

## What is Ethics?

Ethics is the branch of philosophy concerned to discover truths about the moral world. Truths about the moral world, moral facts, are a bit peculiar. Later on it will become necessary to investigate quite precisely just what a moral fact is or could be. But for the moment we will rely on your intuitions. We will presume that you think certain kinds of actions, say recreational torture, are morally reprehensible. That truth, if it is one, expresses a moral fact. Similarly, you are likely to think that certain persons are heroic—for example the men on flight 93 who foiled the suicide hijacker's plans to crash the plane into a building. That truth similarly expresses a moral fact. In general, a moral fact is the possession of a moral property by some thing.

Some common moral properties are these: good, bad, obligatory, permissible, impermissible, evil, heroic, kind, cruel, mean, generous and the like. Some moral properties are less obvious: happy, excellent, better, worse and onerous, for example. All moral properties share something in common—they are evaluative, or, as ethicists say, *normative*. To have a moral property is to be valuable in some respect. The value may be a 'dis-value', in the same way a bad penny may be worse than no penny at all. On the other hand, not all normative properties are moral properties. For example, aesthetic properties, like beauty, pleasantness, and ugliness, are normative because they are valuations, but most ethicists would claim these properties are not moral properties. Similarly, manners, which are rules of behavior, are different in kind from moral constraints on behavior, though both are normative.

Just as only some properties count as moral properties, only some things are suitable for moral evaluation—only some things can, by changing the properties they have, change the facts about the moral world. Among the things that clearly can have moral properties are persons and collections of them, actions and kinds of actions, intentions, and states of affairs comprised by the moral properties possessed by sets of these things. There are many other sorts of things which some have thought to have moral properties: non-human animals, ecosystems, topographical features like mountains or forests, physical objects that have been constructed by persons or non-human animals, and so on. Whether or not such objects can in fact have moral properties remains an open issue in ethics.

Ethicists are in general concerned with a number of very basic questions, and some quite esoteric questions as well. Among the basic questions are these. How are we to differentiate between moral properties and other normative but non-moral properties? What moral properties are there and how are they related to one another? What kinds of objects can have moral properties? What are the moral facts that are true of human beings and their interactions with one another? Many ethicists hope that these questions can be answered by finding a simple but systematic description of our moral world. Such a description would be an *ethical theory* of our moral world, in the same way a comprehensive description of the physical world is a physical theory. The branch of ethics specifically concerned with the construction and testing of ethical theories is called Normative Ethics, and it is the branch we will look at most closely in this course.

### COMPREHENSION QUESTIONS.

1. What is a moral fact?
2. What feature do all moral properties have in common?
3. What does it mean to say that a property is normative?
4. Name two moral properties.
5. Name two normative properties that are not moral.
6. What is a moral theory?

### STUDY QUESTIONS

1. Name two moral properties not mentioned above.
2. Name two normative but non-moral properties not mentioned above.
3. Name two things that you think cannot have moral properties, and say why.

### **Theories of the Moral World: Nihilism and Realism.**

Many, many different theories of the moral world have been offered over the course of human history, and there are many, many others that have never been seriously considered, or even formulated. These theories differ from one another in various ways. Some differences are quite subtle, others are really big and obvious. The biggest differences concern whether there are moral facts to be systematized at all, whether if there are such facts they *can* be systematized, and if they can, the form of the theory that does the job.

Some ethicists, after considering matters, have reached the conclusion that the moral world is empty. These philosophers claim that there are no moral facts at all. Such theories have been given various names; we will refer to all ethical theories which deny the existence of moral facts as *nihilist* theories, and by Nihilism we will mean a commitment to some theory of this sort. We will say that any theory which is not nihilist is *realist*, and by Realism we will mean a commitment to some non-nihilist theory.

Nihilist theories offer a very simple systematization of moral facts, since according to them there are no such facts to systematize. But this leaves the nihilists with a problem. Nearly all human beings, in nearly all cultures, say and do things that seem to make sense only if there really are moral facts. We justify our actions to ourselves and to one another by appeal to moral obligations. We blame one another and ourselves for failing to live up to principles we regard as expressing moral truths. We punish others when we blame them in this way, but regard punishment itself as unjust when we do not hold those who suffer the punishment to be blameworthy. All of us have recognized, or thought we recognized, instances of injustice or unfairness, instances of heroism, of cowardice, and of other virtues. Most generally, nearly all cultures and nearly all people have recognized, in practice, a distinction between the way the world *is* and the way the world *ought* to be. Is all this so much nonsense? If not, what sense can we make of it if there are no moral truths? That is the explanatory problem confronting nihilist theories.

Different versions of nihilism attempt to resolve the problem in different ways. One common way is this. Nihilists claim that utterances and actions which appear to express moral commitments really express merely attitudes. To say that recreational torture is evil is not, on this view, to make a claim that is true or false, or could be true or false. It is instead to express an attitude (namely that expressed by ‘Yuck!’) towards instances of recreational torture, much as we might express an attitude towards boiled oca, fried cockroach, or modern art. Such versions of nihilism have been called ‘Yea! Boo! theories of morality’. This kind of *rational reconstruction* of our everyday behavior and everyday ways of speaking tries to make sense of what we do and say without taking at face value the truth of what we say, or the correctness of the reasons we give for doing what we do.

Realist moral theories therefore seem to have an initial advantage—they can give a common-sense explanation of our everyday appeals to moral facts to justify our actions. According to realist theories, these appeals are claims about how things are in the moral world. As such, they are either true or false. When they are true, our actions *really are*

justified. On the other hand, realist theories face difficulties of their own. Among them are these: what are moral facts and how can we tell what the moral facts are—how can we tell whether a given moral claim is true or not? These are serious problems in their own right. Despite the difficulties moral realism faces, it is the more common view among philosophers, and far the most common view among non-philosophers. In the remainder of this course we will assume that Realism is true—there are moral facts, and claims about them are either true or false. Two sub-disciplines in ethics concern themselves with the dispute between moral realism and nihilism, Moral Epistemology and Meta-Ethics. In the modules that follow, some issues in moral epistemology will make an appearance, but for the most part we will not deal with meta-ethical questions.

#### COMPREHENSION QUESTIONS.

1. What is Ethical Nihilism?
2. What is Ethical Realism?
3. According to nihilistic theories, what do sentences that seem to make moral claims really express?
4. What is 'rational reconstruction'?
5. Why does a nihilistic theory of morality have to give a rational reconstruction of our everyday ways of speaking and behaving?

#### STUDY QUESTIONS.

1. Find a sentence in a newspaper op-ed piece or letter to the editor that appears to express a moral claim, but which you think actually expresses an aesthetic attitude, and explain why.
2. Find a sentence in a newspaper op-ed piece or letter to the editor that expresses a moral claim that you think is false, and say why.
3. Find a sentence in a newspaper op-ed piece or letter to the editor that expresses a moral claim that you think is true, and say why.
4. Suppose a scientist is falsely accused of committing faking data in a published paper, and though no investigation is performed, she is widely held to be guilty and on this basis denied access to federal grants. Assume that nihilism is true. Explain whether and why denying her access to federal grants is wrong.

## Theories of the Moral World: Situational and Systematic Ethical Theories.

All realist ethical theories agree that there are moral facts. But beyond this there is very little agreement. One of the most basic differences among realist moral theories concerns whether ethical facts can be systematized by a simple, comprehensive set of principles.

Most realist ethical theories hold that the moral world can be described by a simple, comprehensive set of principles, but not all do. We will call those theories that deny the possibility of a systematic ethical theory *situational* theories. Among situational theories the most important are versions of *Virtue Ethics*. The first and most important such theory was developed by Aristotle very early in the history of philosophy. According to virtue ethics, the morally correct thing to do in any given situation is whatever a *virtuous* person would do. Virtuous people, according to this theory, have certain features of character, virtues, such as courage, kindness, integrity, self-respect, and so on. These virtues enable a virtuous person to perceive, to ‘see’, what one ought to do in any given situation. No one is perfectly virtuous, of course, but the more virtuous a person is, the better able he or she is to perceive the correct action in more and more situations.

One might think that it would be possible, nonetheless, to provide a theory that explains what makes virtuous actions virtuous. Not so, say the virtue ethicists. Suppose you made a list of virtuous actions and specified for each action all the relevant facts about the contexts in which those actions occur. One could then look for differences in context which might explain why in one context one kind of action is virtuous, while in another context a quite different kind of action is the thing that ought to be done. One might hope that if you looked at enough virtuous actions in enough contexts, one could find principles connecting context to action. Such principles might look like commands: do not kill innocent persons; save lives unless this requires killing innocent persons; be honest unless dishonesty would save innocent lives; be kind unless being kind requires killing an innocent, or failing to save innocent lives, or dishonesty; and so on. But virtue ethicists say this hope is in vain. The features of context that matter are so numerous, and of such importance, that for any set of principles one can find, there is some context in which the virtuous action will be different than that required by the principles.

Because this is so, virtue ethicists say that you cannot teach people how to act as they ought by teaching them a set of principles to follow. Any such set of principles will necessarily lead people astray. Instead, virtue ethicists say that to teach people to be moral one gives them an education that fosters virtues of character. Persons of virtuous character will be able to see the correct thing to do, whatever the situation may be. Even better, being virtuous, they will be motivated to *do* the right thing once they recognize it, since a disposition to act in virtuous ways is itself a virtue.

Situational ethical theories, and virtue ethics in particular, have much to recommend them. But they also face serious problems. One of these is that such theories are, by definition, unprincipled. A second is that virtue ethics makes moral epistemology impossible—there is simply no way to tell whether a given action is morally correct or not, unless you happen to be virtuous, and no way to tell whether or not someone,

yourself or another, is virtuous. Finally, and perhaps most seriously, such theories make it impossible to rationally adjudicate a dispute about what ought to be done in particular cases. If two persons disagree about what ought to be done, this is because they simply perceive the moral world differently. Because the differences lie in perceptions rather than principles, no accommodation can be reached. The situation is exactly analogous to matters of taste. For those who dislike opera or boiled oca or modern art, no amount of reasoning will change their basic perceptual experience—for us dislikers opera sounds bad, boiled oca tastes bad and modern art looks ugly, that is a fact about us that cannot be changed by reasoning (though, of course, it might be changed by a long period of training aimed to change our tastes). If situational ethics is true, then it is just the same when people disagree about moral questions; compromise may be possible, but reasoned resolution is not.

Opposed to situational theories are theories that endorse the possibility of a systematic description of the moral world in terms of principles. We will call such theories *Systematic* ethical theories. The dispute between ethicists that endorse unprincipled, situational ethical theories and those who endorse principled systematic theories is two thousand years old and not likely to end soon. But for various reasons we will assume that it is possible to provide a systematic ethical theory. Our reasons here are varied. Partly they are pedagogical—it would be impossible by means of an ethics course to teach students how to think about ethical problems if situational ethics were correct. Instead, we would need to engage students in along period of quite varied training to inculcate virtues. Since we cannot do that, but we would like to provide you tools to help work through ethical problems, we will assume that situational ethics is mistaken. Also, for various reasons, we think the arguments in favor of systematic theories are just much better than those in favor of situational theories, and so systematic ethics is much more likely to be correct. Though we haven't the space to deal here with the arguments, there are serious arguments on both sides, and interested students are encouraged to consider them as occasion arises.

#### COMPREHENSION QUESTIONS.

1. What is the difference between situational and systematic ethical theories?
2. According to virtue ethics, what makes the morally correct action in a given situation correct?
3. What is the most serious practical problem confronting situational ethical theories?
4. In what sense are situational ethical theories *unprincipled*?

#### STUDY QUESTIONS.

1. Name two virtues of character that are not mentioned above.
2. For each of the virtues you named, describe a context and an action in that context which fails to evince that virtue of character.
3. Is the action immoral? If so, can you say something about why it is immoral?

## Theories of the Moral World: Absolutism vs Relativism.

Both situational and systematic moral theories are *realist* in the sense that they all think there are moral facts. Judgments about whether or not a particular action or kind of action is morally permissible, obligatory or morally impermissible are, for both kinds of theories, either true or false. Systematic theories, unlike situational theories, claim that these facts can be systematized by a set of prioritized moral principles such that actions which conform to the principles are morally correct, while actions that violate the principles are morally incorrect. Among systematic theories, however, there is a big difference between two different ways of understanding these principles.

One kind of theory, known as *moral* or *ethical* relativism says that which principles apply to a particular person depend on who that person is and on their cultural and personal background. Conversely, *absolutist* moral theories say there is just one set of principles, and they hold for everyone, no matter who they are, where or when they live or lived, and no matter how, or even if, they were taught to think about moral issues. In a sense, ethical relativists hold that there are many different moral worlds, each systematized by different sets of moral principles. Which world a given person inhabits depends on the particulars about that person, their culture, education and situation. Absolutists, on the other hand, claim there is just one moral world, just one set of moral principles, and those principles apply to all of us.

There is an easy confusion about absolutist theories which must be avoided. Sometimes people think that an absolutist theory *must* say that what one *ought* to do is entirely independent of any extenuating features of context. This is NOT so. Both relativist and absolutist theories can and do claim that context matters. They differ in *how* context matters. The relativist says that features of context can *change which rules apply to a person*. The absolutist says that features of context which matter *must be written into the rules, which always apply to everyone*. So both a relativist and an absolutist might agree that killing is in general wrong, but that at least sometimes killing is self-defense (an extenuating feature of context) is permissible. But the relativist might, whereas the absolutist will not, allow that it might be that the rule allowing killing in the exceptional circumstance of self-defense will not apply to those from a culture which forbids *all* killing. So, e.g., for the relativist there will be special rules for Amish; the absolutist will deny *this* kind of contextual relativization, and will hold that the rules always apply to everyone.

Ethical relativism has been increasingly popular outside philosophy over the last century, for a number of reasons. One reason is that ethical relativism looks a lot like a different principle, called *methodological relativism*, which is essential for doing certain kinds of humanistic science and humanities, cultural anthropology, history and literary studies of culture for example. The aim, in these disciplines, is to come to an understanding of the culture and practices of another society. Cultural anthropologists recognized early on, and others have recognized since, that it is practically impossible to understand a culture in its own terms if, while studying the culture, one makes moral or aesthetic judgments about that culture or its practices. These judgments are necessarily informed by your

own cultural practices, and so will lead to a deep misunderstanding of your subject. Methodological relativism requires that in studying other cultures one abstain from making any moral judgment of those cultures at all.

Methodological relativism is a perfectly coherent restriction on those who engage in cultural studies. But it is importantly different from *ethical relativism*. Ethical relativism says that though there are moral truths, and these moral truths can be systematized, moral truth and its systematization is relative to specific cultures. There is no one *universal* moral code that applies to everyone. Instead, there are different, equally true, moral codes for each distinct culture. Actions which would be morally heinous if taken by members of one culture may be morally innocuous or even obligatory if taken by members of another culture. Whatever the justification of methodological relativism, it is no reason at all to adopt the quite different claim of ethical relativism.

Aside from this confusion between methodological and ethical relativism, there are three kinds of reasons that have lead philosophers and others in the humanities to adopt ethical relativism, and we need to briefly examine all of them.

Reason 1: The argument from diversity. The first reason is sometimes called the argument from diversity. The argument begins by noting that, as a matter of fact, cultures differ widely in the kinds of actions they endorse as moral, and the kinds of actions they forbid as immoral. Some cultures have permitted infanticide, while others proscribe such killings as seriously immoral; some cultures have permitted, even required, human sacrifice, while others regard such killings as seriously immoral; some cultures have permitted slavery, while others decry it as seriously immoral; some cultures have permitted, or even required, racial or religious discrimination, while others decry it as seriously immoral, and so on. The argument from diversity infers from the fact of such widespread disagreement among cultures that there are no *universal* moral truths, there are only moral truths that pertain to particular cultures.

Reason 2: The argument from moral ontology\*. The second reason concerns what makes a moral truth true. Suppose there were a single, true, universal moral theory. Then by the lights of that theory, the practices of some cultures would be seriously immoral. For example, human sacrifice is either seriously immoral, or if not, practices that forbid human sacrifice are seriously immoral. Since our society forbids human sacrifice, but other cultures have permitted it, either we or they must endorse practices that are seriously immoral. But now consider somebody who endorses the moral theory that, we are assuming, is the one true universal moral theory. What makes that theory so special—why isn't it just another moral code, one among many, endorsed by another culture, one among many? By what right does that culture claim that its moral theory is better than the theories endorsed by other cultures? And what kind of fact could make such a claim true? Those convinced by the argument from moral ontology think there

---

\* 'Ontology' here refers to a specific kind of metaphysical theory, namely a theory about what sorts of things there are. So, for example, scientists differ about ontology when they disagree, as they did in the 1800s, about whether there are atoms, and again when they disagree, as they did for much of the early part of the 1900s, about whether there is any such thing as causation.



could be no such fact. There are many moral standards, each endorsed by some culture, actual or possible, and no standard has any more claim to being a *universal* truth than any other. So the basic argument is this. If there is one true universal moral theory, there must be some set of facts which make it uniquely true. There are no such facts. So there is no one true universal moral theory.

Reason 3: The argument from moral epistemology\*. The argument from moral epistemology is closely related to the argument from moral ontology, but instead of being concerned with facts that *make* one moral theory universally true, this argument is concerned with how we could come to *know* that one moral theory is universally true. The idea here is this. Suppose we make a judgment that some particular moral theory is the one true universally applicable moral theory. When we make that judgment, we do so from the perspective of our own culture and our own prior moral commitments. Our judgments will not be shared by those from different cultures or with different prior moral commitments. Even if we suppose that there is a universal moral truth to be discovered, we ought to recognize therefore that our own judgments about what that truth is might be in error. But once we recognize the possibility of error, we must also recognize that the possibility of error is unavoidable—there is no set of observations, no process of reasoning that can ensure that our final judgments are correct. So at the very least, we are not entitled to judge others and their practices with any certainty; we might after all err in making those judgments. Those who are convinced by reason 3 then go one step further. They say that when there is no method to conclusively verify a proposition, the proposition must be regarded as false or meaningless. It follows for every moral theory that it is false or meaningless to say that this theory is the one true universal moral theory. Hence there is no one true universal moral theory, there are instead many different, equally true, moral theories, each true for different cultures.

#### COMPREHENSION QUESTIONS.

1. What is the difference between moral absolutism and moral or ethical relativism?
2. What is the difference between moral relativism and methodological relativism?
3. Can an absolutist theory allow that features of context count as extenuating circumstances?
4. What is the argument from diversity?
5. What is the argument from moral ontology?
6. What is the argument from moral epistemology?

