be negative: you would be punished and receive a bad grade on the exam. In this misapplication of the test, the analysis focuses on the consequences of publicizing the action and not on how the action colors the moral character of the agent.

Remedy: Any utilitarian calculation (performed under the harm/beneficence test) would include these negative consequences. So this misapplication of the public identification test merely duplicates a part of the harm/beneficence analysis. But this misapplication also misses the fundamental point of whether this action is consistent with the character

and integrity of the MECP. Would the MECP take the exam? We project a moral exemplar into the situation and ask what this person would do. If the action is inconsistent with the character of the MECP, then it should be rejected. It is not that publicizing the action produces bad consequences so we shouldn't do it; this merely requires that we take precautions to prevent it from being publicized. It is rather that the action itself is bad because it stems from a bad person.

Staying focused on the MECP will help you to avoid this problem. Identify the action and then project the MECP into the situation as agent. Is this action consistent with this moral exemplar's character? To answer this question, you focus on the character traits that would lead to this action. Does it stem from the vice of irresponsibility? Then it is, on this test, a bad thing to do. Does it express virtues like honesty, justice, or integrity? Then it is a good thing to do. Taking the exam from your roommate in the *Treasure from Troy* case is wrong because it is inconsistent with what the MECP would do; it contradicts the virtues of honesty and justice that form essential components of the character of the MECP. It is not wrong because the results of publicizing the action—say the teacher finds out—would be negative and would harm the self interest of the agent. The public identification test connects the action with the character of the agent, rather than merely explore the consequences of a possible scenario, i.e., what would happen if the action suddenly were to become public knowledge.

Virtue Ethics

1. Virtue ethics differs from utilitarianism and deontology in that it emphasizes the agent in decision-making, not the action. In virtue ethics, we evaluate the agent, the one performing the action, in terms of certain general dispositions or habits of action that are revealed through what he or she does.
2. Aristotle defines a virtue as the disposition to aim at the mean relative to the person and situation.
 - a. For example, courage is the disposition (habit/tendency exhibited over time) to choose the mean between the extremes of recklessness (too much courage) and cowardice (too little courage).
 - b. The mean, the virtue of courage, is relative to the person (her strength and intelligence) and the situation. A person of little strength and small stature would be foolish to challenge a much larger, stronger person to a fight (relative to person). On the other hand, a soldier has to consider the situation. In a battle where the enemy has overwhelming advantage retreat may be the courageous course of action even though in other situations (where the sides are more evenly matched) it may be cowardly. Risking oneself when the odds are overwhelming is rash while placing oneself in danger (with a more evenly matched enemy) would inspire one's fellows to more effort and turn the tide to one's advantage.
3. Modern virtue theory has recognized that not all virtues can be expressed as the mean between the extremes. In general, a virtue is a habit of action that supports a social or professional practice or a social or professional role. For example, the practice of computing would flourish when its members practice the virtues of integrity, thoroughness, confidentiality, responsibility, curiosity, creativity, documenting one's work, etc.

4. By examining the virtues and vices manifested through her actions, we can put together a picture of an individual's character. If her actions manifest virtues and these virtues hold up even under strong pressure, then we say she is good, a person of integrity.
 - a. Moral exemplars (people who are good and are looked up to as models to be imitated) become important in virtue theory because they help connect a theory about people to a way to evaluate actions. In deciding how to act in a particular situation, one asks how a particular individual (a moral exemplar) would act were she in a similar situation. For example, should I copy the software against the licensing agreement? What would programmer X (someone whom I admire as a good person) do, were he in my situation?
5. When applied correctly, the publicity test captures important parts of virtue theory. If the action you are about to perform were to be publicized so that others would associate you with that action, would you be proud or ashamed? If you are about to tell a lie would you be ashamed of being found out and treated as a liar? The publicity test assumes, along with virtue theory, that a person's character manifests itself through her actions.
6. Finally, virtue theory leads to a consideration of responsibility issues. (Aristotle recognized this in his *Nicomachean Ethics*; after defining human virtue and human happiness, Aristotle turned to a discussion of moral responsibility.) We blame and punish people for their actions when these actions are voluntary and stem from the agent's character traits. However, when actions do not come from the character, when they are compelled by circumstances that overwhelm the agent or ignorance that deprives the agent of crucial insight into the situation, then we excuse the agent for these actions. Moral responsibility consists, in part, of the determination of when we can (and cannot) praise or blame an agent for her actions.