#### STUDY QUESTIONS.

1. Describe a flaw in the argument for diversity: is it possible for the premise of the argument, the claim that different cultures disagree about moral principles, to be true but for the conclusion to be false? Give an analogy.
2. Find two moral principles, endorsed by your culture but not endorsed by one or more other cultures.

---

\* ‘Epistemology’ refers to a theory of knowledge—a theory about what knowledge is, what things can be known, and how one comes to know. An epistemic concern is a concern about how one comes to know. Scientists worry about epistemic issues when they worry about proper method (e.g. where and when will regression give information about causes, and where and when are experiments required).

3. What reasons can you give for thinking that your culture is correct to think that these principles are true *for everyone*, whether or not they are members of your culture; what reasons can you give for thinking that these principles are not true for everyone?
4. Find a moral principle that you think is true for anyone in your culture, but not true for someone from a specific different culture: that is, find a principle such that you think that if anyone from your culture violated it they would be doing something wrong, but that someone from some other specific culture could violate without doing anything wrong.
5. What reasons can you give for thinking the principle you found for question for is a *moral* norm, rather than an *aesthetic* norm, i.e. a rule of manners, or something similar, rather than a constraint on ethical conduct?

## Theories of the Moral World: Responses to Relativism.

Though some of the arguments for relativism are very powerful, none are completely compelling. In this module we review some difficulties in the reasons given for relativism. In the next we will consider some reasons for adopting realism.

Consider the first argument for relativism, the argument from diversity. Though initially compelling, the argument is in fact invalid. It is perfectly possible for the premises to be true while the conclusion remains false. The argument from diversity therefore does not give compelling reason to adopt moral relativism.

By definition, an argument is said to be *valid* only if it is impossible for matters of fact to be so arranged that the premises are true and the conclusion false. Arguments that are not valid are said to be *invalid*. Validity and invalidity, in this logical sense, are matters of structure rather than of content. That is, the relations between premises and conclusion which guarantee that *if* the premises are true *then* the conclusion must also be true are relations that hold in virtue of the *syntactic* or *grammatical* structure of the sentences which are the premises and the conclusion, rather than in virtue of the *subject matter* of those sentences. Consequently, one can test the validity of an argument by substituting nouns systematically throughout the sentence. For example, consider the argument W:

W) All whales are mammals.  
Some mammals live in the ocean.  
Therefore,  
All whales live in the ocean.

W happens to have true premises and a true conclusion. Nonetheless, the argument is invalid because it would be possible for an all powerful being to arrange matters so that the premises were true but the conclusion false. We can see this by considering the grammatically identical argument H.

H) All horses are animals.  
Some animals live in the rainforest.  
Therefore,  
All horses live in the rainforest.

H differs from W only in that throughout the argument ‘horses’ has been substituted for ‘whales’, ‘animals’ for ‘mammals’ and ‘rainforest’ for ‘ocean’. The result, argument H, has true premises and a false conclusion. H must necessarily be invalid—since it actually has true premises and a false conclusion it is certainly possible that it have true premises and a false conclusion. And the invalidity of H, combined with the fact that it has the same *syntactic structure* as W (guaranteed by the fact that we can convert W into H by substituting noun expressions throughout), entails the invalidity of W—it is possible for W to have true premises and a false conclusion, even though as a matter of fact its premises and conclusion are all true.

H can be used to show that W is invalid because they share a structure or *form*. The form of both arguments is this:

All X are Y  
Some Y are Z

Therefore

All X are Z.

By appropriate substitutions of noun expressions for X, Y and Z one can turn get W or H from this form.

The argument from diversity is similarly shown to be invalid. It has the form:

Cultures disagree in their beliefs about X.

Therefore

There is no universal truth about X.

To get the argument from diversity from this form, substitute 'moral principles' for X.

The resulting argument from diversity has a premise that is known to be true. It is unknown whether the conclusion is true, and the argument is offered as a reason for thinking the conclusion *is as a matter of fact* true.

But instead of 'moral principles' substitute 'germ theory of disease'. We then get the argument G.

G) Cultures differ in their beliefs about the germ theory of disease.

Therefore

There is no universal truth about the germ theory of disease.

This argument has a premise that is clearly true, since not all cultures endorse the germ theory of disease. But the conclusion is clearly false. This shows that G is invalid—it does not have a structure such that the truth of the premises guarantees the truth of the conclusion. Consequently, the argument from diversity too has a structure such that the truth of the premise does not guarantee the truth of the conclusion. Hence the argument from diversity is invalid.

The argument from moral ontology is more difficult to address. But it too is invalid. In effect, the argument fails to consider two alternatives open to an absolutist moral theory. The argument from moral ontology demands of absolutist theories that they provide some set of facts which would make one moral theory true as against alternatives. Absolutists have two options here. First, they could offer a reductivist theory, according to which moral truths, moral facts, can be reduced to non-moral facts, and these facts could be non-moral normative facts, or even non-normative facts. Such a reduction would look much like a reduction of, say, thermodynamic facts to facts about quantum statistical mechanics, or a reduction of chemical facts to facts from more basic physics. Of course, if such a reduction is to work, sooner or latter absolutists need to provide it. Several reductive strategies have been tried. Perhaps the most popular seeks to reduce facts about morals to facts about rationality. The idea is that non-moral actions turn out, in the end, to be *irrational* in some specific sense. While no such reduction has too date been entirely satisfactory, neither is it clear than no such reduction can be given. Perhaps the most influential such attempt comes from Kant, but many others hold to something like this view.

A second alternative is for an absolutist simply to deny the legitimacy of the demand for some fact which 'makes' the correct moral theory true. To see the motivation here, consider a similar demand made of a fundamental theory in physics—say the true theory

unifying gravitational and quantum mechanical phenomena. For such a theory we can give *evidence* that it is true, namely that observable phenomena conform to the predictions of the theory. But if we ask, why should it be this theory, rather than any of the other possible theories, that is true, there will be no answer—that question is beyond the scope of the scientific enterprise. If an answer were possible from within science, our theory wouldn't any longer be fundamental, since we could explain it by appeal to some even more fundamental theory. And if we ask why that more fundamental theory, rather than any other, then again we ask a question whose answer is beyond the scope of science. Similarly, the normative ethicist can say, 'Look, we absolutists are in the business of finding out what the true universal moral theory is. We are not, and cannot be, in the business of telling you *why that theory rather than some other* is the true theory. Whatever facts, if any, account for this are not moral facts, and hence fall outside the scope of moral theory. All we can do is give you *evidence* about which theory is true, and that is sufficient for our ends.' And in fact there is a special domain of philosophy, called Meta-Ethics, and one question that arises in that domain is exactly what, if anything, 'makes' moral facts true.

The argument from moral epistemology is perhaps the most interesting and challenging of the arguments for moral relativism. The central charge is that sentences whose truth cannot be conclusively verified are either false or meaningless, and for any moral theory the claim that just that theory is a universally true moral theory cannot be verified. The challenge is correct at least in the second part. As far as we know, there is no way of conclusively verifying any moral claim, even for example the claim that recreational torture is morally wrong. On the other hand, there are very good reasons for rejecting the first part of the challenge, namely the view that sentences which cannot be conclusively verified are false or meaningless.

First note that on many views about how science works, this is true of scientific claims as well. For example, the most influential philosopher of science among scientists was Sir Karl Popper. Popper held that a scientific theory can be conclusively *falsified* or disconfirmed, but it can never be conclusively verified. If Popper is right, and so too is the argument from moral epistemology, then scientific truths are just as relative, just as culture dependent, as moral truths. A consequence of this view, for example, is that prior to Copernicus, the sun and the planets *really did* orbit the earth. Those who are unwilling to accept this result should be similarly unwilling to accept the argument from moral epistemology.

Second, note that there are many claims we accept, or anyway think might be true, that cannot be conclusively verified. The most obvious case is that of 'other minds'. Presumably you think that other people have thoughts, feelings and desires in much the same way you do. You appeal to these features of others, to their states of mind, to explain and predict their actions, and people begin doing this very early in childhood. But think for a minute about what it would take to *conclusively verify* the claim that other people have minds. Their thoughts and feelings are not available for direct inspection; they must instead be inferred from behavior. And surely it is possible, physically, for the same behaviors to be produced by physical systems that do not have minds. There seems

no way to be absolutely certain that others have minds. Nonetheless, all sane people do think this. To that extent, we should reject the central epistemological tenet of the challenge from moral epistemology: it is simply not true that all sentences which cannot be conclusively verified are either false or meaningless. Consequently, the argument from moral epistemology does not give compelling reason to reject moral absolutism.

#### COMPREHENSION QUESTIONS.

1. What does it mean to say that an argument is valid?
2. What does it mean to say that an argument is invalid?
3. How can we tell that the argument from diversity is invalid?
4. What options does the argument from moral ontology fail to consider?
5. What does it mean to 'reduce' moral facts to non-moral facts?
6. Why think that some theories must be fundamental in the sense that there are no facts that 'make' the theory true?
7. What is wrong with the idea that a sentence is only true if there is some method that could in principle conclusively verify it?

#### STUDY QUESTIONS.

1. Find an argument in some op-ed or letter to the editor, and assess its validity.
2. Find a sentence in science that expresses a fact that can be reduced to facts in a different domain of science.
3. Find a claim in some science which you think is fundamental, i.e. not reducible and unlikely to be reduced by future theories.
4. Find a claim that most would assent to that cannot be conclusively verified by any method.

## Theories of the Moral World: Reasons for Absolutist Theories.

Although no conclusive argument for absolutism can be given, there are arguments in its favor. We will discuss two of them. The first aims to give us reasons for thinking an absolutist theory might be possible, the second aims to give us reasons for preferring absolutist theories to relativist theories.

The first argument we will call the argument from *Shared Moral Commitments*. The idea here is that behind the initial disagreement amongst cultures about moral truths lurks widespread agreement about quite general moral principles and sources of moral value. The differences which are initially so apparent turn out to be no more than exceptions to moral rules that are generally endorsed.

So, for example, some cultures have in the past permitted infanticide, while many other cultures, ours included, have not. Initially, that looks like a serious and fundamental difference in the moral views of different cultures. But very few cultures have allowed that infanticide at parental whim is morally permissible—even those cultures that have legally allowed infanticide have regarded infanticide as *morally justified* only in special circumstances. In particular, infanticide is regarded as permissible when, and only when, caring for the extra infant would endanger the family or family-group as a whole. So, Eskimo culture permitted infanticide when caring for the extra infant would be likely to impair the ability of the group to feed everyone. That is, all cultures accept two general moral principles: Do not kill innocent persons, and Care for infants and children. Also, all cultures recognize exceptions to these general rules. But cultures differ in the exceptions they recognize, because they differ in the environmental circumstances they face. Eskimo families routinely faced situations in which not everyone in the family or group could be fed, and this forced unhappy choices about how to distribute the limited food supply. Not surprisingly, the culture recognized exceptions to the two general rules which permitted the culture as a whole to survive. We, conversely, do not routinely face such circumstances, and so regard infanticide generally as impermissible. But even we have recognized special exceptions to the general rule—e.g. we have provide palliative care only for certain infants in terminal condition, and even for some infants that are not terminal but severely and untreatably disabled, as in the most severe forms of spina-bifida.

The reasons for which all stable cultures have adopted rules requiring care for children and forbidding the killing of innocent persons, as well as other rules, e.g. rules requiring truth telling, certain kinds of social cooperation, norms of fair play and so on, are fairly straight-forward. Absent such general rules, with only special exceptions, requiring or forbidding such behavior, a society is bound to collapse. If children are not cared for, the next generation will not survive; if killing is not generally seen as immoral, there will be random killing, and a consequent lack of social interaction, and so on.

The absolutist may hope, then, that behind the initial disagreement lies general agreement about some fundamental ethical principles. These principles, and the recognized exceptions to them, can perhaps be systematized in a way that makes good sense to most

people from most cultures, once each takes into account the environmental circumstances faced by other cultures, and takes seriously the moral choices confronting individuals in those circumstances.

We should point out, however, that attractive as this vision is, it faces some serious difficulties. The most obvious concern is how to prioritize different general principles, each of which is widely shared. Consider for example moral injunctions not to kill and moral injunctions to be pious, i.e. respect the gods. For many cultures piety has required human sacrifice, and hence a violation of the injunction not to kill. For them, the injunction to be pious takes precedence over the injunction not to kill. For others, the order of priority is just the reverse. Even if there is shared commitment to general principles, there may be fundamental and ineradicable differences in views about their prioritization.

The second argument for moral absolutism gives us reason to prefer an absolutist ethical theory over alternatives. We will call this the argument from *Reflective Equilibrium*. The basic idea is this. There are certain claims about moral judgment, and certain moral judgments, that you have which are inconsistent with moral relativism. Either your beliefs are false, or moral relativism is false. The argument from reflective equilibrium assumes that when push comes to shove, you are more inclined to hold on to your beliefs than to adopt moral relativism.

There are a number of such beliefs that many, perhaps most, of us hold that are inconsistent with moral relativism. We will give two such here. First, most of us are inclined to think that some cultures engage or have engaged in practices that really are morally reprehensible. Examples include the execution of Jews in Nazi Germany, slavery in the 18<sup>th</sup> and 19<sup>th</sup> centuries and in West Africa in the 20<sup>th</sup> century, and the systematic racial discrimination of Jim Crow or Apartheid. Our judgments about the moral status of these practices, however, are made from the perspective of a quite different culture than that in which the practices occurred. We are not living in Nazi Germany, 18<sup>th</sup> century North America, West Africa, the Jim Crow South, or South Africa circa 1980. Our judgments are informed by moral principles, *norms*, adopted by us, but not by the cultures we are judging. From the perspective of relativism, therefore, these judgments are a mistake—one can judge a practice only with respect to the moral principles that apply to its practitioners, and these are the principles which are *true for their culture*. Since each of the relevant cultures adopts principles that permit the practices, the practices are *not* seriously immoral for members of those cultures. If ethical relativism is wrong, then our moral judgments about these apparently reprehensible practices are simply in error.

Similarly, most will judge that the changes in our own society over time sometimes count as *moral progress*. E.g., our society became more moral when slavery was outlawed, when women were legally able to hold property and to vote, and when Jim Crow fell. But the idea of progress implies *improvement*, i.e. that old practices, now forbidden, were morally *worse* than current practices. But again, any such judgment must be made using a moral theory *which would not have been accepted when the practices in question*



*occurred*. Hence, for the relativist, such judgments about moral progress rest on a fundamental mistake, and so are themselves in error.

Consistency requires therefore that either we reject ethical relativism, or that we reject our judgments about moral progress and our moral evaluations of such intuitively immoral practices as the Holocaust. Consistency, logic, alone does not tell us which we should reject, of course. There is nothing *incoherent* in a serious, consistent relativism. But for many its cost will be prohibitively high. In any case, we have here reason to *prefer* an absolutist moral theory over a relativist theory, provided they are otherwise equal. And that really is the aim of the argument from reflective equilibrium.

#### COMPREHENSION QUESTIONS

1. What does the argument from shared commitments establish?
2. In what way is it possible for two cultures to share commitments about general moral principles, and yet have unreconcilable differences about how to apply those general principles?
3. What does the argument from reflective equilibrium give us?

#### STUDY QUESTIONS

1. What must, in general, a consistent relativist refrain from doing?
2. Find at least one other kind of moral judgment not mentioned above that relativists cannot consistently make, but which many or most of us will be inclined to make at least sometimes.
3. How might differences about prioritization of general principles be resolved?

## The Structure of Ethical Theories

The kinds of moral claims humans make are enormously varied. But much of our moral discourse can be captured by three basic kinds of moral assertion. First, we often make *moral evaluations* of particular persons or particular actions those persons take. We might claim, for example, that the ousting of the Taliban from Afghanistan was morally justified, but the invasion of Iraq was not, or that Clinton was blameworthy for his false testimony, but not for his decision to intervene in Bosnia. Such judgments, or denials of them, concern particular actions and agents, in particular contexts. Any such judgment is of course contentious, in the sense that not all will agree. But if absolutism is correct, then for every action, and every agent who so acts, there is only one complete, true moral evaluation of that agent and that action in those circumstances. It is the job of an absolutist moral theory to offer a principled systematization of all such complete, true moral evaluations.

Second, we often make claims about the moral status of kinds of actions. These claims typically take the form of norms, or principles governing morally correct action. For example, we might claim that killing and lying are wrong, saving lives and charitable contributions obligatory, and so on. Sometimes such *general principles* are claimed to permit exceptions, again specified by kinds of actions: e.g. lying is wrong in general, but white lies are permissible, or killing is wrong in general, but killing in self-defense is permissible. And again, sometimes in recognizing exceptions we do so by implicitly prioritizing general principles, as e.g. ‘lying is wrong unless it is a white lie, or necessary to save innocent lives’ implies that the general principle requiring us to save lives takes precedence over the general principle requiring truth telling. If absolutism is correct, then there is a unique, true, collection of prioritized general principles. Further, by applying these principles to particular actions and particular agents in the circumstances, we ought to be able to recover our ethical evaluations. That is, the set of prioritized general principles in conjunction with the relevant circumstantial facts, ought to entail each and every true moral evaluation.

Third, we make claims about *value*—we say, and seem to think, that certain things, persons, actions, achievements, and institutions among them, are morally valuable. So, e.g., we talk about the value of life and of democratic government. Claims about value come in two kinds. Some things are valuable in and of themselves—human life, for example, or freedom. Such things are said to be *intrinsically valuable*. Other things are valuable for what they produce, for their consequences. On some views, this is true of democratic government—democracy is valuable not in its own right, but because it is the system of government that best conduces to human liberty and effective pursuit of individual interests. According to such views, liberty and individual interests are valuable in their own right, intrinsically valuable, while democracy is valuable only because *it serves those ends*. This sort of value is called *instrumental value*.

If there are true general principles governing moral behavior, then these principles can sometimes conflict, as when we can fulfill an obligation to save lives only by violating an obligation to tell truths. To resolve the conflict, we must prioritize those rules. Such a

prioritization implicitly amounts to an ordering on the importance of various values. If we think, for example, that one ought to lie when by doing so one will save lives, one is implicitly committing to the view that the value of human life is greater than the value of human autonomy. This might be because one thinks that human autonomy is only instrumentally valuable, or because one thinks that, while it is intrinsically valuable, its value is less important than that of life itself. Many have also held the alternative view that life itself is only instrumentally valuable, while autonomy is intrinsically valuable, and accordingly on such views it is held to be impermissible to lie even to save lives.