Chapter 12 includes a table that discusses several virtues by presenting a definition of the virtue, a description, its location between excess and defect, emotions and beliefs that support the virtue, and moral exemplars who manifest the virtue.

Reversibility Test:

The *Reversibility Test* captures a central idea in Kantian formalism, the idea of universalizing one's action (see the text box on Kant). It is also the main idea behind the Golden Rule. Positively, it tells us to do to others what we would have them do to us (the Christian formulation: ref). Negatively, it tells us not to do something to somebody that we would be unwilling to have them do to us (Rabbi Hillel: ref). Thus, we are treating reversibility as a key procedure to asking the universalizing question: "Would I recommend that all persons in this situation act this way?"

We also add the guidance that this test is about treating others with respect. Again, we do this in the spirit of Kant who sees the issue of treating others as ends (rather than merely as means) as an issue of respecting the autonomy of all humans. Kant sees the universalizing question and the ends question as essentially the same, and so we combine them into this test. We think it helps make the test a more faithful encapsulation of Kantian ethics.

Steps in Applying the Reversibility Test

1. Set up your analysis (identify the agent, describe the action, identify those who will be affected by your action (the stakeholders) and list the impacts your action will have on these people).
2. Using your stakeholder analysis, identify the key relations that you are going to reverse.
3. Reverse roles between the agent (you) and these stakeholders: put them in your place (as the agent or doer of the action) and yourself in their place (as the one subjected to the action).
4. Answer this question: If you were in their place, would you still find the action acceptable?

Cross checks for Reversibility Test (These questions help you to check if you have carried out the reversibility test properly.)

- Does the proposed action treat others with respect? (Does it recognize their autonomy or circumvent it?)
- Does the action violate the rights of others? (Examples of rights: free and informed consent, privacy, freedom of conscience, due process, property, freedom of expression)
- Would I recommend that this action become a universal rule?
- Am I treating others in this situation only as means to my own ends?

Let's return to *Treasure from Troy* to see how the reversibility test fits in. As we will see, the set up is important for carrying out the reversibility. In fact, most problems with this test come either from not setting the test up properly or from not sticking to the set up developed.

Set Up: (The same as above)

- **Agent:** You, the student studying for the computer engineering exam.
- **Action:** Taking the computer engineering exam from your roommate and looking at the questions in order to do better on the exam tomorrow.
- **Stakeholders:** You, your roommate, the hacker classmates, the rest of your classmates, your teacher, the university
- **Consequences:**
 - You'll know the test questions in advance.
 - You'll have a competitive advantage over those who haven't seen the test and do better if the exam is curved.
 - You might get caught and punished.
 - Those who haven't seen the test will be placed at a competitive disadvantage.
 - The teacher (if she doesn't find out) will be deceived.

Using this list, our analysis will reverse three relations: the relation between you and your roommate, you and your classmates, and you and the professor. You might choose some

other relations as important to reverse (e.g. you and your potential clients or employer when you graduate)

You and your classmates: Here we need to make a distinction between those who are in on the conspiracy (those who have the exam in advance) and those who are outside (those who will not have access to the exam before the test). If you were in the position of those on the inside, and they in yours, you would probably want them to take the exam copy offered to them. But the same considerations discussed just above apply here. We can distinguish between treating the stakeholder with respect and responding to the moral quality of their actions and wants. Your classmates in on the conspiracy (who have advance copies of the exam) want you to take the exam. Tell them respectfully why you are not going to do this. Then refuse the exam.

Turning to those outside the conspiracy, we find different results. If you were in their place (you did not have access to the exam) and they in yours (they were being offered the exam) would you still want them to take the exam and cheat on the test? Clearly not, because this would impose an unfair burden on you; you, who had not seen the exam before the test, would be competing with others who had. Thus, viewing the exam in advance would imposing an injustice on them that you would not have them impose upon you. On this relation, the action clearly fails the reversibility test.

You and your professor: This relation also fails to reverse. You would be treating the professor in a way you would not have her treat you. First, the students who used the program to gain access to the professor's computer did so without her knowledge and authorization. Consider this analogy. Imagine that your professor entered your house without your authorization and took photos of your personal correspondence and distributed these photos to her colleagues. You would find this unacceptable. So, by analogy, gaining unauthorized access to the professor's files to copy her exams is not reversible.

But others did this, not you. If you accept the proffered exam, you are merely taking advantage of a situation wrong brought about by others. What's wrong with this? Suppose again that your professor broke into your house and took photos of your personal correspondence. Then she offered these photos to another professor, say, your ethics professor; she wants to use them to show him what a slacker you are. Should your ethics professor look at the photos? He didn't take them in the first place; it was your computer engineering professor who did this. So he could argue that he was merely taking advantage of something done by somebody else, and he could really make use of the photos. Yet if he were in your position, he would not want you looking at photos of his private correspondence. So by analogy, taking the exam copy from your roommate is not reversible with your professor; you would not want her distributing materials that belonged to you without your knowledge or consent.

There is another argument here. The professor believes you are seeing the exam for the first time when she hands it out during the test. She also believes that your responses on the exam represent your knowledge and not the skill you and others have in hacking into

others' computers. She grades you accordingly. That these beliefs are unfounded constitutes a deception. You would not have others deceive you, so you should not deceive them. Consequently, taking the exam in order to cheat fails the reversibility because it exploits a deception, and you would not have others deceiving you.

3. *You and your roommate*: Before we reverse this relation, let's consider a variation:

Variation: Suppose your roommate asks you to find a way of getting an advance copy of the exam. He knows that you are good with computers and are aware of invasive programs that you can use to gain access to the information stored in another computer. So he asks you to use your knowledge to hack into the professor's computer and download a copy of the exam. He really needs to pass this course, and the teacher gives such hard exams. He strongly desires that you do this. Knowing that you are taking computer ethics, he asks you to perform the reversibility test. If you were in his position wouldn't you want him to do the same for you?

Something is wrong here, and it doesn't take much reflection to see what it is. Your roommate has attempted to force upon you a desire that you don't have. You don't want to hack into the professor's computer—he does. If you wanted him to do this, then—to be consistent—you should be willing to do the same for him. But—and this is the central point—you don't want to. Your roommate is using the reversibility test to manipulate you. He wants you to do something that you wouldn't do on your own. So he has asked you to set aside your standards, your judgment, and your freedom of will in order to lend yourself to his plans. In short, he wants you to abandon your autonomy.

Now, look at taking the computer engineering exam from your roommate in order to do better on the exam tomorrow. Your roommate offers you the advance copy of the exam. He wants you to take it. Knowing that you are taking computer ethics, he tells you that if he were in your position, he would take the exam. So you should too.

In responding to your roommate, it is important to distinguish between treating him with respect and doing what he wants. His wanting you to take the exam does not oblige you to do so. It requires, instead, that when you respond to the moral quality of his act you still treat him with respect. Recognizing and respecting your autonomy—not capitulating to the wrongness of your act—is what you would want your roommate to do were he in your place. So, refusing to take the exam and giving your roommate a reasoned account of why you are choosing this course of action is consistent with the reversibility. You can disagree with your roommate and still treat him with respect. The bottom line here is that the reversibility test should never require that you sacrifice your own autonomy to comply with someone else's wishes. Not taking the exam is reversible. It might be possible to defend the position that taking the exam was reversible too, in this relation. So perhaps it is a toss up for this relationship, since both actions might be reversed. But taking the exam is clearly not reversible in the first two relations we tried. This carries out the reversibility test for the most difficult relation and it also give us a clue to how to perform the action of refusing the exam: do so in a way that respects the autonomy of the other person.