The upshot of this is that by designating the things which have intrinsic moral value, and prioritizing those values, one can ground a set of general principles and an ordering of them in importance. So complete ethical theories include claims, which we will call *fundamental principles* which specify what sorts of things have intrinsic moral value and/or intrinsic moral disvalue. From such fundamental principles, perhaps using facts about human nature, one can derive general principles and an ordering of them in priority. Such orderings in effect specify the kinds of exceptions which must be recognized by general principles enjoining us not to kill or lie, or enjoining us to act beneficently (i.e. aid others when we can do so by small sacrifices of our own interests). And from such general principles and facts about particular contexts, one can derive evaluations of particular actions or agents. So the full structure of a complete ethical theory begins with some set of fundamental principles, from which are derived a set of general principles and prioritizations of them, and from these one can derive ethical evaluations.

Commonly, we find ourselves unsure about what the correct complete moral theory is, but quite sure that this theory, whatever it is, will include various general principles. But because we are unsure of the true complete ethical theory, we will be unsure of which principles take precedence in particular contexts. Hence, we will be unsure what exceptions ought to be recognized. Much of the work in ethical evaluation then turns on figuring out the best way to reconcile various general principles, all of which seem right but which happen, in certain contexts, to conflict. How one reaches such a reconciliation is the subject of the next module.

Before leaving the subject of the structure of ethical theories, we need to take note of some features of general principles. There are three broad kinds of general principles—those that say a certain kind of action is *impermissible*, those that say a certain kind of action is *morally required*, and those that say a certain kind of action is *morally permissible*. Rules of the first kind generate obligations, or duties, *not to do* certain kinds of things. Rules of the second kind generate obligations, or duties, *to do* certain things, whenever an occasion arises on which it is possible to do them. Rules of the third kind say that there is no moral requirement to perform or refrain from actions of a certain kind. When rules of this kind apply, our choices are morally innocuous (as, for example, in choosing between wonder bread and home hearth bread at the grocery store).

Sometimes the obligations entailed by a general principle can be formulated as obligations to particular persons. For example, a general principle against lying generates

an obligation I have to my wife to honest up with her about my opinion of her new shoes (with which she happens to be delighted, but which I think not to be so attractive). When this is so, a failure to fulfill an obligation is said to *wrong* someone in particular, in this case my lying to my wife would wrong her. Other obligations are not like this. So, a general principle requiring contributions to charity generates an obligation, but not an obligation to anyone in particular. I can fulfill my obligation to be charitable by contributing to Doctors Without Borders, Habitat for Humanity, or my local church. So long as I contribute to any one of them, the others have no claim against me. Hence, if I fail to be charitable at all, there is no one in particular I have wronged. The church is not harmed by my failure to be charitable, since I could have fulfilled that obligation by contributing to Habitat for Humanity. The same goes for Doctors Without Borders, and so also for Habitat for Humanity. Since no one in particular is harmed, no one in particular is wronged. Nonetheless, *I* have done something wrong. The failure to be charitable is, as it were, a victimless crime. When obligations are such that the failure to fulfill them wrongs particular persons, the obligations are said to generate *perfect* duties. When obligations are such that the failure to fulfill them wrongs no one in particular, they are said to *imperfect* duties.

We should also consider the strength of the obligations entailed by general principles. Sometimes the obligations are undefeatable, in the sense that, in every context these obligations must be satisfied, no matter what the cost to us or others. Such obligations are typically, though not always, perfect duties. When such obligations are perfect duties, they correspond to *rights*. So, for example, if I lie to my wife, I am not just violating a perfect duty to her, I am violating a right she has, against me, to my honest opinion. Other obligations, again typically but not necessarily imperfect duties, do not correspond to rights. By obligation to be charitable does not correspond to any right had by Habitat for Humanity against me to my charitable contribution. For example, though I have an obligation to be charitable, and am currently flush with dollars, I might not be obligated to contribute to any charity if I required all of my current funds to achieve some life-long aim, say climbing Mount Everest, or in order to fulfill some other obligation, say helping my children through college. In general, then, rights must always be respected, no matter what, so the obligations generated by rights must always be fulfilled, no matter what, if this is possible. But duties of beneficence, for example, need not always be fulfilled—if the cost to me is sufficiently large I need not act beneficently, and if I do act beneficently (say by risking my life to save a child from a burning building), my actions are above and beyond the call of duty. Such actions are said to be *supererogatory*.

#### COMPREHENSION QUESTIONS

1. What is a fundamental principle?
2. What is the difference between intrinsic and instrumental value?
3. What is a general principle?
4. Why do general principles have to be ordered in priority?
5. What is an ethical evaluation?
6. Distinguish between perfect and imperfect duties.
7. What is a duty of beneficence?

8. What is the difference between duties of beneficence and duties generated by rights?
9. What does it mean to say that an action is *supererogatory*?

#### STUDY QUESTIONS

1. Name something you think has intrinsic moral value, and explain why you think this.
2. Name something you think has instrumental moral value, and explain why you think this.
3. Name something you think has *both* intrinsic and instrumental moral value, and explain why you think this.
4. Formulate two general principles. Say whether they generate obligations to refrain from acting in certain ways, obligations to act in certain ways, or instead say that certain kinds of actions are permissible.
5. Formulate two general principles that generate obligations. Are the obligations perfect or imperfect? Do they correspond to rights, or not?
6. Specify two kinds of actions you think are supererogatory, and say why.

## Testing Ethical Theories

Just as frequently as people engage in ethical reasoning, formulate moral claims, and act on their evaluations, people disagree with one another about the correctness of such reasoning, the truth of such claims, and justifiability of their actions. Given the wide differences in opinion about what is and is not ethically permissible behavior, how are we to decide the right of it in any given situation?

For those who endorse the idea of a systematic ethical theory—the idea that there is one, universally correct, set of standards by which to judge the morality of behavior—there are a number of strategies to employ. The first step is to note that any ethical evaluation of an action, a person, a proposed law or policy must be *principled*. So there must be a general principle applying to the case at hand. Further, this principle, being part of a universal, systematic theory, must always hold. That is, while there may be contexts in which it does not apply, anywhere it does apply it must apply in full force. Finally, it is not an arbitrary matter where such rules apply and where they don't. If a rule is held to apply in one context but not in another, then those contexts must differ in certain respects, and those differences must be morally relevant—there must be a complete ethical theory which explains *why* the principle applies in one context and not in the other by showing that the principle serves the *intrinsic values* recognized by the theory in the one context, but does not do so in the other. This is just what it means to insist on a *systematic* ethical theory. Principles cannot be applied willy-nilly, but instead apply, where and when they do, because behavior in conformity with them conduces to intrinsic moral value in just those contexts.

This gives us two procedures for testing an ethical theory. Both procedures work by eliminating alternatives. As in science, progress in ethics mostly consists not in finding the truth, but in ruling out falsehood. When assessing a moral theory, or any other moral claim, the first test is to ask whether the theory or claim is self-consistent. If a theory is logically contradictory, then it must be in error, whether the theory is a mathematical theory, a scientific theory, or a theory of morality. Ethical theories will be inconsistent if they recognize two or more distinct general principles, these principles are not prioritized, and it is possible for them to generate conflicting recommendations in particular circumstances. So for example, if an ethical theory requires that we respect our dead, and also requires that we obey the laws of our country, but does not prioritize these two principles, the theory *must* be wrong. To see this, consider the ethical dilemma facing Antigone. She must, by the first principle, bury her dead brother. She must, by the second principle, refrain from this, because her country (her city) has ruled that her brother must remain unburied. According to the moral theory, Antigone must do what cannot be done—both bury and not bury her brother. Situations like Antigone's, where general principles, taken to be equally compelling, conflict are called *moral tragedies*, and we will have more to say of them later. When a theory generates a moral tragedy, we can know that it is mistaken.

A second way to test a moral theory uses *reflective equilibrium*. Recollect the reflective equilibrium arguments for absolutist moral theories. Such arguments work by noting a

conflict between our intuitive evaluations of particular cases and the general principles we intuitively endorse. We think, for example, slavery is wrong not just for us, but for everyone. But if we endorse ethical relativism, this claim is mistaken. So one or the other must be false—either the theory, here ethical relativism, must be wrong, or our intuitive evaluation of the practice of slavery must be wrong. Once such conflicts have been noted, we have two options. We can change our intuitive evaluation, or we can change the principles we endorse. The first option will be appealing to the extent that we can explain why our intuition in the particular case is mistaken. The second will be appealing to the extent that we can see a way to reformulate our principles in ways that preserve our judgments about what is intrinsically valuable. By carefully looking for and resolving such conflicts, we can hope to reach a final state in which our intuitions about particular cases are fully consistent with the principles we endorse. When we have reached such a state we are said to be in reflective equilibrium—or considered, reflective, judgments about particular cases accord with our considered, reflective, judgments about general principles and intrinsic values.

The good thing about reflective equilibrium arguments is that they can be applied to partial theories. I do not need a full ethical theory, or anything like it, to work out principles governing just killing for most contexts of interest. I might begin by noting that there is a general rule against killing ('Don't kill human beings'), while remaining unclear whether human life is itself an intrinsic or instrumental value; instead I hold that, for one reason or the other, I know not which, the correct ethical theory will include an obligation not to kill. I then note that this duty will be a perfect duty, since if I violate it I wrong my victim. Moreover, it is a perfect duty that corresponds to a right, namely the right to life. I can then ask, do all killings violate a right to life, or are there some killings that don't, and hence are permissible? Perhaps I think that killing in self-defense is permissible, and that capital punishment is also permissible. Those intuitions conflict with my principle above, so something has to change. I could resolve the conflict by changing my principle: 'Don't kill *innocent* human beings'. In effect, I have said that certain human beings do not, after all, have rights to life, and hence killing them does not violate a right to life. And that might accord pretty well with my views about intrinsic value—I might think that, after all, whatever it is that makes human life valuable in general is missing in the lives of non-innocents.

So far, so good, but perhaps I also thinking killing in a just war is permissible, even when those killed are innocent victims. Then my principle will need further modification. I might then endorse the principle 'Don't kill innocent human beings, except when necessary for the prosecution of a just war.' Again we have saved our intuitions, but this reformulation of our general principle also requires that we change our view about rights to life. The prior change in principle required that we change our view about who has a right to life. But the current change in principle requires that we change our view about *what obligations a right to life entails*. In essence, we take a right to life to be a right against being unjustly killed. On the principle 'Don't kill innocent persons' it might be that all killings of beings with rights to life are unjust, but this cannot be the case for our new principle 'Don't kill innocent persons unless necessary for the prosecution of a just war'. The innocents surely do have a right to life, it is just that their deaths as a result of

'collateral damage' do not violate their rights to life. By further consideration of cases, values and principles, I might hope to find a general principle about killing, and an associated conception of the right to life, which accord with those cases of killing I am inclined to regard as permissible. If so, I will have reached a local reflective equilibrium—local because it concerns only the moral issues surrounding killing.

The search for reflective equilibrium, however, does not necessarily decide between alternative theories. This is because when a conflict is discovered, we always have a choice: modify the principles we endorse, or modify our intuitions. There is nothing incoherent, for example, in deciding to reject the above intuitions about just killing, as might a Quaker, for example. We might just decide to stick with the principle 'Don't kill human beings', come what may. So if the most basic intuitions about intrinsic value differ amongst some group of persons, it may be that they can never, by searching for the reflective equilibrium, reach a resolution of their differences regarding general principles or specific evaluations. But in the search they will, at least, become very clear about *why* they differ.

There are a number of other constraints on moral theory. One is this—the kind of ethical theory that matters is an ethical theory for us, for human beings. If a theory demands that we do what *cannot* be done by humans, then it is a bad moral theory. Its recommendations cannot guide our actions, since it will recommend that we do things we are simply powerless to do. Another constraint is that we must somehow manage to live together in societies which, so far as possible, allow us to pursue our own ends, provided those ends are themselves morally permissible. If a moral theory requires that people behave in ways which would make social interactions *impossible* to sustain, then that is good reason for rejecting the theory. We could follow the recommendations of the theory, it is true, but in so doing we would make life itself impossible, and that cannot be right.

Whether in science or in ethics, data are said to *underdetermine* the truth if more than one theory is consistent with the data. In general, underdetermination cannot be avoided, either in science or in ethics, without making substantive background assumptions about, respectively, the physical and the moral world. Such assumptions are not objectionable, so long as they are not currently being challenged. So if all sides to a dispute in ethics agree that certain assumptions are true, it is legitimate to treat them as if they *are* true, even if one has no independent argument for them. But when those assumptions *are* disputed, they must be defended. In ethics, such defense will nearly always employ arguments of the above kinds, using *other assumptions* that *are* agreed to by all sides of the dispute. Hence, when thinking through moral issues in which different parties disagree, it is important to ask what moral claims will be taken to be true by everyone in the dispute. Those claims for a common set of background assumptions, and will be essential for reaching a local reflective equilibrium.

#### COMPREHENSION QUESTIONS.

1. What is *moral tragedy*, and what can we say about a theory if it permits moral tragedy?



2. What is reflective equilibrium?
3. When we find a conflict between principle and intuition, what two choices do we have in resolving the conflict?
4. What further constraints on a correct moral theory are there?
5. When are we entitled to make an assumption when resolving an ethical dilemma, and when are such assumptions *not* legitimate?

#### STUDY QUESTIONS.

1. Construct a new example of a moral tragedy from the perspective of your own personal ethical theory (you might consider here conflicts between or among your obligations to family, to work, your duties of beneficence, conflicts between obligations not to kill and obligations to save lives, or other specific obligations).
2. Consider the rule 'Lying is morally impermissible', where by a lie we mean speaking falsely when one knows or believes that one is speaking falsely. Consider a number of cases in which you think the rule might not apply, i.e. cases in which you think lying might be permissible, and use these cases to construct an alternative general principle about lying. Consider both cases which force a modification of principle, and cases in which principle forces a modification of intuitions, and for each modification, try to say what values, intrinsic or instrumental, are served by that modification.
3. It is often said that 'ought implies can', by which it is meant that people cannot be morally obliged to do what it is impossible for them to do. See if you can construct an argument for why this is plausible.

## Consequentialist Ethical Theories

Ethical theories can be categorized into three broad families: consequentialist theories, deontological theories, and virtue theories. Because virtue theories raise problems about situationalism, and because virtue theories have not been developed as much as other types of theory, we will not discuss them in this course. The remaining two types of theory have been very popular historically, and elements of both are found in our everyday moral reasoning.

The consequentialist family of theories starts with a basic insight: all human actions are performed for the sake of some result or consequence. We brush our teeth in order to prevent cavities, we hug our children because it makes them and us feel good, and we look at art because it enriches our lives. When we are trying to decide what action to perform out of the many choices in front of us, we consider the results of each action. For example, I could eat some leftover meatloaf or go out for fast food tonight. Which shall I do? I have to compare the consequences of each action: one action will be less expensive, healthier, and quicker. The other will be tastier and more fun. As a rational person, I try to choose the action that will have the best consequences over-all. In fact, isn't that the very definition of rational action: doing the action that has the best results? Suppose I hated the taste of fast food, and disliked being in fast food restaurants. Then I would have no *reason* to choose the fast food dinner, and it would be irrational for me to choose that option.

Well, if all rational action is aimed at the best consequences, then can't we simply define *moral* action as action that is aimed at the best *moral* consequences? This is the basic consequentialist approach: figure out what the morally best states of affairs are, and then define right action as whatever actions produce those states of affairs. Another way to put the point is this: consequentialists first figure out what things are *good* or valuable, and then they define right action as whatever action maximizes the good. This makes consequentialist theories very straightforward because there is only one simple moral rule to follow: always act so that you maximize the good in the world.

But what is good in the world? What are the morally best consequences? Consequentialists have disagreed over the answers to these questions, and that has led to two main types of consequentialist theory.

Utilitarianism: "The Good" is pleasure and the absence of pain. The morally right action is whichever one maximizes the total amount of pleasure in the world.

The argument: a good case can be made that pleasure is the only thing humans value for its own sake. We may also value education, loving relationships, art, power, health, and wealth, but isn't our ultimate reason for valuing these things that they make our lives more pleasurable? If all works of art were ugly and made us feel sick to our stomachs, we wouldn't value art, right?

Since each person values pleasure (and the avoidance of pain), then the most valuable thing for everyone is the greatest total amount of pleasure.

Since moral action is supposed to aim at the over-all most valuable state of affairs, moral action should aim at maximizing pleasure (and minimizing pain).

Preference-Satisfaction Consequentialism: “The Good” is preference-satisfaction. The morally right action is whichever one maximizes over-all preference-satisfaction.

The argument: almost as soon as utilitarianism was formally developed (by Jeremy Bentham and John Stuart Mill in the mid- 1800s), critics attacked its definition of “The Good.” It’s not difficult to come up with examples of presumably rational people who value other things more than they value pleasure. If we offered people a choice between a very happy, long life as a severely mentally handicapped person and a moderately happy, shorter life as a person of average mental ability, most people would choose the latter. This shows that we prefer certain kinds of abilities to sheer pleasure. And what about masochists, who seem to pursue pain for its own sake?

Rather than try to make the argument that every rational person really values pleasure above all else, why not just define “The Good” as the fulfillment of whatever desires people happen to have? Some prefer pleasure, some prefer pain, some prefer knowledge, and so forth. The morally right action is whatever maximizes the satisfaction of all those preferences, whatever they may be. This has become the most popular form of consequentialism.

Both forms of consequentialism ascribe one basic moral right to all human beings: the right to have their pleasure or satisfaction taken into consideration equally with everyone else’s. A *right* is a legitimate claim that a person has on other people. When one person has a right, other people have corresponding obligations to that person. For example, most people believe that they have a right to control their own property. This right generates corresponding obligations in other people: everyone is obligated not to steal my property or damage it for no reason. Similarly, a right to life generates an obligation for others not to kill me except in special circumstances. On the other hand, it doesn’t necessarily generate an obligation for other people to preserve my life at all costs; nobody is obligated to give me one of their kidneys, even if I need it to survive. The interesting thing about consequentialism is that it reduces all of these rights to one fundamental right: the right to have our interests considered equally with the interests of others. We don’t have a fundamental right to life, only to have our interests taken into account. It will almost always be the case that the best state of affairs will be obtained by not killing innocent people, but there may be exceptional circumstances where others’ interests will outweigh an innocent person’s interest in staying alive. As long as we count that innocent person’s interests fairly in our calculation, we have met our obligation to that person.