Conclusion: We have seen that this action fails to reverse in two of our three relations. There are other stakeholders, and we could pause to apply the test to them. But a definite pattern has emerged. The results of the reversibility test converge internally and externally. Internally they converge in that reversing the different relations has produced similar results. Externally, they converge with the information we have already obtained from the harm/beneficence and public identification tests.

Kantian Formalism

Immanuel Kant put forth a highly influential moral theory in the 18th century. While his theory has close affinities with modern *deontology* (duty-based ethics) and modern rights theory, we will present it here in its original form. As such, Kantian formalism can be summarized in three central theses.

1. *Kantian formalism is a non-consequentialist ethical approach. For any non-consequentialist approach, the moral value of an action lies in the action's formal characteristics, not in the consequences of the action.*
2. *Kant next classifies actions according to their motive and their conformity to duty. Four different action-types result:*
 - a. *Actions that conform to duty but have motives other than duty. (I save a drowning boy because I expect a reward.)*
 - b. *Actions that conform to duty and are motivated by duty (duty for duty's sake). (I save the drowning boy with no expectations of reward and in spite of my intense fear of drowning.)*
 - c. *Actions that transgress duty for motives other than duty. (I wish nothing but ruin on the drowning boy's father, so I let the boy drown even though I could save him.)*
 - d. *Actions that transgress duty but were motivated by duty. (These actions are hard to characterize, but they appear to be miscarriages of implementation. (I try to save the drowning boy but am unable to carry out my duty-motivated intention; the boy drowns, I drown and nothing good remains but the purity of my initial motive.)*
3. *Kant's fundamental principle of morality, the categorical imperative, focuses on duty for duty's sake, (type b above). This imperative tells us to act only on that maxim (a rule one applies to oneself) that can be converted into a universal law (i.e., applied to all people, in all places, and at all times).*
4. *Violations of duty (which are also violations of the categorical imperative) exhibit a special kind of internal inconsistency: when I violate a duty—say tell a lie—I will duty as the universal law and seek to make myself the exception. (When I lie to you, I want you and everyone else to be a truth-teller, since I want my lie believed. If we were all liars, lies wouldn't work.) Hence, violations of the categorical imperative treat others with disrespect by bending them to our purposes. Kant converts this disrespect into another formulation of the fundamental principle of morality, the formula of the end: always treat others (yourself included) as ends and never merely as means.*
5. *As with the harm test and utilitarianism, we would argue that the reversibility test captures important parts of Kantian formalism, particularly when it is used across*

multiple roles. Checking to see that what one is about to do is reversible with a wide range of others is a kind of iterative approach to universalizing the action. It is also an important component of recognizing and respecting autonomy, i.e., of treating others as ends and not merely as means.

Problems, pitfalls and temptations of the Reversibility Test

The overly enthusiastic masochist: A masochist takes a class in computer ethics and is struck by the close relation between the reversibility test and the Golden Rule. He wishes to put this important moral principle into practice. Armed with the idea that he must do to others what he would have them do to him, he goes about inflicting physical pain on everyone he meets. But for some strange reason, the recipients of his good will do not seem to be all that appreciative.

Remedy—respect the person and respond to the act: We have already encountered this problem above in the *Treasure from Troy* case. Your roommate offers you the exam. He wants you to take it. If you were in his place (and he in yours), you would require that he do the same. But we saw that reversibility requires two things: that we respect the person with whom we are reversing and still respond to the moral quality of his (or her) action.