We use consequentialist reasoning all the time. In fact, it's hard to imagine how we'd get along without it. Suppose a building is on fire, and you can either go down one hallway and rescue the three children trapped down there, or you can go down the other hallway and rescue one child trapped down there, but you don't have time to do both. Shouldn't you try to save the three children? Of course it's bad to let the one child die, but it would be even worse to let the other three children die. When we make the decision to let the one child die, we haven't violated any of his/her rights because we have considered his/her interests equally with the other three children's' interests; it just happened in this case that the interests of the other three outweighed the interests of the one.

To take a more mundane example: suppose you have one coconut cream pie and four children sitting around the table. How should you distribute the pie? At first glance, the most *fair* option seems to be to divide the pie evenly between all four kids. Don't they each have a right to their share of the pie? But suppose one of the kids isn't really hungry and doesn't like coconut anyway; then isn't the best option to divide the pie three ways between the other three kids? The best rule to follow in this case is the simplest: do whatever will maximize the total amount of pleasure or preference-satisfaction. This respects the rights of all four kids because we've taken account of all their interests.

Despite the fact that there is much to recommend consequentialist theories, not least the frequency with which we employ consequentialist considerations in everyday moral reason, it does face some serious difficulties. As already noted, among them is that consequentialist theories seem to imply that we have far fewer rights than we are normally inclined to think we have. Various versions of consequentialist theories have been proposed which avoid this problem (though it is a matter of debate whether they resolve the problem correctly). But other problems are more difficult. Perhaps the most important of these is this. In order to do the right thing, according to consequentialist theories, one has to know what the consequences will be of each possible course of action. Very often, we are in no position to know any such thing. Will reducing the amount of televised violence available to children also cause those children to become less violent adults, or will such restrictions merely make unhappy those who happen to enjoy violent programs? Will restricting power-plant mercury pollution really make the public healthier and smarter, or will it just make them poorer because they must pay more for electricity? Will a policy of engagement or a policy of aggression be most effective in convincing North Korea to stop building nuclear weapons? We can only take our best guess, and act on it. If we choose wrongly, then we will have done something morally incorrect from the consequentialist perspective—even if we gave the decision our best efforts in good faith. But that makes it look as if consequentialism demands of people that they do what they cannot reliably do, namely predict the future. As a general rule, you are not blameworthy for failing to do what you could not possibly have done. It looks like consequentialism might violate that rule

So there are many reasons to endorse consequentialist theories, but also reasons to be unconvinced that they are true. As a matter of fact about human psychology, most of us *sometimes* engage in consequentialist moral reasoning, for *some* kinds of moral problems, but also sometimes *reject* consequentialist reasoning as mistaken for other kinds of moral

problems. When we do so, we generally appeal to a different kind of moral theory, called *deontic* moral theory.

#### COMPREHENSION QUESTIONS.

1. What is consequentialism?
2. How does Utilitarianism differ from Preference-Satisfaction Consequentialism?
3. What is the argument for Utilitarianism?
4. What is a right?
5. What is the relation between rights and obligations?
6. Describe two problems a Utilitarian theory must solve?

#### STUDY QUESTIONS.

1. Is the argument for Utilitarianism valid? Why or why not?
2. Can you think of a case where you believe it would be morally *wrong* to maximize over-all satisfaction?
3. Do you think humans have other rights besides just the right to have our interests taken into account? Why or why not?

## Applications of Consequentialism

In this section we will consider some simple applications of consequentialism.

Case 1. Igor, Boris and Svetlana are physicists engaged in a high energy physics experiment. Their job is to analyze the performance of the particle detectors employed in the experiment. Igor and Boris on the one hand, and Svetlana on the other, disagree about the appropriate sensitivity to attribute to the detectors. Both assessments find that the detectors are sensitive enough to perform the experiment, but the overall significance of any finding will depend on which sensitivity is employed in further calculations. However, all three agree that if Irina, the PI for the experiment, learns that Igor, Boris and Svetlana disagree in their assessments, she will halt the experiment until the disagreement is settled. Assuming preference-satisfaction consequentialism, in their report to Irina, should the detector team go with the majority view, or should they report both assessments?

Analysis of Case 1. According to preference-satisfaction consequentialism, the detector team ought to act so as to maximize desire satisfaction. The desires that matter are those of the individual detector team members, Igor, Boris and Svetlana, the PI, Irina, and the remaining members of the collaboration, who we will call the *remaining members*. Since the results of the sensitivity analysis cannot affect the actual run of the experiment, the detector team's decision here has only one consequence of interest for the remaining members: if they report dissension to Irina, the preference had by all the remaining members to go forward with the experiment (provided the detectors are working, which the detector team believes them to be) will be frustrated, and these preferences will be satisfied if the majority view is reported. So the interests of the remaining members are best served by reporting only the majority view. So too, the preferences of Igor and Boris are best served by such a report—both their preference for continuing the experiment on time, and their preference for their own analysis of the detector's reliability. But Svetlana has two interests, the timely completion of the experiment, but also her preference for her own assessment of the detector's reliability. The first, but not the second, will be satisfied by a majority report, the second but not the first would be satisfied if dissension is reported to Irina. Irina herself has two preferences—that the experiment be completed in a timely fashion, and also that it not go forward until a reliability assessment has been agreed to by the whole team. Again, the first but not the second is satisfied by a report that includes only the majority view, while the second but not the first is satisfied by a report that includes both the majority view and also Svetlana's minority opinion. On balance, then, reporting the majority view to Irina is the preferred solution. Such a report satisfies preferences for Igor, Boris and the remaining members that otherwise would not be satisfied, and it does so without frustrating other interests they have. Such a report does frustrate an interest had by Irina and Svetlana, but it also satisfies another interest each has, and the alternative offers no advantage in this respect, since it too frustrates one interest and satisfies another had by Irina and Svetlana. A report including only the majority view therefore maximizes preference-satisfaction.

Notice that in applying the consequentialist theory there are a variety of steps. A) Identify each possible action open to the agents who are to act. B) Identify all the people

who would be affected by at least one of the open alternative actions. C) For each affected person (including the agents themselves) and for each alternative action, find the desires had which are frustrated by the action, and those which are satisfied by the action. D) Sum up the effect of each action over each person, and choose that action which maximizes preference satisfaction.

There are sometimes difficulties which do not emerge in the above case. Two of these are important. First, while the desires of each *person* are equally important, not all desires of the *same* person are equally important. For example, it might be that Irina *much* prefers the experiment run to completion than that it run only once the reliability assessment has been agreed to, or conversely. When desires differ in importance *to the person who has them*, each must be proportionally discounted or upgraded in our calculations, and that requires a deft judgment. Second, we sometimes cannot be sure what the consequences will be, we can at best guess about the chances of various outcomes for each open alternative course of action. In this case, judgments about which preferences will be satisfied and which frustrated must be discounted by our estimate of the chance of each outcome. Again, that requires a deft judgment about probabilities. For each kind of problem it is sometimes the case that the formal machinery of *Decision Theory*, *Game Theory* and related theories of rational choice can be used, but these theories are beyond the scope of this module.

Case 2. Jane, Doug and several other graduate students are taking an introductory course in *Drosophila* genetics focusing on laboratory techniques. Doug is an advanced student finishing his dissertation work, and has secured a terrific post-doc position at NIH assessing the efficacy of alternative cancer treatments (his main interest), but Doug needs one more course to fulfill his degree requirements (the birth of his second child in his first year forced him to withdraw from the course then). The term project for the course involves collecting data to find the norm of reaction for bristle number against natal temperature in an inbred strain of fly. Jane has discovered that Doug has manufactured his data rather than actually measuring the bristle numbers on the flies he has reared. She judges that if she says nothing, his fraud will remain undiscovered. The instructor for the course, Professor Immanuel, is a stickler for academic rules—she will certainly flunk Doug, and will try very hard to have him dismissed from the university, if she learns of his fraud. In either case, Doug would certainly lose his post-doc position, and in the latter case would have to give up biology altogether. Jane herself is very angry at Doug's deception, and would be happy to see him punished. Further, she believes that there is a good chance that if Doug is not reprimanded at all, he will continue to engage in fraud in the future, for she has noticed that Doug tends to be somewhat lazy and slipshod with what he regards as unimportant routine matters. Jane judges that she has two options: report Doug's fraud to Dr. Immanuel, or say nothing to Dr. Immanuel. From the perspective of Utilitarianism, what ought Jane do?

Analysis. Many persons are affected by Jane's decision: she herself, Doug, Professor Immanuel, the other students in the class, Doug's wife and children, the faculty and students who know Doug, the people for whom he will be working at NIH, the person who would take his spot at NIH if Doug does not, and the patients whose treatment

depends on the results of the NIH study and their families and friends. Jane herself will clearly be happier if Doug is punished. Doug will clearly be much happier if Jane remains silent. Professor Immanuel is harder to judge. Being ignorant of Doug's fraud, she will not be *unhappy* if Jane remains silent. On the other hand, if Jane reports the fraud, Professor Immanuel will be happy that the fraud is discovered, but unhappy that a promising student has ruined his career. The other students in the class, and the faculty and students who know Doug are in a similar position. Perhaps on balance, Professor Immanuel will be happiest if Jane reports Doug's fraud, while the other students and faculty will be happiest if she doesn't. Note that in saying this we are NOT claiming that, if Jane does report the fraud, anyone of the faculty or students will *wish she had remained silent*; rather we are assuming only that these people will be happier if she remains silent than if she does not.\* Doug's family is clearly *much* happier if Jane remains silent. The person who would replace Doug at NIH is at least slightly happier if Jane reports the fraud, since this particular post-doc is quite prestigious. Doug's future NIH employers and the patients whose treatment will be affected by the NIH study are the most difficult to assess. If Doug goes on to do good work at NIH and engages in no further fraud, then these people are largely unaffected one way or the other by Jane's decision—Doug does good work, but so would his replacement have done. But if Doug engages in *further* fraud, then matters are quite different: as a result of Doug's further fraud, it is quite possible that people will die because they receive an ineffective treatment when a better treatment is available. And others, who would have been diagnosed as untreatable, will receive a painful and otherwise useless treatment during the last months of their lives. Occurrences of both kinds will negatively affect the friends and families of these patients. On balance then, we have the following. Neglecting the possibility of future fraud, remaining silent is preferred—outing Doug will make Jane, Professor Immanuel and Doug's replacement somewhat happier, but at the price of making Doug and his family *much* unhappier, and making many faculty and students marginally unhappier. But if future fraud occurs, very many people will be made very, very much less happy than they otherwise would. So Jane's decision should depend on how likely she thinks it is that Doug will engage in fraud in the future if unpunished now. Fraud, if it occurs, will engender much more unhappiness than that experienced by Doug and his family if Jane reports Doug. So the chance of fraud need not be very high at all for Jane to be obliged to report Doug. So if Jane thinks future fraud is at all likely, she should report Doug.

#### STUDY QUESTIONS:

1. Dr. Tierlieber has just published data that offer strong, but not conclusive, support to the hypothesis that the increase in mean temperature during this decade has played a significant role in producing the increased frequency and severity of red-tides, but the data fall far short of showing that global warming will have catastrophic ecological consequences, and do not bear at all on the question of whether or not some anthropogenic signal can be found in the data on increases in

---

\* This is an important difference between preference-satisfaction and utilitarian versions of consequentialism: for utilitarians, it is happiness that matters, even if the persons in question would *prefer* a course of action which makes them less happy, as is the case with the friends of Doug, who will be happier if Jane does not report, but would nonetheless prefer she report Doug if they knew of the situation.



mean temperature. The popular press, however, is currently misinterpreting Dr. Tierlieber's data as supporting these claims. Dr. Tierlieber believes that current evidence for the claims that global warming is occurring, that it is being caused, in part, by human actions, and that it will have catastrophic ecological consequences, is strong, but not conclusive. He believes that it is a matter of good policy to take action now to limit emissions of carbon and methane, but that such policy can only be established with stronger public awareness of global warming, an awareness that is being fostered by the press's misinterpretation of his data. Assess whether Dr. Tierlieber should take steps to correct the press's misinterpretation of his data assuming preference-satisfaction consequentialism.

2. The University of Arizona's Astronomy Department has formed a consortium with 4 other institutions, including the Vatican and the Smithsonian, to build a series of new telescopes on the Pinaleno Mountains in southeastern Arizona. The Pinaleno range was chosen for a number of reasons. They are far enough from Tucson and Phoenix not to be polluted by light from those cities, yet are close enough that that the building and staffing of the observatory is economically feasible. Further, uniquely among mountain ranges so situated, the Pinalenos, at 10,000 feet in elevation, are high enough that atmospheric distortion of starlight is minimized. The chosen location provides the habitat for a sub-specific population of red squirrel that has been isolated for 10,000 years, since the end of the Pleistocene glaciation. Because the population size varies, depending on the size of the conifer cone crop, between 100 and 500 animals, the population is endangered by inbreeding and drift, both of which, given the small population size, can swamp the effect of selection. The proposed observatory will destroy nearly a quarter of the remaining prime habitat for this population of squirrel. From a utilitarian perspective, should the US Forest Service grant permission for the construction of the telescopes? Were the proposed project a ski resort rather than an observatory, would your answer be any different? If so, why?

## Deontological Ethical Theories

One of the problems with consequentialism is that it seems to have *counter-intuitive* results in some cases. A theory has counter-intuitive results when the theory implies something that is in conflict with our everyday intuitions. When this happens, we need to use the process of reflective equilibrium to decide whether to alter the theory or our intuitions.

One of the counter-intuitive results of consequentialism is that it implies that there is no type of action that is always, absolutely wrong. Any action can be the right action if it produces the best consequences. For example, consequentialism says that in principle it would be right to kill an innocent person if that would prevent a greater harm from occurring. It might even be right to kill one person and harvest his/her organs to save the lives of five other people needing organ transplants.

This clearly goes against our ordinary intuitions. We usually think that there are certain kinds of actions that are always wrong and that should never be performed, even if it would create greater happiness in the world. We prohibit torture, even if it would be the most effective way to get information, and we prohibit nonconsensual medical experimentation, even if it would lead to quicker cures for terrible diseases. These are rules we think we should follow regardless of the consequences.

The idea that there are certain moral rules that trump considerations of happiness lies at the base of deontological theories. ‘Deontos’ is a Greek word meaning “that which is binding” or “duty.” Deontological theories start with the insight that morality seems to have a special status for humans: we think of moral requirements as binding us, as requiring that we do or refrain from doing certain things no matter what. This makes morality very different from other normative systems. For example, etiquette is normative and gives us rules: you “should not” put your elbows on the table while eating, for example. But we think of this rule as contingent: it’s a good rule to follow if you want to be polite within a certain culture, but if you don’t care about being a polite person then there’s no real reason to follow the rule. Similarly, art criticism is normative, and an art critic could give us many good reasons for thinking a Rembrandt painting is much more beautiful and valuable than a velvet painting of Elvis. But if you prefer velvet paintings you’re not obligated to change your preferences.

Morality is different. When morality tells us that we should not kill innocent people except in self-defense, morality is not telling us it would be “nice” or “polite” or “in good taste” to refrain from killing people. It’s telling us we are bound and obligated not to kill innocent people (except in self-defense), no matter what. Even if we wake up in a bad mood one day and really feel like killing just one annoying person, we’re still obligated not to do so. Even if the innocent person in question is annoying everyone, and the world would be better off without him/her, we’re still obligated not to kill that person. Even if we could use the person’s organs to save five other people’s lives, we’re still not allowed to kill him/her.

We saw earlier that obligations correspond to rights. When someone has a right to something, that corresponds to obligations that other people have toward that person. So another way to think about deontological theories is to think of them as theories that ascribe a variety of rights to us, not only the right to have our interests taken into account. Most deontologists think we have a right to life, which obligates others not to kill us, even if killing us would create more happiness in the world. These rights place limits on the extent to which a moral person can pursue over-all happiness or satisfaction. A deontologist can act to maximize over-all happiness, but not if it violates one of her duties.

So deontologists start with the idea of duty and try to figure out what our most basic moral duties are. Some deontologists try to find one unifying principle that underlies and explains all our moral duties; we'll refer to these theorists as *monistic* deontologists. Others settle on a list of various duties; we'll call these theorists *pluralistic* deontologists. We can think of the Ten Commandments in the Old Testament as a very early deontic list of duties. Professional codes of ethics can also be seen as deontic lists.

The most difficult theoretical problem for deontologists is making an argument that we really do have certain moral obligations. They have to show that there are certain rules or laws that are binding on us regardless of how we feel about them, what the consequences will be, etc. Where do these laws come from? How do we know them? Why do they have authority over us? All deontologists have to answer these questions. Furthermore, monistic deontologists have to argue that their one fundamental principle does really explain all our other moral duties. On the other hand, pluralistic deontologists have to tell us how their various duties relate to each other. What should we do if two duties conflict with each other? Are some duties more important than others?

One thing deontologists have noticed is that there seem to be different classes of moral duties. There are very strict duties that forbid us from ever doing certain things: duties not to murder, steal, etc. Then there are other duties that have more leeway: for example, we may have a duty to help others, but usually we get to decide when, how much, and whom to help. Traditionally, the strict duties in the first group have been called *perfect duties*, and the less strict duties in the second group have been called *imperfect duties*. Most deontologists think of the perfect duties as trumping the imperfect duties. If someone is drowning, you have a duty to help him or her, but if you would have to kill someone else in order to reach the drowning person, then you're not allowed to do that. This solves some of the problems deontologists have explaining what we should do when we seem to have conflicting duties.

However, arguing for perfect duties raises a new set of problems for the deontologist: what if the consequences of doing your duty would be catastrophic? Most deontologists think of us as having a perfect, strict duty not to kill innocent people. But what if an evil alien came to Earth and told us that unless we killed one innocent person (with a painless lethal injection), the alien would immediately destroy 10 million people? How could it be rational to stick to the rule in this case? A similar thought experiment can be devised for any perfect duty the deontologist proposes (don't torture babies, don't steal, don't

experiment on unwilling humans, etc.) This supports the idea that in the end the consequences do trump the rules.

As we saw in our earlier discussion of moral tragedy, sometimes our intuitions about what to do in particular cases seem to be irresolvably conflicted. We endorse two general principles, both of which seem to apply, but which yield opposite recommendations about the morally correct choice. For example, what should one do when one's obligation to be pious conflicts with one's obligation not to kill, as in the story of Abraham and Isaac? Sometimes moral tragedy involves conflicts between two deontic principles. So, for example, we have a deontic obligation to help catch criminals. We also have a deontic obligation to help our family. These can conflict when a family member is a criminal—if your brother seeks your aid in escaping justice, ought you to provide that aid? At least as commonly, however, moral tragedy involves a conflict between consequentialist and deontic obligations. It is clear that one ought to choose to save three innocent children from a fire rather than choose to save just one, when all four cannot be saved. But what if the one is *your* child? Or what if, to save the three, you must actively kill the one? Very often, arguments for or against deontic as opposed to consequentialist theories turn on an analysis of cases of moral tragedy.

#### COMPREHENSION QUESTIONS.

1. What are some of the counter-intuitive implications of consequentialism?
2. How are moral requirements different from other normative requirements?
3. How are deontological theories different from consequentialist theories?
4. What's the difference between monistic and pluralistic deontology?
5. What's the difference between perfect and imperfect duties?

#### STUDY QUESTIONS.

1. Name some counter-intuitive implications of deontological theories.
2. Describe a time when you believe you had conflicting moral duties. How did you decide between them?
3. Are there certain actions that you would never be willing to take, no matter what? Why or why not?

## Deontological Theories: Kantianism, I

The most famous monistic deontologist is the German philosopher, Immanuel Kant (1724-1804). Kant argues that there is one fundamental moral principle that generates all of our moral obligations. That principle is called the Categorical Imperative: Act only on that maxim that you can at the same time will to be a universal law. In other words, you should only act on principles that you could will everyone to follow. This is an imperative, meaning a command that we must follow. And it's categorical, meaning that it applies to everyone, at all times, without exception.

To make use of the Categorical Imperative, what you need to do is test your maxim or principle of action. A maxim states what you are going to do, and what your purpose is. For example, I could form this maxim: "In order to get some money that I need, I will ask for a loan from this wealthy person and falsely promise to pay the money back in the future." This would be a maxim of lying to someone in order to get something I want. Now imagine that this maxim became a universal law: whenever anyone needed some money, he/she would make a false promise in order to borrow the money from someone else, but would never intend to pay it back. Could I will this as a universal law? Actually, a world organized by this universal law is *inconceivable*. If everyone made false promises to get money, then the very words "I promise" would lose their meaning. It would be impossible to make promises because saying the words "I promise" wouldn't mean anything to the other person. So nobody could make false promises to get money, because there would be no such thing as promises. Turning my maxim into a universal law would make it self-destruct, and so the Categorical Imperative says that my maxim is immoral and must not be acted on.

The basic idea here is very compelling: morality requires that we not make exceptions for ourselves or act in ways in which we couldn't will other people to act. Kant thinks his Categorical Imperative captures the intuition behind the age-old Golden Rule, but in a more precise form. He also thinks the Categorical Imperative explains all our other moral duties. For example, murder is wrong because you could never will a maxim of killing innocent others as a universal law. A universal law of lying would self-destruct like the false promising case above, so it's wrong to tell lies.

Kant even has an ingenious way of explaining why some duties are perfect and some are imperfect. Consider the maxim a selfish person might have: I'll never help anyone else unless it directly benefits me. Can this be willed as a universal law? Unlike the false promising case, the world where everyone follows this selfish maxim is *conceivable*. It would be a very different world from the one we live in, but there's nothing conceptually impossible about it. But could we *will* such a world? No, because the point of our original maxim was to indulge ourselves and have everything we want without the bother of helping others. However, if we willed that as a universal law, we'd be willing a world where nobody helps us when we need help, and that conflicts with our original purpose of indulging ourselves.

Kant says that when there's a maxim that cannot even be conceived as a universal law, we have a strict and perfect duty not to act on that maxim. When there's a maxim that can be conceived as a universal law, but that we cannot will to be a universal law, then we have an imperfect duty not to act on that maxim. That means we do have a duty to help other people, but there's some leeway built into that duty. We don't have to go around helping others 24 hours a day, but we do have to go around not murdering others 24 hours a day. The Categorical Imperative explains both kinds of duty.

The crucial contrast between Kantianism and consequentialism is that for Kant the only important thing about your actions is the principle that you act on, not the consequences of the action itself. Suppose you come across a drowning man, and you pull him out of the water and resuscitate him. As it turns out, this man is actually a serial killer and he goes on to kill 17 people before he is caught. Did you do the right thing? A consequentialist would have to say that your action was wrong because it did not maximize over-all happiness in the world. But Kant would look at your principle of action. A maxim of walking by and not trying to help a drowning person could never be willed as a universal law, so you had a moral duty not to act on that maxim and to try to help this person. The consequences ended up being bad, but your action was good in itself and you did your duty.

Kant's theory has been very influential in ethics, but there are some serious problems with it. One of the most important problems is that it's very difficult to apply the Categorical Imperative to actual cases. The Categorical Imperative seems to allow certain kinds of actions that we intuitively think are wrong, and to forbid other actions that we intuitively think are permissible. Here are some examples:

- I could form this maxim: "I will kill John A. Smith on January 10, 2007 because the world would be better off without him." Is this very specific maxim universalizable? Well, if everyone tried to kill John Smith on that date there would be a big mess: we'd be getting in each other's way and so forth. But there's nothing *inconceivable* about such a world. Could I will such a world? Sure, because all I want is for John Smith to be dead. So this maxim seems to pass the Categorical Imperative test, yet it's clearly an immoral maxim.
- On the other hand, what about this maxim: "I will work out at the gym at 5 a.m. every day because hardly anybody else is there and I want a quiet workout." Is this maxim universalizable? Well, if everyone went to the gym at 5 a.m. in order to avoid the crowds, there would be huge crowds there at 5 a.m. So this maxim is self-defeating when it's universalized, and we can't will it as a universal law. The maxim only works because most people don't act on it. So the maxim doesn't seem to pass the Categorical Imperative test, yet it intuitively seems like a perfectly permissible maxim.

Kantians have worked very hard on this problem, but there has been no final resolution of it.

### COMPREHENSION QUESTIONS.

1. What is the Categorical Imperative?
2. How does the Categorical Imperative show that making false promises is wrong?
3. What is the most serious problem with applying the Categorical Imperative?

### STUDY QUESTIONS.

1. Think of some specific examples that we haven't discussed yet where Consequentialism and Kantianism would give different advice on how to act.
2. Which do you think is most important in determining the morality of an action: the consequences of the action or the motive behind the action? Defend your view.

## Deontological Theories: Kantianism, II

So far we have only looked at Kant's first formulation of the Categorical Imperative. Kant went on to provide several other formulations of it, which he claimed were equivalent to each other in meaning, though stated in different words. The formulation we've already looked at is called the Formula of Universal Law: *Act only on that maxim that you can at the same time will to be a universal law*. The most important alternative formulation Kant gives is called the Formula of Humanity: *Act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means*. Many people find this second formula to be easier to apply and more intuitively plausible, though the concepts involved in it ("humanity," "means," and "end") are much trickier.

The Formula of Humanity tells us that we must never treat humanity solely as a means. What does Kant mean by "humanity?" He is referring to our *rational nature*, which is our capacity to guide our own actions on the basis of reasons. This capacity is what sets us apart from other animals. Other animals are guided by instinct and can't freely choose how to act; humans also have instincts, but we are able to stand back from our instinctual urges and desires and make a free choice of how to act. This freedom is known as autonomy, and it is rationality that makes autonomy possible for humans. Autonomy in turn makes us morally responsible. If a pit bull attacks and kills a child, we may put the dog to death for safety's sake, but we don't *blame* the dog or hold it morally accountable for its action. We don't think the dog *chose* to kill the child; it had to act the way it did because animals simply act on whatever desire is strongest at the moment. But if an adult human attacks and kills a child, and the person was in his/her right mind at the time, then we do hold him/her responsible; even if the person was very angry at the child, we think of the person as having the ability to *choose* not to act on that emotion. In this case, the person chose to do something that was clearly impermissible.

Kant thinks of our rational nature as the most important and valuable thing about us. Although he uses the word 'humanity' to refer to this capacity, there's no essential link between having rational nature and being biologically human. If we ever encountered aliens who had rational nature, we would have to treat them as ends as well. But what does it mean to treat humanity as an end and not solely as a means? To see what Kant is getting at, imagine two different ways you might use someone as a means of transportation: you could hire a taxi driver to take you to your destination, or you could own a slave and command him to drive you to your destination. In both cases, you are using another person as a means for your own ends. But in the case of hiring the taxi driver, you are also treating that person as an end in himself. That's because you are only using him in a way that he is freely choosing to be used. He has made the choice to be a taxi driver, and you are respecting his choice by treating him in accordance with his own choice: paying him the required fee, not requiring him to do anything except drive you to your destination, etc. By contrast, owning a slave treats another person solely as a means to your own ends. The person has no choice in the matter; even if you "pay" the slave with food and board, you are not properly valuing that person's ability to choose for himself how to live his life.



The Formula of Humanity tells us that we may pursue our own ends, and we may enlist the help of others in pursuing our own ends, but we must limit ourselves in that pursuit. We have to be limited by everyone else's freedom to choose their own actions. No matter how valuable our goal is, or how much happiness it would create, we are never allowed to use other people purely as means to reach that goal. Furthermore, this rule applies to how we treat ourselves: we must never undermine our own rational nature for the sake of some lesser goal. For example, the Formula of Humanity would forbid extreme drunkenness or drug use, because these activities undermine our own rationality purely for the sake of pleasure.

It is trickier than it sounds to treat humanity as an end in itself. It is not enough to act in accordance with other people's choices, because there are cases where other people's choices are not truly free and autonomous. Treating someone else's humanity as an end requires us to respect and promote that person's autonomy, their capacity to make free choices for themselves. Here is an example: in order to get some money from someone, I could hold a gun to his head and tell him that I'll let him live if he hands over the money. Suppose the person does hand over the money. Does that mean he freely chose to give me the money, and that my action respected his autonomy? Definitely not, because his "choice" in this case was coerced. I have created a situation where the costs of refusing my request would be prohibitively high, and so I've manipulated this person's ability to choose so that he will choose exactly what I want him to choose. Even if the man would have given me the money anyway (suppose he's generous and would have just given me the money if I had asked in a polite way), he is still not freely choosing to give me the money now, so I am not respecting his autonomy.

There are many more subtle ways to undermine other people's autonomy. For example, we can use emotional manipulation to make people feel compelled to choose the outcome that we want or we can withhold information so that the person will make the choice we want. The Formula of Humanity forbids all such actions unless they are necessary for promoting the other person's autonomy in the long-run. That's why it can be okay for us to manipulate our children into making certain choices (say, by using a system of punishments and rewards to get them to obey family rules), as long as we are doing it to help the children grow into autonomous adults. The Formula of Humanity allows this kind of *benevolent paternalism* if it promotes autonomy.

Kant claims that the Formula of Humanity is equivalent to the Formula of Universal Law. The two formulas do give the same answers in most cases. For example, the Formula of Universal Law says we shouldn't make false promises because a maxim of false promises would be inconceivable as a universal law. The Formula of Humanity also says we shouldn't make false promises. The reason is this: if we lie to someone in order to get her to do something, then we are treating her ability to choose how to act purely as a means to our own end. We're preventing her from making her own choice about how to act, and we're doing this in order to meet our own goal. By contrast, if we asked someone for money and told her the truth about our inability to pay it back, we would be treating that person both as a means and an end. We're still using her as a means to get

some money, but we're doing it in a way that allows her to make her own choice about whether to give us the money.

It's no accident that the two formulas give the same answer in most cases: treating someone purely as a means would involve acting on a maxim you yourself could not will as a universal law, since you could never will for others to treat you in a way that doesn't respect your own rational nature. It would be a contradiction to choose to be treated in a way that undermines your own ability to choose. Conversely, whenever you act on a maxim you couldn't will to be universal law, you're acting in a way you wouldn't want others to act. So you're basically using other people's law-abiding behavior as a means to your own ends, and doing it in a way they couldn't agree to. So there is a strong conceptual link between the two formulas. However, the Formula of Humanity may be the superior formula, because it doesn't have some of the problems we saw earlier in the Formula of Universal Law. We saw that a maxim of going to the gym at 5 am doesn't look universalizable; but the Formula of Humanity explains why that maxim is still permissible: going to the gym at 5 am to avoid crowds *does* rely on other people not acting the same way, but it does it in a way that other people could agree to. Nobody is being used against their own free will in this case.

Unfortunately, the Formula of Humanity brings out a deeper problem in Kantianism: what do we do about beings that don't have rationality and are not ever going to become autonomous? For example, what about terminally ill babies, people in permanent comas, or severely mentally handicapped people? Is it okay to treat them solely as means? Would it be alright to run medical experiments on them, since they don't have any "humanity" that needs to be treated as an end? Our deontological intuitions tell us that these less-rational beings also should be treated as ends and should not be used purely as means to our own goals, but Kantianism doesn't give us much theoretical support for that intuition.

#### Comprehension Questions.

1. What is the Formula of Humanity?
2. What is the difference between treating someone as an end and treating them solely as a means?
3. What is the connection between the Formula of Humanity and the Formula of Universal Law?
4. What is the main problem with the Formula of Humanity?

#### Study Questions.

1. Describe a case where consequentialism would morally require that we treat someone solely as a means.
2. In the case you just described, do you think the consequentialist analysis or the analysis by the Formula of Humanity is superior? Why?
3. What would the Formula of Humanity say about telling white lies? Come up with your own example.

## Applications of Deontological Theories

Case 3. Dr. Feuereven has developed a new statistical method for solving a particular measurement problem, a method he believes to be demonstrably more reliable than either of three currently accepted standard procedures. Using this method, Feuereven has reanalyzed several published data sets, and new data he has just finished collecting, and produced the most precise measurements to date. The precise measurements yielded by this technique conclusively confirm a theory favored by Dr. Feuereven's erstwhile colleague and most ardent critic, Dr. Hopfield. Any of the more standard techniques yield less precise results that appear to support Dr. Feuereven's preferred theory. Dr. Feuereven believes that the editors of *Science* will reject a paper that employs the new statistical method, precisely because it is new. He believes, however, that *American Naturalist* would publish such a paper. Because he much prefers publishing in *Science*, he submits and eventually publishes a paper in *Science* using the measures generated by the old techniques to support the theory he now believes is false. Has Dr. Feuereven acted in an ethically defensible manner? Does your evaluation of his actions change if we assume that he is correct in his assessment of what the editors of *Science* would have done had he submitted a paper employing the new statistical method? Under this assumption, are the editors of science acting in an ethically responsible fashion?

Analysis of Case 1. Let's apply the Formula of Universal Law to this case. To apply this formula, we first need to know what Dr. Feuereven's maxim of action is. In this case, his maxim is: I will submit a paper to *Science* that makes what I think is a poor argument for a theory that I believe is false because I want to publish in *Science*, and I don't think *Science* will publish an article making the argument that I think is good. It might also be part of Dr. F's maxim that he doesn't want to publish an article supporting the theory of his rival, but to make Dr. F's maxim as unobjectionable as possible, let's imagine that his rivalry with Dr. Hopfield is not the motivating factor here.

Our next task is to universalize this maxim: what if everyone submitted scientific papers that they believed were of poor quality and had false conclusions simply in order to get a publication in the journal of their choice? It's hard to imagine how such a world would be possible. If people didn't aim at the truth in their scientific writings, their articles wouldn't count as scientific. So-called scientific journals would actually be in the business of publishing fiction or editorials. And even if such a world is conceivable, it would still contradict Dr. F's own purposes. He wants to publish in *Science* because it's a respectable journal and he wants his article to be taken seriously. But if people were always submitting articles they had good reason to think were shoddy, there would be nothing respectable about any journal and nobody would take any of the articles seriously. Dr. F's purpose is only achieved in this case as long as most other members of the scientific community act more honestly. So his maxim is impermissible, according to the Formula of Universal Law.

But suppose Dr. F is correct that the editors of *Science* are biased against new methods and wouldn't have published his other, better, paper? One interesting thing about Kant's view is that even if others are acting immorally, that doesn't excuse immoral action on

your part. All that matters is what your own maxim is, so it would still be wrong for Dr. F to submit a dishonest article, even if he has to do so in order to get published. The best Dr. F can do is fight the bias in *Science* in a permissible way: by publishing editorials, writing a letter to the journal, etc.

Of course, if the editors of *Science* really are biased in this way, they are also acting on immoral maxims. Each editor's maxim might be this: I will turn down any article that employs methods that are too new, no matter how well-defended that method is and how good the quality of the article is, in order to avoid taking a risk and appearing irresponsible. What if all journal editors did this? Such a world is conceivable; it would simply be a world in which science progresses much more slowly, or scientists find different outlets for publicizing their work. But can a *journal editor* will such a world? The original point of rejecting new articles was to preserve their own credibility and standing in the important world of scientific journals. But if scientific journals did not serve the role they serve, of publishing important advances and new methods, then there would be no point to maintaining a credible scientific journal. As editors of a scientific journal, these people have certain values and goals which would be undermined by systematic bias. Their goal of running a respected journal is only met in a world where most people running journals do so without certain kinds of bias. So their maxim of turning down good articles is impermissible.

Case 4. A small liberal arts college with a biology department of four members loses two of those members during the same year. The department seeks to replace them and advertises two positions, one in ecological field studies, the other in medical anatomy and physiology. The best applicants for the two positions happen to be recent graduates from the same major university, at which they formed a homosexual marriage. The department turns the applicants down on the grounds that as a couple they would form a voting block with undue political influence given the size of the department. The applicants object that had they 'stayed in the closet' there would have been no basis for their rejection. Have they been unfairly wronged by the department's decision?

Analysis of Case 4. Let's use the Formula of Humanity for this case. When using the Formula of Humanity, we don't need to do a universalization test; we just need to figure out whether the agent's actions treated someone's humanity solely as a means and not at the same time as an end. In this case, the action involves refusing to hire two otherwise qualified people because the two people are married and might team up in departmental decisions, which would give them more than the usual amount of power in the department.

This case is complicated. The first thing we should sort out is what the Formula of Humanity says about homosexuality versus heterosexuality. The short answer is: nothing. As long as someone has rational nature, he or she must be treated as an end. Since sexual orientation has no bearing on whether one has a rational nature, it cannot by itself affect the basic right a person has to be treated as an end.

Next, we need to figure out what a hiring department owes to any applicant as a rational being. The Formula of Humanity implies that all legitimate agreements must be abided by. If rational people entered a cooperative activity with certain rules in place, it would be unfair for either party to change the rules without the other's agreement. When someone does not play by the agreed-upon rules, or does not act as she agreed to act, she is using the other person's rational choice to join the cooperative venture purely as a means to her own purposes.

What are the agreed-upon rules in this case? The applicant has implicitly agreed to submit an honest application, to bargain in good faith, and to disclose certain information that is relevant to the hiring decision (where she earned her degree, what her research plans are, etc.) However, some information is protected by law; an applicant is not obligated to disclose a disability, marital status, race, ethnicity, or age, even if that information would be relevant to the hiring decision. The hiring department, on the other hand, has agreed to conduct a fair, good faith job search, to evaluate each individual applicant on his/her merits, and to hire the best applicant *all things considered*. The department is allowed to consider such things as collegiality, over-all departmental functioning, and conflicts of interest.