Respecting the person requires that we treat him or her as an end and never merely as a means. We must recognize and respect that person's autonomy, i.e., the capacity to develop intentions, design plans, and carry them out. Since autonomy includes the capacity to know along with the capacity to act, recognizing and respecting autonomy obliges us to provide others with pertinent information and to obtain their consent whenever we want them to join with us in a cooperative venture. In short, recognizing and respecting another's autonomy requires that we refrain from circumventing that autonomy by manipulating, deceiving, or compelling that person.

But respecting the person does not mean that we have to capitulate or consent to the wrongness of what that person is about to do. It requires, instead, that we continue to respect the person while responding to the wrongness of the act. Take an example. A team of Japanese engineers come to Puerto Rico to work with Puerto Rican engineers on a joint engineering project. But a member of the Japanese team, who holds traditional views on women, refuses to work with a female engineer on the Puerto Rico engineering team. He demands that she be taken off the project and makes this request to the head engineer of the Puerto Rican team. What should he do?

1. He could reassign the woman engineer to placate the Japanese engineer. He reverses with the Japanese engineer; were he in the Japanese engineer's place, he would want his cultural beliefs respected. But when he reverses with the woman he is about to reassign, he runs into a problem. He would not have her to do this to him were he in her place, so the action is not reversible in this relation. Are there other options?
2. He could tell the Japanese engineer to take it or leave it; in other words, to work with the woman on his team or go back home. This would respect the woman engineer on his team but it would fail to recognize or respect the cultural beliefs

- of the Japanese engineer. Are there any options available where he treats both stakeholders (the Japanese engineer and the Puerto Rican woman engineer) with respect?
3. He could try to reason with the Japanese engineer. He could tell him that the woman in question is highly qualified and that women engineers are quite common in Puerto Rico. He could try to negotiate with the Japanese engineer through the other members of the Japanese team. They have not objected to working with the woman engineer, so it is possible that they could persuade the reluctant engineer that he should follow suit. They might be able to appeal to his higher sense of duty or point out other loyalties that he has that supercede his traditional view on women. This option would respect the Japanese engineer by working with his autonomy instead of circumventing it. On the other hand, it responds to the wrongness of his discrimination toward the woman. So it is possible to respect the agent while responding to the moral quality of his action. (BTW: this was the action taken in this instance. It succeeded.)

So the overly enthusiastic masochist has really failed to respect the autonomy of the agents upon whom he is inflicting pain. He has not considered the possibility that they may want something different from what he wants. And here lies the problem with the Golden Rule, if it is simplistically applied. He has imposed his highly idiosyncratic standards upon others and, in the process, has treated them with disrespect. The bottom line here is that the reversibility test requires that we distinguish between respecting the person (the stakeholder with whom we are reversing) and responding to the moral quality of what that person is doing. It is possible (and probably obligatory) to find ways of reconciling the two even when this may prove difficult. We need to balance respecting the person with responding to the moral quality of his or her action, even when this action is wrong and requires a negative response. In the case of the reluctant Japanese engineer, the Puerto Rico team leader respects the Japanese engineer by seeking to persuade him rationally; but he refuses to give into the request to discriminate against the woman on his team.

If we build in certain insights from Kantian Formalism, we can find yet another way to formulate the remedy to the problem of the enthusiastic masochist. The masochist correctly recognizes that the Golden Rule involves reversibility because it asks us to place ourselves into the shoes of those who will be affected by our actions. But, he fails to recognize that it also requires universality. This is the insight that Kant expresses through the Categorical Imperative: act only on that maxim that can be converted into a *universal law*. The masochist's maxim (inflict pain whenever you can) fails when we attempt to universalize it; the absurdity of everybody being obliged to inflict pain on their fellows is readily apparent. The masochist has reversed with others but the standard or maxim he has used in this reversal cannot be universalized.

Drifting Off Target (again): Often we are tempted to reduce the reversibility test to the harm test. We do this by converting the question, "Would I think this a good choice if I were among those affected?" into the question, "What impact would my proposed action have on others?"