The department is in a difficult situation in this case. Considered individually, each applicant is qualified and has no personal relation to any *current* member of the department. Furthermore, it would have been permissible for the candidates to conceal their marital status during the hiring process. So respecting each applicant as a rational being would seem to require hiring each applicant, and thus hiring both applicants. However, a department is allowed to consider potential conflicts of interest, and in this case there may be such a conflict. As a department member, each applicant would have an interest in doing what's best for the department, and in keeping some information confidential from other members of the department (for example, if three members meet to discuss tenure for the fourth member, some of the details of that discussion will be kept confidential from the fourth member). But as a married person, each applicant would also have an interest in maintaining household harmony and in sharing maximal information with the other. These interests can and probably will pull in opposite directions. Some of these conflicts can be mitigated, and often are in departments in which there are married couples; for example, people recuse themselves from discussions involving their spouse. But in a four-member department, there is really no way to circumvent the conflict of interest between doing what's best for the department and doing what will maintain marital harmony, so the department has a legitimate reason not to have any married couples as members. Since conflicts of interest are legitimate considerations in hiring decisions, and the applicants know this, it would not violate the applicants' rationality to refuse to hire both of them.

#### STUDY QUESTIONS:

1. Walter Jones' teenage daughter wishes to become a large animal vet. In order to provide her with experience caring for large animals, Walter has bought his daughter a horse, and 40 acres of prime grazing land on which to pasture it. Walter shows the

pasture to a friend, who happens to be a botanist working for the state natural history society. The botanist discovers a thriving population of endangered gentians on Walter's land, and informs him of this fact, and the further fact that while horses love gentians, gentians are not viable in habitat subject to grazing by large herbivores. Walter's view is that the land is his, bought and paid for, to be used as he sees fit. He prefers to use the land to pasture his horse rather than as a sanctuary for some endangered plant, and so informs the botanist. Should the botanist report his discovery to responsible state and federal authorities, or should he remain silent?

2. Dr. Green is conducting a study of wolverines in a remote part of Canada. He has a bright young woman apply to work with him specifically on the study of wolverines. He tells her about the primitive field conditions in which they will work and she is still excited about the study. During the study Dr. Green forms a strong emotional attachment to the student, and she seems to be forming the same type of attachment to him. Dr. Green is faced with the decision of having one or both of them discontinue research or face the probability of a romantic attachment forming. If the work is discontinued by even one of them, much valuable data will be lost. What should Dr. Green do? What should the student do? Would your answers be the same if the professor were female and the student male? Does the marital status of the people involved make a difference in your answer?

## Hybrid Theories: Rule-Consequentialism

As we've just seen, consequentialism and deontology both have their strengths and weaknesses. Consequentialism supports our intuition that human happiness is the most important thing, and that we shouldn't stick to inflexible moral rules when that will have disastrous consequences. On the other hand, it allows us to do seemingly terrible things to people if that will maximize over-all happiness. Deontology gives each individual a protected sphere of rights and prohibits us from treating others in terrible ways for the sake of the over-all good. On the other hand, it also forbids us from performing certain types of actions no matter how dreadful the consequences of not performing those actions will be. Is it possible to combine the best aspects of both kinds of view to develop something stronger?

One type of theory that tries to do this is Rule-Consequentialism. A rule-consequentialist is still a consequentialist, and he/she holds that preference-satisfaction is the most important thing in the moral universe. But instead of requiring people always to maximize preference-satisfaction, the rule-consequentialist says people should consistently follow a certain set of rules. The rules are determined by figuring out which rules would maximize over-all satisfaction if followed consistently by everyone. For example, a society that prohibits killing innocent people will be much happier in the long-run than a society that allows people to kill others whenever they think that will maximize happiness. So a rule against killing innocent people is a good rule to have, and everyone should follow it, even if in a particular case it might be better to kill an innocent person. A similar case can be made for rules against lying, stealing, torturing, forced medical experimentation, etc.

The strength of rule-consequentialism is that it explains why people have certain duties not to treat each other in terrible ways, even if that treatment would maximize happiness in a certain case. Yet, this view can also make room for exceptions to the rules in extreme cases. Since the reason a rule is good is that it maximizes happiness in the long-run, there can be exceptional circumstances where the rule is no longer a good one to follow. The rule about killing innocent people is a good one we should all follow. However, if an alien did come and threaten to kill millions if we didn't kill one innocent person, we would have a good reason to give up the rule about killing innocent people since that rule would no longer maximize happiness in the long-run.

Rule consequentialism also provides ways of avoiding the problem of predicting the future. In order to know what the right action is on any given occasion, you need not work out the detailed consequences of each possible course of action on the spot. Instead, you can simply consult a short list of rules. In order to discover which rules, of course, you need before hand to have figured out which rules, if generally followed, would maximize happiness. Difficult as that task may be, it looks to be a bit easier than working out the future consequences of every possible action, and moreover, need not be done on the fly.

Unfortunately, this strength is also a weakness. It's true that we're all much happier when there's a general rule prohibiting killing innocent people. If we knew that others might kill us at any time, we'd all be on edge all the time and cooperation and trust would be impossible. But suppose you're in one of the exceptional cases where killing an innocent person would be best. Suppose you could kill someone and harvest his organs, and you could do it completely secretly so that it wouldn't break down trust in society. What could possibly be *your* reason for following the rule in this case? If you break the rule right now, you'll save 5 people's lives without causing the rule itself to become ineffective for everyone else. The rule itself maximizes happiness, but that very logic supports your decision to break the rule in this one case. And the same logic can be used to justify breaking all the rules whenever that would maximize happiness. Thus, rule-consequentialism seems to be no different in practice from other forms of consequentialism.

The problem here seems to be that our moral intuitions pull in two different directions. We have strong consequentialist and strong deontological intuitions, and it is going to be very difficult, if not impossible, to develop a theory that fits with all our intuitions. For now, it may be best to keep both types of theory in our moral "tool kit" and use whichever one seems most appropriate to our current circumstances. That requires a lot of individual judgment.

#### COMPREHENSION QUESTIONS.

1. What is Rule-Consequentialism?
2. How does it differ from other forms of Consequentialism?
3. What is the main problem with Rule-Consequentialism?

#### STUDY QUESTIONS.

1. Find a statement in a newspaper op-ed piece or letter to the editor that clearly relies on consequentialist reasoning to support a moral claim.
2. Find a statement in a newspaper op-ed piece or letter to the editor that clearly relies on deontological reasoning to support a moral claim.
3. BONUS: Find a statement in a newspaper op-ed piece or letter to the editor that clearly relies on Rule-Consequentialist reasoning to support a moral claim.
4. Analyze the case of Dr. Green using a rule-consequentialist theory:

Dr. Green is conducting a study of wolverines in a remote part of Canada. He has a bright young woman apply to work with him specifically on the study of wolverines. He tells her about the primitive field conditions in which they will work and she is still excited about the study. During the study Dr. Green forms a strong emotional attachment to the student, and she seems to be forming the same type of attachment to him. Dr. Green is faced with the decision of having one or both of them discontinue research or face the probability of a romantic attachment forming. If the work is discontinued by even one of them, much valuable data will be lost. What should Dr. Green do? What should the student do? Would your answers be the same if the professor were female and the student male? Does the marital status of the people involved make a difference in your answer?



## Research Ethics

### Professional Codes: Introduction

Many professional associations have a code of ethics which is taken to govern the professional behavior of practitioners. Among such professions, of course, are Physicians and Lawyers, but so too are such professions as chaplains, architects, engineers, home builders, librarians, and translators. Not surprisingly then, many professional associations of scientists also have such a code, among them the American Psychological Association, the American Sociological Association, the American Physical Society, American Geological Institute and the American Chemical Society. The status of these codes of conduct varies in two respects. First, there is variation in the extent to which the professional association in question *enforces* these codes by punishing, formally or informally, members of the profession who violate them. So, for example, a lawyer or medical doctor who violates the code adopted by her profession risks losing the ability to practice her profession. These societies are able to use the force of law to ensure that the codes they adopt are not violated. This is not true of most other professional associations.

Second, there is variation in the way the professional code is understood to relate to the moral requirements anyone has to live up to, whether or not she is a member of the profession. One way to put the issue is this: does the professional code confer obligations that conflict with, and sometimes override, the obligations we have just in virtue of being humans? Some have thought so, for at least some professions. So for example, it used to be universally held, and is still often held, that physicians may intentionally mislead patients, provided certain other conditions are met. It used to be that physicians would routinely withhold a diagnosis if a patient was terminal, but some kind of treatment was regarded as possible. The idea was that the patient might survive if he complied with the treatment recommendations, which he would not do if he knew the diagnosis was terminal. Such intentional misleading in everyday life would be recognized as seriously immoral (if your auto-mechanic did it, for example). So here it is thought the professional obligation to save lives *trumps* the everyday obligation to speak truly about one's diagnoses to those who have employed you to provide such a diagnosis. When a professional code does trump everyday morality in this way, there is said to be *role differentiation*.

Codes for scientific associations fall somewhere in the middle of both ranges of variations. Violation of NSF, NIH or NEH standards on scientific integrity can lead to loss of the ability to apply for grants, and in some cases civil or criminal legal action. On the other hand, the scientific association is not directly involved in such actions in the same way as a state bar association, for example. Similarly, while some think that scientists are role differentiated, most don't, and even those who do think that such role differentiation as does exist is minimal.

Professional codes of conduct are not ethical theories—the codes do not aim to say what moral obligations a professional *has*, but instead aim to say what moral obligations a

profession *will be held* to have. So a code does not so much say ‘professionals have these obligations’ so much as it says ‘we believe professionals have these obligations, and we will judge our members accordingly’. But even so, the relation between a professional code of ethics and the truth about morality is important. First, because sometimes we can have very good reason to think that the professional code requires behavior that is, in fact, seriously immoral. This may result from role differentiation, or from an ethical code that *presupposes* a mistaken ethical theory. The question then arises as to which course of action should be taken. Second, because the obligations recognized by a given professional code generally have their foundations in more than one kind of ethical theory, it is often the case that these codes will sometimes demand that professionals do two different things in one and the same circumstance. Before the professional can act, she must resolve the conflict. And third, professional codes change over time. They change in response to what the members of the profession regard as morally proper behavior for members of the profession. That is, what standards entomologists, say, are held to depends on what standards entomologists think they should be held to. Since professionals change the codes they endorse, those changes ought to be directed by moral theory. If a code endorses what now seems to be immoral behavior, or fails to proscribe behavior that now seems to be immoral, it ought to be changed. To change it, the professionals must recognize the inadequacy of the current code. And that requires that the current code be subject, to some extent, to evaluation *as if* it were an ethical theory. In this section we will begin doing that for some particular codes of scientific ethics.

Here is the code of ethics from the American Chemical Society.

“**Chemists Acknowledge Responsibilities To:**

- **The Public**  
Chemists have a professional responsibility to serve the public interest and welfare and to further knowledge of science. Chemists should actively be concerned with the health and welfare of co-workers, consumer and the community. Public comments on scientific matters should be made with care and precision, without unsubstantiated, exaggerated, or premature statements.
- **The Science of Chemistry**  
Chemists should seek to advance chemical science, understand the limitations of their knowledge, and respect the truth. Chemists should ensure that their scientific contributions, and those of the collaborators, are thorough, accurate, and unbiased in design, implementation, and presentation.
- **The Profession**  
Chemists should remain current with developments in their field, share ideas and information, keep accurate and complete laboratory records, maintain integrity in all conduct and publications, and give due credit to the contributions of others. Conflicts of interest and scientific misconduct, such as fabrication, falsification, and plagiarism, are incompatible with this Code.
- **The Employer**  
Chemists should promote and protect the legitimate interests of their employers, perform work honestly and competently, fulfill obligations, and safeguard proprietary information.
- **Employees**  
Chemists, as employers, should treat subordinates with respect for their professionalism and concern for their well-being, and provide them with a safe, congenial working environment, fair compensation, and proper acknowledgment of their scientific contributions.

- **Students**  
Chemists should regard the tutelage of students as a trust conferred by society for the promotion of the student's learning and professional development. Each student should be treated respectfully and without exploitation.
- **Associates**  
Chemists should treat associates with respect, regardless of the level of their formal education, encourage them, learn with them, share ideas honestly, and give credit for their contributions.
- **Clients**  
Chemists should serve clients faithfully and incorruptibly, respect confidentiality, advise honestly, and charge fairly.
- **The Environment**  
Chemists should understand and anticipate the environmental consequences of their work. Chemists have responsibility to avoid pollution and to protect the environment."

Some parts of the code are clearly inspired by deontic considerations: the injunction to serve clients faithfully, to respect associates, to give credit for the contribution of others, and to avoid conflicts of interest fall in this area. Others are less clear, but are still probably best explained by appeal to elements of a deontic ethical theory. The injunctions to 'offer tutelage' to students, to respect students, and to work in good faith for the interests of one's employer seem to fall into this category. But other elements of the code are best explained by appeal to elements of a *consequentialist* theory, at least to first appearance. For example, the injunctions listed under *The Environment* and many of those listed under *The Public* are probably best so explained.

Given the disparate sources of the duties imposed by the code, it should be unsurprising that these duties can, and often will, conflict. What if best serving the interests of your employers will require developing a highly carcinogenic chemical? One knows with reasonable certainty that this chemical, if used with any frequency, will sometimes be released. Hence developing the chemical will arguably contravene the injunctions not to pollute and to serve the public interest and welfare. But not to do this will violate the injunctions to advance science and to serve the interests of one's employer. One can seek a resolution in a simple minded way by counting up the obligations satisfied or violated by different courses of actions, and then choosing the action that satisfies the most obligations and violates the least. But such a course supposes that the obligations are all equally important, and this is unlikely. So individual chemists must, in such situations, come to some decision about the relative priority of these various obligations, and a weighting of them, in order to find the ethically correct course of action.

This consequence is important in several ways. As a practicing researcher, one *cannot* avoid confronting ethical problems, and working through them, simply by consulting the guidelines of your profession. The guidelines are ambiguous. If a researcher does not think through such dilemmas, or does so uncarefully, she is in effect gambling with her moral fate. She may, by luck, choose a defensible course of action. But more likely, she will choose a course of action that is sub-optimal in ways that make it immoral.

Also notice that in the case at hand, the dilemma arises from a conflict between deontic and consequentialist obligations. This sort of conflict is very frequent; it will arise all the time in lots of different contexts. So resolving the relative importance of deontic and consequentialist demands turns out to be crucial to finding the right course of action in lots and lots of cases, cases that any researcher is likely to confront with some frequency over the course of her career. As a general rule, we are, and should be, inclined to take consequentialist demands more seriously when we are considering matters of *policy* (which will affect many people) than when considering contexts in which there are no policy implications. But this general rule of thumb is no more than that—there are many cases of ‘local’ non-policy decisions in which consequentialist considerations might be quite important, and lots of policy questions where deontic considerations are at least as important as consequentialist considerations. Which kinds of demands matter most should always be determined by considering which sorts of demands *best serve the intrinsic values* at stake. So if the question at hand concerns how to treat a particular employee or colleague or student, respect for autonomy will be a value of essential importance, though not necessarily the only value of importance. Respect for autonomy will also be important if the question involves a choice of *policy* governing the treatment of many students, colleagues or employees, but other, consequentialist, values will also be essential here: which policies lead to the best educational outcomes on average, which lead to the most efficient progress in science, and so on.

We need to make a final note about terminology. There is, as we have seen, a distinction to be drawn between what one ought to do in a given situation, and what the code of ethics governing your profession *says* one ought to do in that situation. Often, the true ethical theory and the professional code will recommend the same action. But sometimes they will not, and so it is always important to ask whether the action recommended by the code *really is* morally correct. Many people who write about professional ethics but are not themselves by profession ethicists use a particular terminology to do this. Behavior recommended by a professional code is said to be ‘ethical’, behavior contrary to such recommendations is said to be ‘unethical’. Conversely, behavior which is permissible by the lights of the true moral theory, whatever it is, is said to be ‘moral’, and behavior contrary to that theory is said to be ‘immoral’. That terminology is **not** used by professional philosophers—for them ‘ethical’ and ‘moral’ are synonyms and so too are ‘unethical’ and ‘immoral’. The first pair denote behavior that is consistent with the true moral theory, the second behavior that is inconsistent with the true moral theory. We, being professional philosophers, use the terms in this way. When you read other discussions of professional ethics, one thing you will need to do is determine which terminology the author in question is using.

#### Comprehension Questions.

1. What is *role differentiation*?
2. Are professional codes of conduct *ethical theories*?

3. Can professional codes be assessed as *if they were* ethical theories, and if so why would we care about doing that?
4. Will professional codes be grounded in one systematic, ethical theory? How can we tell?
5. What is the difference between ethical behavior and moral behavior?

Study Questions.

1. Find a code of conduct promulgated by some association of researchers in your area.
2. Which of the rules in this code do you think presupposes a deontic motivation?
3. Which of the rules presupposes a consequentialist motivation?
4. Which of the rules could be justified either by appeal to a deontic theory or by appeal to a consequentialist theory?
5. Does the code prioritize the various rules?
6. Are there any rules that you think are mistaken—i.e. do not demand enough, or demand too much, of researchers? If so, why?
7. Pick two unprioritized rules and say which one you think ought to take priority, and say why. Consider some examples in which the rules can conflict, and apply your prioritization there. What intrinsic values are best served by your prioritization?
8. Does the code as it stands imply any role differentiation? If so, say in what respect this is so. If not, say whether you think it *ought* to do so, and why.
9. Go to the letters sections of Science or Nature, and find a discussion of some ethical issue which proposes a general rule for some kind of research activity (publishing, sharing data, informed consent for research subjects, rules about advocacy, etc.). Is this rule a good rule? Why or why not?

## Research Ethics: Misconduct

There are a wide range of ethical dilemmas researchers may confront as part of their research activities. These range from scientific misconduct, to issues of ownership of data, techniques or results, proper publication practices, record keeping, abuses of power, fair treatment of research subjects, whether human or animal, and the moral propriety of particular lines of research, and conflicts of interest. In this section we focus on the first of these areas, scientific misconduct.

Misconduct is typically held to include three broad categories of immoral behavior: fabrication of data (making it up), falsification of data (reporting only some of the data when the omitted data are relevant), and plagiarism (stealing the data, arguments, conclusions or written work of others). For example, the Office of Research Integrity (ORI) defines scientific misconduct as follows:

“Sec. 93.103 Research misconduct.

Research misconduct means fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results.

(a) Fabrication is making up data or results and recording or reporting them.

(b) Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

(c) Plagiarism is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit.

(d) Research misconduct does not include honest error or differences of opinion.” ([http://ori.dhhs.gov/documents/FR\\_Doc\\_05-9643.shtml](http://ori.dhhs.gov/documents/FR_Doc_05-9643.shtml))

Its standards for a finding of conduct seem somewhat broader, but also more permissive in that they are sensitive to common practice:

“Sec. 93.104 Requirements for findings of research misconduct.

A finding of research misconduct made under this part requires that--

(a) There be a significant departure from accepted practices of the relevant research community; and

(b) The misconduct be committed intentionally, knowingly, or recklessly; and

(c) The allegation be proven by a preponderance of the evidence.” ([http://ori.dhhs.gov/documents/FR\\_Doc\\_05-9643.shtml](http://ori.dhhs.gov/documents/FR_Doc_05-9643.shtml))

When confronting a specific rule in a code of conduct, there are three general questions we want to work through. First, the rule will concern a specific kind or kinds of behavior, and it is essential to ask what reasons we may have for avoiding such behavior or engaging in it—that is, what intrinsic values will be served by such behavior, and to what values will such behavior be a disservice. Given answers to this first question, we can ask about the cogency of the rule: will following this rule *best* serve the relevant values, or would some other rule be better. And once we have settled on the rule, we can ask how to apply it in particular cases.

The ORI rule regarding scientific misconduct (namely, don't engage in falsification, fabrication or plagiarism as defined above) really concerns two distinct kinds of behavior: reports that mislead readers about the data one has gathered, and reports that mislead readers about who actually generated the reported results, whether these be data, methods, conclusions or descriptions of any of these. The intrinsic values relevant to each kind of behavior are in some respects different, so we will consider each in turn.

Why should we avoid *making up* data, or misreporting the data by omitting some of it or by misdescribing the procedures by which the data were generated? There are a number of values relevant here. First, if the research has been funded by a granting agency, such misrepresentation is a kind of fraud. In effect, the researcher has contracted to do a job; the job is not done at all, or is ill-done, when the misrepresentation occurs. So there is a deontic value here, that of keeping one's promises. Second, such misrepresentation misleads readers, so it is therefore contrary to the deontic value of honesty. Further, this dishonesty is contrary to at least three consequentialist values. Readers will make decisions about what further experiments to perform, and what assumptions can be made in performing them, on the basis of the reported data and results. When the data are falsified and the conclusions mistaken, those decisions will be bad, wasting the time, money and effort of those readers, and slowing the progress of science. Further, when the mistakes are discovered, others will have to backtrack, identify the mistakes, correct the misconceptions and so on, further wasting time, money and effort. Finally, at least for some research, especially in medicine, social science and engineering, results will inform private decisions and public policy. When decisions, private or public, are made on the basis of false conclusions, people are prevented from effectively satisfying interests and preferences they otherwise would have been able to satisfy.

In some ways, the consequentialist motivations for avoiding falsification and fabrication are the strongest—people can be (and have been) severely harmed by such misconduct. For example, for many years Great Britain based its educational policy on sociological conclusions defended by fabricated data. That policy determined the educational fate of primary school students quite early, preventing many from pursuing college or university training. Since many of these students could have done very well at university, and would much have preferred to go to university and to pursue careers that are available only to those with university degrees, the educational policy caused them much irreparable and lifelong harm. On the other hand, the consequentialist motivations are subject to a worry—if the conclusions one supports using falsified or fabricated data turn out to be *correct*, then rather than wasting time, effort and money, and working irreparable harm, the misconduct in fact saves time, effort and money, and leads to *better*, not worse, private and public decision making.

This worry suggests that sometimes we can have very good *moral* reason for engaging in fabrication or falsification—if we are certain our conclusions are correct, but also that they will not be accepted unless data are fabricated or falsified, then it may seem as if we have very strong consequentialist reasons for fabricating or falsifying. But this would be wrong. We cannot be certain our conclusions are true, indeed we cannot even have *good*

*reasons* for thinking our conclusions are true, unless we have unfalsified, non-fabricated data to support those conclusions. Thus, it is a necessary condition for having good moral reasons to engage in this sort of conduct that such conduct is entirely unnecessary.

The motivation for avoiding plagiarism is somewhat different. Since plagiarism does not introduce new data or conclusions, its consequential effects seem negligible—at first glance, only the plagiarist and her victims are helped or harmed in any substantive way by the behavior. But there are significant deontic values at stake. Plagiarism is a form of theft, and to the extent that there theft is contrary to deontic values—property rights or respect for autonomy, there are good reasons to avoid it. And once this is recognized, a consequentialist value can be seen. If plagiarism is common, people will be less inclined to disclose useful discoveries, since by disclosing them rather than privately exploiting them, they make theft possible. Since scientific progress depends on free and open communication, plagiarism is a disservice to the progress of science, and hence to all those whose lives are made longer and better by the advance of science.

These then are the values served by avoiding misconduct—the reasons we have for not engaging in it. Misconduct is contrary to the deontic values of keeping one’s promises, of honesty and of property rights, it is contrary to the consequentialist value of scientific progress, and to the consequentialist values of enabling informed and effective decisions in both public and private spheres. Finally, of course, given that misconduct is often discovered and punished, it is contrary to one’s own self-interest, and this is a value that must be considered from the consequentialist perspective.

How well does the ORI rule serve these values? Reasonably well. We might have doubts, for example, had we discovered one or more special cases in which misconduct seemed, on the whole, morally justifiable. But we found no such cases. We might otherwise have doubts if we thought the rule would count as misconduct behavior which, on reflection, we judge really isn’t misconduct. For example, some have claimed in commenting on the ORI rule that misconduct really depends on an *intention to deceive*. The idea is that if there is no deceptive intent, we might have bad science, but not scientific misconduct.

There are two ways to interpret this particular response. On one understanding, it is a mere semantic quibble. Suppose we have behavior which satisfies the ORI definition, but which is *not* intended to deceive. Is such behavior morally improper? If so, then the objection is merely a quibble about *naming*. The objector agrees that all behavior that falls under the ORI definition is morally wrong, she just wishes to call some of this behavior by another name than misconduct. There is here no substantive objection to the ORI rule. Differently, the objector might be claiming that behavior which satisfies the ORI definition, but which is *not* intended to deceive, is not immoral. That is real difference.

So we should ask—how much, if any, difference does an intent to deceive make for the permissibility of ‘misconduct’ as defined by the ORI? We think not much. First, the presence or absence of an intent to deceive can have no affect whatsoever with respect to



the consequential values. If a fabrication, falsification or plagiarism is immoral given the intent to deceive, it is just as immoral without it, from the perspective of any consequential theory. Second, even with respect to the deontic values, the absence of an intent to deceive can have no effect on whether or not one is keeping one's promises. So fabrication and falsification of data in work supported by grants will remain deontically immoral for that reason. Matters are harder when considering respect for property rights and the value of honesty. It may be that, absent an intent to deceive, no deontic rule about respect for privacy rights or honesty is flouted by 'misconduct' as defined by ORI. But since these concerns seem to us much the least pressing reasons for regarding 'misconduct' to be immoral, we think ORI has come pretty close to getting it right.

This leaves us only with the third question: how do we apply the rule in specific cases? For fabrication and plagiarism, this is simple. Do not make up data, and do use the ideas, methods or words of others without citation. Falsification is more difficult, since one must often make judgments about what data to report, and what data to analyze. There are a number of reasons for this. When data are gathered in observational rather than experimental contexts, there are likely to be missing data, or one may have doubts about the reliability on one or more survey questions, or one or more observers, and so on. In experimental contexts, one may need several trial runs before becoming confident that one is operating a laboratory device properly, and once trials have begun, one might have good reason to think a particular observation results from a failure of some piece of equipment, or from contamination, and so on. Bad data should be removed; on the other hand, it is not always obvious which data are bad. There are some good rules of thumb. When there are methods for testing the reliability of the data (test-retest consistency, and inter-observer reliability, for example), they should be used, and the results reported. When data are excluded, this should be noted, along with some reason for their exclusion. If excluding a datum or set of data materially affects the analytic results, this should be noted. If suspect data are included, one should note that the data are suspect, and why, and also whether the inclusion of them materially affects the analytic results or introduces a possible source of error that might otherwise have been avoided.

A general, principled test is this: if I omit (or include) this suspect datum, without noting that I am so doing, is it possible that in so doing I will materially mislead a reader in ways that cause her to over (or under) estimate the strength of the argument for my conclusions, or lead her to engage in, or forego, or change, studies she otherwise would not have engaged in, foregone, or changed. If the answer is yes, disclosure of the assumptions *I* am making is essential. If not, there is less pressing *moral* reason for including the extra information.

The rules of thumb and the general test are important, but so too is experience in reasoning through specific cases. We shall do that in the next section.

#### COMPREHENSION QUESTIONS

1. What three activities constitute scientific misconduct?
2. What intrinsic values are relevant to assessing the moral status of fabrication and falsification?

3. What intrinsic values are relevant to assessing the moral status of plagiarism?
4. Why is intent to deceive irrelevant in a definition of misconduct, according to the above?

### STUDY QUESTIONS

1. Sahotra is a statistical epidemiologist reviewing the efficacy and danger of a new ‘morning after’ contraceptive. Sahotra believes the conclusions reached by his report will be decisive in determining whether or not the drug is made available over the counter. Sahotra himself believes that contraception of any sort, and especially contraceptions which lead to the spontaneous abortion of a conceptus, are seriously immoral. If Sahotra includes all available data in his analysis, the standard methods for assessing the ‘effect of treatment’ show the new drug to be at least as safe as carrying the pregnancy to term for the population at large. However, the data appear to show a non-linear, and possibly significant, relationship between the drug and maternal mortality if one conditions on other risk factors. The sample of women in the relevant risk group is small. By excluding certain outliers and using a method that assumes a *linear* relationship between dosage and mortality, Sahotra can produce an analysis which indicates that the drug is not safe for women in this risk group. However, if he either does not remove the outliers, or uses methods appropriate when the relationship between dosage and morbidity is *non-linear*, the analysis will not support that conclusion. Is it morally permissible for Sahotra to remove the outliers and use the linear method in producing his analysis? Explain.
2. Inga is a social scientist working on the relationship between exposure to TV violence and adult aggression. She has data from a longitudinal study. If she stratifies the sample by sex (i.e. essentially treat the sample of women as a different population from the sample of men), she discovers different relations between exposure and adult aggression in men and women: for men, all models that fit the data include a causal relation between childhood exposure and adult aggression, but for women there is not such relation, instead all models that fit the data for women include a causal relation between *adult exposure* and adult aggression. However, if she does not stratify the data, all the models that fit the data include no relation between childhood or adult exposure and aggression. There are then two possibilities—exposure, whether in childhood or adulthood, really has no effect, and the within group associations are a matter of sample error, or exposure has an effect, but it is different in women and men. The first conclusions are supported by an analysis without stratification, the second by an analysis with stratification. Which analysis should Inga report in her published findings, or should she report both, and why? What intrinsic values are served by applying and following our general rule of thumb in this case?
3. Antony is writing up the results of field-work on upland sandpipers. The data are pretty interesting, because they bear on different theories about how kin selection works. In particular, they seem to show that mate-selection in the upland sandpiper allows for kin selection without geographic isolation of kinship groups. In order to explain the relevance of his findings, Antony must introduce

some of the relevant theoretical background. Unfortunately, Antony finds it quite difficult to write about complicated theoretical ideas (he is much more comfortable describing the data and the conclusions). Antony asks a fellow grad-student Alyssa to ghost write the theoretical section, which she does. Is it misconduct if Antony publishes the paper without Alyssa as a co-author? Is this true even if Alyssa agrees to that arrangement?

## **ORI Cases**

In this section we reproduce ORI summaries from two cases of misconduct, and after each a copy of the code of ethics for the discipline to which the accused belongs. After reading through each case summary and the corresponding code of ethics, try to identify which injunctions in the code were violated, and what intrinsic values are served by those injunctions. Both summaries can be found on line in full at <http://ori.dhhs.gov/misconduct/cases/>

### **Case Summary - Regina D. Horvat, PhD**

Federal Register: June 25, 2004 (Volume 69, Number 122) Page 35629

DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Office of the Secretary, Findings of Scientific Misconduct

**SUMMARY:** Notice is hereby given that the Office of Research Integrity (ORI) and the Acting Assistant Secretary for Health have taken final action in the following case: Regina D. Horvat, Ph.D., Northwestern University: Based on the report of an inquiry conducted by Northwestern University (NU Report), the respondent's admission, and additional analysis conducted by ORI in its oversight review, the U.S. Public Health Service (PHS) found that Regina D. Horvat, Ph.D., former Postdoctoral Fellow, Department of Cell and Molecular Biology at NU, engaged in scientific misconduct in research supported in part by the following National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH) grants: F32 HD041309, RO1 HD38060-01A1, and T32 HD007068 ...

Specifically, PHS found that: Dr. Horvat falsified a western blot of an immunoprecipitation (IP) assay presented as Figure 5B in a manuscript ("Inhibition of Luteinizing Hormone Receptor Desensitization Suppresses the Induction of Ovulatory Response Genes in Granulosa Cells") submitted to Molecular Endocrinology. Dr. Horvat falsely labeled an autoradiogram in her laboratory notebook with a piece of tape to misrepresent the data from a different IP experiment that was actually conducted on October 31, 2001, as the experiment described in Figure 5B. Further, Dr. Horvat falsely used Figure 5B in an oral presentation at a national scientific meeting; and Dr. Horvat falsified the intensity of the band in Lane 6 of a luteinizing hormone receptor (LHR) Western blot experiment to quantitate the level of LHR immunoprecipitated with an arrestin2 antibody in cells treated with hCG for 30 minutes in the PowerPoint figure, prepared in response to the initial review of the Molecular Endocrinology manuscript. This manuscript was withdrawn.

Dr. Horvat has entered into a Voluntary Exclusion Agreement (Agreement) in which she has voluntarily agreed for a period of three (3) years, beginning on

June 2, 2004: (1) To exclude herself from serving in any advisory capacity to PHS including but not limited to service on any PHS advisory committee, board, and/or peer review committee, or as a consultant; and (2) That any institution which submits an application for PHS support for a research project on which the Respondent's participation is proposed or which uses the Respondent in any capacity on PHS supported research, or that submits a report of PHS-funded research in which the Respondent is involved, must concurrently submit a plan for supervision of the Respondent's duties to the funding agency for approval. The supervisory plan must be designed to ensure the scientific integrity of the Respondent's research contribution. Respondent agrees to ensure that a copy of the supervisory plan is also submitted to ORI by the institution. Respondent agrees that she will not participate in any PHS-supported research until such a supervision plan is submitted to and accepted by ORI.

...

Chris B. Pascal,  
Director, Office of Research Integrity.

Here is the code of ethics from the American Society for Biochemistry and Molecular Biology.

American Society for Biochemistry and Molecular  
Biology  
*(approved by the ASBMB Council in January 1998)*

Members of the ASBMB are engaged in the quest for knowledge in biochemical and molecular biological sciences with the ultimate goal of advancing human welfare. Underlying this quest is the fundamental principle of trust. The ASBMB encourages its members to engage in the responsible practice of research required for such trust by fulfilling the following obligations.

In fulfilling OBLIGATIONS TO THE PUBLIC, it is EXPECTED that:

investigators will promote and follow practices that enhance the public interest or well-being;

investigators will use funds appropriately in the pursuit of their research;

investigators will follow government and institutional requirements regulating research such as those

@ Bruce Glymour and Amy Lara  
2005

ensuring the welfare of human subjects, the comfort and humane treatment of animal subjects and the protection of the environment;

investigators will report research findings resulting from public funding in a full, open, and timely fashion to the scientific community; and

investigators will share unique propagative materials developed through publicly-funded research with other scientists in a reasonable fashion.

In fulfilling OBLIGATIONS TO OTHER INVESTIGATORS, it is EXPECTED that:

investigators will have actually carried out experiments as reported;

investigators will represent their best understanding of their work in their descriptions and analyses of it;

investigators will accurately describe methods used in experiments;

investigators will not report the work of others as if it were their own;

investigators in their publications will adequately summarize previous relevant work;

investigators acting as reviewers will treat submitted manuscripts and grant applications confidentially and avoid inappropriate use; and

investigators will disclose financial and other interests that might present a conflict-of-interest in their various activities such as reporting research results, serving as reviewers, and mentoring students.

In fulfilling OBLIGATIONS TO TRAINEES, it is EXPECTED that:

investigators serving as mentors will provide training and experience to advance the trainees' scientific skills and knowledge of ethical research practices;

investigators will provide appropriate help in advancing the careers of the trainees;

investigators will recognize research contributions of the trainees appropriately;

investigators will encourage and support the publication of results of trainees' research in a timely fashion without undisclosed limitations; and

investigators will create and maintain a working environment that encourages cultural diversity.

<http://www.mcponline.org/misc/ethics.shtml>

Here is a second case, importantly different in part because publications had to be retracted.

### **Case Summary - Craig H. Gelband, Ph.D**

Federal Register: November 10, 2003 (Volume 68, Number 217) Page 63797-63799

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Office of the Secretary Findings of Scientific Misconduct

SUMMARY: Notice is hereby given that the Office of Research Integrity (ORI) and the Assistant Secretary for Health have taken final action in the following case:

Craig H. Gelband, Ph.D., University of Florida: Based on the reports of two investigations conducted by the University of Florida (UF) (UF Reports) and additional analysis conducted by ORI in its oversight review, the U.S. Public Health Service (PHS) found that Craig H. Gelband, Ph.D., Associate Professor, Department of Physiology, College of Medicine at UF, engaged in scientific misconduct in research. Publications and manuscripts containing the falsified data cited support from National Institutes of Health (NIH) grants, or falsified data was included in NIH grant applications, as follows:

R29 HL52189-01A2, then R01 HL52189-05, "Regulation of renal vascular cells in hypertension".

R01 HL56921, "AT1 receptor control in chronic hypertension".

F32 HD08496, "Role of MNA V2.3 in uterine contraction".

R01/R37 HL49254, "Ionic & pharmacological regulation of vascular cells".

F32 HL08531, "Hormonal regulation of renal artery ionci currents".

P01 DK41315, "Regulatory mechanism in colonic motility-program project".

R01 HL69034-01, "Mechanisms of cerebral resistance artery constriction."