Remedy: Focus the reversibility test, not on the consequences of your proposed action, but on whether this action will treat with equal respect all those who stand to suffer impacts of your action. Does it undermine or circumvent their ability to make decisions for themselves? Then it would fail the test. Does it deceive, manipulate, or coerce them? Does it adopt a paternalistic stance by purporting to make decisions for them that they should be making for themselves? These are all instances where the reversibility test is violated. Results do enter into answers to these questions but the test does not turn on them. The central issue is whether the proposed action recognizes and respects the autonomy of others.

Rights Theory

Kantian formalism gives the human individual special respect which cannot be overridden even in situations where overriding would bring about substantial public utility. The respect due to the individual is grounded in autonomy, literally the individual's capacity to give the law to him or herself. Recognizing and respecting the sphere of autonomy that surrounds the individual, gives rise to a structure of rights and correlative duties.

1. Definition: *A right is an essential capacity of action that others are obliged to recognize and respect.*
 - a. All rights claims must satisfy three requirements. They must be (1) *essential* to the autonomy of the individual and (2) *vulnerable* so that they require special recognition and protection (on the part of both individuals and society). Moreover, the burden of recognizing and respecting a claim as a right must not deprive others of something essential. In other words, it must be (3) *feasible* for both individuals and social groups to recognize and respect rights claims. (refs)
2. Definition: *A duty is a rule or principle requiring that we both recognize and respect the legitimate rights claims of others.* Duties attendant on a given right fall into three general forms:
 - a. The *duty not to deprive* others of their rights or the duty not to violate the rights of others. Computing specialists have the duty not to deprive others of their rights to privacy by hacking into private files.
 - b. The duty, when feasible, *to prevent others from violating rights.* Hence, computing specialists have the duty to take measures to prevent individuals from using computing systems and computer technology to deprive persons of their privacy. This might entail helping a client to encrypt stored data. It could involve advising clients on how computing systems could be misused to violate privacy and require that computing specialists develop counter-measures to avoid these abuses.
 - c. Finally, when others have their rights violated, we have *the duty to aid them in their recovery.* So, for example, a computing specialist might have to serve as an expert witness in a lawsuit in which the plaintiff seeks to recover damages for having her right to privacy violated. Part of this duty would include accurate, impartial, and expert testimony.
3. A close inspection of the first two definitions reveals that *rights and duties are*

correlative; for every right there is a correlative series of duties to recognize and respect that right.

4. This allows us to develop a system of computing rights and correlative duties.
 - a. Specific rights (such as free and informed consent) can be defined. Moreover, some of the conditions of recognizing and respecting rights claims can be set forth.
 - b. Rights can be justified by showing that they are *essential*, *vulnerable*, and *feasible*.
 - c. Right holders can be specified.
 - d. Correlative duties and duty holders can be specified.
 - e. Finally, the correlative duty-levels can be specified as the duties not to violate rights, duties to prevent violations (whenever feasible), and the duties to aid the deprived (when this is feasible).

We have developed several computing rights in this manner in charts in Chapter 12.

Ethical Approaches vs. Comprehensive Theories

In what sense are the ethical approaches (utilitarianism, Kantian formalism, and virtue ethics) that underlie the ethics tests (harm/beneficence, reversibility, and publicity) theories? The temptation is to map all the characteristics of full-blown theories onto these ethical approaches. However, ethical approaches are more like different perspectives on a multi-dimensional object than mutually exclusive accounts of the same phenomena. Instead of contradicting one another, they complement each other; each compensates for the limitations of the others.

An analogy will help here. When we go to buy a house and the view from the outside is different from the view from the inside, we don't conclude that one view contradicts the other. What we have here are not two contradictory accounts of the same phenomena, but two distinct views of a multi-dimensional object. So in forming a comprehensive picture of the house we combine the different perspectives; we synthesize the two different, partial views into a complete and comprehensive view of the whole house. Something analogous happens when we turn to ethical approaches. Each approach offers a view of a different aspect of the moral action:

- Reversibility focuses on the *internal dimension of the action*, its formal characteristics such as consistency, reversibility, and universality.
- Harm/Beneficence offers an account of the *outer dimension of the action*, the action viewed in terms of its consequences or results.
- Publicity gives us insight into the action seen from the *standpoint of the agent*, the person performing the action. It views the action as the manifestation of the agent's character. Or, if the action is "out of character," it leads us to reflect on those circumstances which prevented the agent from acting responsibly in the situation.

This analogy conveys the importance of bringing all the approaches to bear in one's analysis of the situation. Each approach addresses the limitations of the others, because each complements the others. The proper ethical analysis employs all the approaches, uses each to address the inadequacies of the others, and works to harmonize and balance the differences that may emerge.

Codes of Ethics Test

The *code of ethics test* asks that the agent benchmark the proposed course of action with the recommendations of a professional code of ethics. Engineers, for example, should look at the impact of their decisions on public health, safety, and welfare; almost all engineering codes identify this as the area of paramount responsibility.

Steps in Applying the Code of Ethics Test

1. Identify the provisions in the code that are relevant to the case at hand.
2. Answer the following questions: Does your proposed course of action violate any of these provisions? Does it promote any?
3. Check for inconsistencies, i.e., instances where an alternative satisfies or promotes some code provision but not others. If there are inconsistencies, look for priority rules. For example, many codes hold public health, safety, and welfare *paramount*.

Problems, pitfalls and temptations of the Code of Ethics Test

The code is not specific enough: The code says nothing specific about the particular set of actions you are considering. In other words, there is a gap between the generality of the code provisions and the particular situations to which they are applied.

Remedy: This is more a characteristic of codes of ethics than a problem in their application. Codes are not about the answers to specific situations, but more about principles that are valued by the profession. Decision-makers will require some moral imagination to connect the principles to specific situations, and even ethicists with lots of moral imagination may not find much in a code that applies specifically. Take the stakeholder approach listed above as a way of opening up your imagination. Carry out the other tests (harm/beneficence, publicity, and reversibility) before tackling this one. Those tests may produce results that make this one more clear.

Guidelines for Interpreting Ethics Codes

- Most codes can be divided into sections organized around the relations between professionals and stakeholders of that profession. Key stakeholders are the public, clients, employers, users, peers (one's fellow professionals), and the profession itself.
- Internal conflicts can emerge between provisions that address these different relationships since each stakeholder group is organized around different goods and values.
- Be on the watch for inconsistencies. When confronting them, check the language for priority rules. For example, when codes hold public health and safety *paramount*, they mean it trumps the other responsibilities with which it may come into conflict.
- When the code itself does not provide guidelines for resolving inconsistencies, bring in other tools such as the ethics tests discussed above. Code provisions are largely based on these tests (or on the ethical approaches that these tests embody).

Hence they should always be used to check and supplement appeals to ethics codes. The tests help us to (1) resolve conflicts between code provisions, (2) fill in gaps left by the codes (gaps that arise when we move from the generality of code provisions to specific situations in which they are applied), and (3) resolve ambiguities in interpreting provisions in the codes.

- Identify the logic of different code provisions.
 - For example, some provisions serve to present *ideals of the profession*, goals toward which professionals should aspire. In ideals of the professions, the specifics are left undecided: they rarely specify circumstances such as when, where, for whom, and in what manner they are to be realized; these circumstantial matters are left up to the individual professional. Hence, while we should strive to realize these ideals, we cannot be punished for failing to do so since, lacking specification of circumstances, violation cannot be detailed.
 - On the other hand, *principles of professional conduct* are specified according to circumstances. Hence they can be violated, and professional societies frequently back them with sanctions. Work to discover the logic of code professions by sorting them into ideals of the profession and principles of professional conduct.
- Ethics codes often do not define the key concepts they contain. Public health, safety, and welfare may be held paramount but it is not clear who is the *public*, in what consists their *health, safety and welfare*, and what it means to say that all this is *paramount*. To understand codes of ethics, we must make an effort to clarify these concepts.