Specifically, PHS found that:

I. Dr. Craig H. Gelband falsified data based on contractile tension recording in antisense experiments on the angiotensin enzyme (ACE), purportedly using renal arteriolar smooth muscle tension preparation:

A. by falsely labeling the tension recordings in Figures 5, 6, and 7 in a publication by Wang, H., Reaves, P.Y., Gardon, M.L., Keene, K., Goldberg, D.S., Gelband, C.H., Katovich, M.J. & Raizada, M.K. "Angiotensin I--converting enzyme antisense gene therapy causes permanent antihypertensive effects in the SHR. "Hypertension 35[part 2]: 202-208, 2000 (subsequently referred to as the "Hypertension 2000 paper 1"), when he had earlier reported the same contractile records as being from experiments on the angiotensin receptor (not the enzyme), in Figures 6, 7, and 8 of an earlier mini-review by Martens, J.R. & Gelband, C.H. "Ion channels in vascular smooth muscle: Alterations in essential hypertension. "PSEBM 218:192-200, 1998 (subsequently referred to as the PSEBM paper);

B. by falsifying three of the four sets of the mean data that were in fact the same for both the F0 and F1 mean data in Figures 5 and 6 of the Hypertension 2000 paper 1. Dr. Gelband also dishonestly provided the institution with the falsified/fabricated tables of the mean data and the associated false standard error values as evidence that he had conducted the experiments for Figures 5 and 6; and

C. by falsifying EC50 values in Table 1 in NIH grant application HL52189-05; the EC50 values had been interpolated from the falsified mean and SEM data shown in Figures 5 and 6 in the



Hypertension 2000 paper 1.

II. Dr. Gelband falsified data in the reporting of research, misrepresenting current/voltage (I/V) data to be results from totally different experimental models or preparations in six publications (including one manuscript "In-Press") and in NIH grant application HL52189-05, specifically:

A. as Figure 1A, in Gelband, C.H., Wang, H., Gardon, M.L., Keene, K., Goldberg, D.S., Reaves, P., Katovich, M.J., Raizada, M.K. "Angiotensin 1-converting enzyme antisense prevents altered renal vascular reactivity, but not high blood pressure, in spontaneously hypertensive rats. "Hypertension 35 [part 2]:209-213, 2000 (subsequently referred to as the "Hypertension 2000 paper

2").

B. as Figure 2, in Martens, J.R., Fergus, D.J., Tamkun, M.M., England, S.K., Gelband, C.H. "Identification of voltage-gated K<sup>+</sup> channel genes contributing to the decreased renal arteriolar K<sup>+</sup> current in hypertension." J. Biol. Chem (MS

M01389200), online, in press (subsequently referred to as the "JBC paper"). J. Biol Chem Online (submitted and withdrawn).

...

III. Dr. Gelband falsified traces of potassium currents in Figure 4 of the J. Biol. Chem paper (see PHS Finding II) where they were claimed to have been recorded from smooth muscle cells from rats treated with antisense to potassium channels, and/or in Figure 3 of the Hypertension Online paper (see PHS Finding II) where they were claimed to have been records from rat renal cells treated with phorbol esters and PKC inhibitors. Furthermore, the potassium currents were recorded from neurons, not from smooth muscles as falsely reported in these publications. Dr. Gelband falsified data in the proposing of research by submitting the falsified traces of potassium currents as Figure 9 in NIH grant application HL52189-05.

...

VIII. Dr. Gelband fabricated laboratory research records for four Western blot experiments during the investigation, withholding from the institution his associate's notebook from which he had removed four labeled autoradiographic films from separate and different experiments, and using the removed films to fabricate a laboratory notebook containing falsified Western blots, which he provided to UF as evidence that he had conducted the experiments under investigation.

The terms of this Agreement are as follows:

(1) Respondent agreed to exclude himself voluntarily from any contracting or subcontracting with any agency of the United States Government and from eligibility or involvement in nonprocurement programs of the United States Government referred to as "covered transactions" as defined in the debarment regulations at 45 CFR part 76, for a period of ten (10) years, beginning on October 3, 2003.

(2) Respondent agreed to exclude himself voluntarily from serving in any advisory capacity to PHS including but not limited to service on any PHS advisory committee, board, and/or peer review committee, or as a consultant, for a period of ten (10) years, beginning on October 3, 2003.

(3) Within 30 days of the effective date of this Agreement, Respondent agreed to submit letters of retraction to the following journals concerning the specified data in the listed articles:

A. Hypertension 2000 paper 1: Figures 5, 6, and 7 merited retraction. A retraction has been submitted relevant to this paper.

B. Hypertension 2000 paper 1: Figure 1A merited retraction. A retraction has been submitted relevant to this paper.

C. JBC paper: Figure 2 and Figure 4 merited retraction. It has already been withdrawn.

D. Hypertension Online paper: Figure 4A and Figure 3 merited retraction. It has already been withdrawn.

E. Hypertension 1999 paper: Figure 3 must be retracted.

F. PNAS paper: Figure 4A and 4B must be retracted.

G. Circ. Res. 1999 paper: Figure 5A must be retracted.

H. Circ. Res. 1995 paper: Figure 4C or 4B must be retracted.

Chris B. Pascal,

Director, Office of Research Integrity.

The code of ethics for the American Physiological Society reads, in part, as follows.

Code of Ethics

(Adopted by Council October 1989)

Membership in the American Physiological Society includes the acceptance of and the responsibility to uphold the following Code of Ethics.

The role of the physiologist is to advance the field through teaching, research, and service. In the process physiologists shall be honest in their reporting of research findings and ethical in their dealings with others. Moreover, physiologists shall be humane in the treatment of human and non-human subjects. Physiologists shall also have the professional responsibility to bring to the attention of appropriate authorities apparent violations of these principles.

Physiologists recognize the Society's responsibility to consider breaches of ethical behavior and to take any response deemed necessary in accordance with the Society's Bylaws, Article IX, Section 4 and as defined in the Operational Guide.

<http://www.the-aps.org/about/opguide/appendix.htm#code>

## **Workplace, Personal Relations and Mentoring**

As in any workplace, there are a variety of ethical issues that arise from differences in power which affect scientists. In the latter stages, during graduate school and post-doc appointments, training in science resembles an apprenticeship more than in many other professions. The principle investigator in the lab and the dissertation advisor assume responsibility for training students and post-docs that extends beyond suggesting relevant course work or suitable research projects, and into informal, but deeply important, professional practices: how to organize and run a lab, what kinds of analytic techniques to use, when and where to publish, the care to be used when keeping records, the access to preliminary data or conclusions given to colleagues, and so on. Commonly scientists take it to be an obligation to the profession to provide mentoring, and they also think that mentoring may be done well or ill, and when done ill, this sometimes involves moral transgressions. In this section we will consider some of the moral dilemmas concerning interpersonal relations that can arise within the lab or research group.

Whenever there are power imbalances within a group, the first and often overriding concern is that power not be abused. There are two constraints on the abuse of power. The first is the intrinsic deontic value of autonomy. One way of abusing power is in essence a failure to respect the autonomy of whoever is being abused. A second value that is also often of crucial importance is that of paternalistic beneficence. ‘Paternalism’ denotes the act of deciding for another because that person is not competent to decide for himself. Since the exercise of autonomy depends on competence, a non-competent decision maker cannot choose autonomously, freely, and so it does not flout the demand to respect autonomy if others decide for the non-competent person. However, when we engage in paternalistic decisions, those decisions will be legitimate only if we decide on the basis of our best judgment about *the interests of the person for whom we decide*. That is the demand of beneficence. Furthermore, as we saw in the section on Kantianism earlier, the ideal is only to engage in paternalistic behavior when it’s necessary to promote the person’s eventual autonomy.

There are also secondary consequential and deontic concerns. The obligation to provide mentoring itself is motivated by the essential role mentoring plays in training the next generation of scientists, which is instrumentally important for the advancement of science. Further, given the understanding that advisors and principal investigators are *supposed* to provide mentoring, and that this understanding is used to justify the relatively low pay and relatively long working hours imposed on graduate students and post-docs, advisors and principal investigators have a deontic obligation to their students and post-docs to provide mentoring, i.e. to actually train students in ways that lead them to make good professional choices. However, the deontic value of respect for autonomy often conflicts with the social climate within a lab or research group, which climate is often paternalistic and sometimes quite authoritarian. Lab directors will often have the final say in what gets investigated, how it gets investigated, who does the work, what work is published, whose names appear as authors, what inventions get patented, whose names are on the patents, and where students and post-docs apply when leaving the lab or group. And while individual scientists may solicit and act according to comment from

members of the research group, there is typically no *legal* requirement that they do so. That is quite a lot of power.

Commonly, there are at least minimal institutional restrictions on how this power is exercised, but in practice, once a student or post-doc has joined a group, she will not be in a position to resist or override the recommendations of the lab director or advisor, and she will have very little choice but to comply if she wishes to continue in her profession. This deep power imbalance imposes obligations on both student and research director.

Let us deal briefly, first, with the obligations of the student. The student properly has strong prudential reasons to make sure she is comfortable with the working environment in a lab or research group, and with the standard practices of the director, before joining a group. But these obligations are moral as well. For example: John knows that a particular PI, Dr. Sally, insists on an otherwise perfectly moral policy, say not allowing leaving post-docs to take samples of their breeding strains with them. Suppose John intends to do work developing a new strain of bacteria, and further intends to use this strain for his own work once his post-doc is finished. It would then be immoral for John to join Dr. Sally's group. It might be that John could, after completing his post-doc, wrest control of a sample of the new strain from Dr. Sally by threatening a law-suit. If Dr. Sally thinks the cost of defending the suit is greater than the cost of breaking her policy, she might accede to an ultimatum from John. But John would be immoral in issuing it, since he knew when joining the group exactly what Dr. Sally's policy was.

The general rule of thumb here is this: before joining a group, a student or post-doc must do what she can do discover what the working practices of the group are; if she reasonably thinks she *cannot* accommodate herself to those practices, and they are otherwise morally permissible, she should not join the group. Having joined a group, she has a deontic obligation to the PI to abide by the procedures, provided doing so is otherwise morally permissible, or voluntarily leave the research group. This does not mean that advice, comments or complaints, solicited or not, cannot be directed to the PI. It does mean that the student or post-doc ought to abide by the rules, everything else being equal. That obligation, however, can be overridden. Special considerations apply if the practices of the research group turn out to be immoral in their own right. The student must then make a decision, often a quite difficult decision, about whether or not she will comply and whether or not she will take action to change the lab practices. This is an especially difficult topic, and we will treat it under the topic of *whistleblowing*.

The converse obligations imposed by the power imbalance on PIs are much greater in scope, and often tricky to diagnose. Again, the crucial intrinsic values here will be respect for autonomy and paternalistic beneficence. This means that the PI, in reaching decisions that affect her students, post-docs and lab workers, must consider *their* interests as well as her own. And sometimes that will mean sacrificing *her own* best interests in order to further the interests of those whom she mentors. To fail in this respect would be to use others as mere means, rather than as both means and ends in themselves. Further, one cannot read off from the fact that a student or post-doc has chosen a given course of action that acting in accordance with that choice does respect the student's autonomy:

autonomy is exercised by choosing *freely*, i.e. without coercion. Because of the deep power imbalance in most research settings, the choices of students and post-docs are often *not* free choices.

To illustrate, consider some examples. If a student and his major professor form a sexual or emotional relationship, it may be (and often is) the case that *both* parties in fact prefer to engage in the relationship rather than not. Indeed, it may be a case of ‘true love’, with deep and lifelong importance for both participants. But the student’s choice is not, cannot be, free. Whatever reasons he has for entering the relationship, they are not independent of the context in which that relationship occurs. And that context makes certain choices, otherwise open to the student, prohibitively costly. So much is itself not enough to show that such liaisons are immoral; it is enough to show that the choice to engage in such a relationship, on the part of the student, is *not* an exercise of autonomy.

Suppose an advisor recommends that a student apply to post-doc positions at the universities of Nebraska and Montana, with respectable programs in theoretical population ecology, but not to Harvard or Princeton. The advisor reasons that the student is not quite capable of performing to the standards imposed at the latter institutions, but will exceed those imposed at the former. Further, in the advisor’s judgment, the student himself does not work well under pressure, especially as he judges that others think he is not working up to snuff. All of this suggests to the advisor that the student’s prospects for a happy and successful career are best served by an appointment at Nebraska or Montana, and would be seriously compromised by an appointment at Harvard or Princeton. She conveys all this to the student by way of justifying her decision to write letters for him to Nebraska and Montana but not to Harvard or Princeton. On reflection, the student endorses her recommendations, i.e. he takes it that the advisor has in fact offered very good advice. Given this, it may seem obvious that the student’s decision not to apply to Harvard or Princeton is autonomous. But that is not so. The decision may be wise, prudent and even correct. So too, the advisor’s behavior may be wise, may in fact serve the student’s best interests, and may even be morally correct. But the student’s decision is not, and cannot be, autonomous. He knows that an application cannot be successful without a strong letter from his advisor, and given what the advisor says to him, he knows that any application to Princeton or Harvard is doomed. His choice not to apply is, therefore, coerced.

The lesson here is that judgments about whether a choice is autonomous, or whether an action respects autonomy, do not concern the *actual* motivations or desires of the agents in question, but rather their *counter-factual* behaviors: had the context been different, had it been free from the power imbalance, would the agent have chosen otherwise, or for different reasons? If so, the decision is not autonomous. Scientists in positions of power must choose policies for regulating behavior in a lab or group, and must treat individual students and post-docs, in a way that best reconciles two distinct and often conflicting values. Those in positions of power must offer as much latitude for *autonomous* choices as possible, and respect those choices when offered, while at the same time acting for the best interests of the agents *precisely because those agents are not in a position to make autonomous choices well*. The balance is delicate, and often quite subtle differences in

behavior can make all the difference between moral propriety and serious violations of trust.

A resolution of the conflict between the value of autonomy, and the importance of paternalism when otherwise autonomous agents are not in a position to decide well, cannot be achieved by any one rule. But there are some good rules of thumb to guide considerations about which actions best serve the relevant intrinsic values. First, certain failures to respect autonomy are absolutely intolerable. When power imbalances exist, it is an infringement on autonomy to engage in any social behavior that would not otherwise be tolerated, including for example sexual behavior or joking or teasing with discriminatory or derogatory content. The problem is not that the participants don't, or won't, all enjoy the exchange (this may or may not be true). The problem is that the exchange is not free, and *were the power imbalance absent*, all parties *would be* competent to make an informed, autonomous decision about whether or not to engage in the conduct. To impose such conduct in the work place is immoral, even if all parties consent, and even if all parties would consent were there no power imbalance. It is immoral even in these circumstances because such conduct fails to properly respect autonomy.

Second, when those without formal power *are* competent to make decisions that primarily affect their own interests, it is improper not to permit and respect those choices, even if those with power would, for whatever reasons, have them choose otherwise. So, for example, if a student can see why a choice is bad, even disastrous, for his career (say applying to Princeton rather than Nebraska) but wishes to pursue that choice rather than any more prudent course of action, he ought to be permitted that choice. Not to do so fails to properly respect autonomy.

It can be enormously difficult to differentiate between cases in which a student would be competent to make an autonomous decision were the power imbalance absent but a) is not competent to make such a decision when the power imbalance is present, and cases in which a student would be competent to make an autonomous decision were the power imbalance absent and a') is competent to make an autonomous decision even when the power imbalance exists. In particular, when a student would be competent without the power imbalance, and *also in that case would decide as he in fact does decide with the power imbalance*, it is natural to suppose the decision, even given the power imbalance, is autonomous. But this need not be so. If the power imbalance either substantively constrains the options open to the student, substantively changes his evaluation of the options, or substantively changes his perception of the decision context, the choice is arguably not autonomous. This is so even if it agrees with the choice he would have made absent the power imbalance. Though recognizing where and when those who lack power are none the less competent to make autonomous decisions is difficult, it is also crucial.

Third, we must consider what to do when those without power are not in a position to choose autonomously, and would not be in a position to do so absent a power imbalance. Typically, these situations arise when students or post-docs suffer from a lack of training

or information, and so lack the knowledge required to make an informed choice. Sometimes, this impairment cannot be rectified simply by imparting information—learning often requires sustained training. Here it is impossible for those lacking power to make an autonomous decision, with or without the power imbalance. Two cases then arise: the decision primarily affects the interests of the person without power, or it affects the interests of others in the lab as well. In the first case, those making the decision ought to make it in ways that best pursue the interests of the person on whose behalf the decision is made. This is the consequence of obligations stemming from paternalistic beneficence. In general, the more this person can be involved in the decision, the better. But in the end, those with the power to make the decision are obligated to use that power, and to do so in whatever way best serves the interests of the person for whom they are deciding.

In the second case, it is proper to consider, in consequentialist fashion, the interests of all in the lab, including the interests of those with power. Here, and only here, is it permissible for PI to override the preference of a graduate student or post-doc in order to pursue interests and aims which are not those of the student or post-doc.

We now work through two cases.

Case 1. Joseph and Sandra are both finishing doctoral students under Dr. Menendez. Joseph has designed and run a set of elegant experiments in nanotech machines, and has secured a position as a materials scientist in a large corporation as an experimental bench scientist. He is quite satisfied with this position, having little interest in theory and no interest in an academic career at all. These results themselves suggest a possibility to Sandra, which she has developed into a fundamental theoretical advance, but demonstrating the theory depends crucially on Joseph's results. She hopes to parley a paper reporting the advance into a position in theoretical physics at MIT. Dr. Menendez asks Joseph if he will publish his results in *Physics of the Solid State*, a good but not first rate journal mostly featuring experimental results, and allow Sandra to use his results in a *Nature* paper, on which he will not appear as an author. Sandra's paper will appear before Joseph's, and she will garner much of the credit as a result—perhaps enough to capture the MIT job. Joseph agrees, since he has no interest or need for the enhanced reputation he could achieve by insisting on co-author status on Sandra's paper. Given Joseph's agreement, is Dr. Menendez' plan permissible?

Analysis. No. Were Joseph's results already in the literature, it would be perfectly reasonable for Sandra to publish and cite those results as supporting evidence. But they are not already in the literature, and so she may not legitimately do that without Joseph's consent. Joseph does consent of course, and indeed arguably would consent even if he were in no way dependent on Dr. Menendez' good will. But he is in fact so dependent—if Dr. Menendez so chooses, he could deny or delay Joseph's defense so as to compromise his standing offer at the large corporation. Hence Joseph's decision is *not* autonomous. In these circumstances, Dr. Menendez is obligated to consider Joseph's best interests, even if that is to the detriment of his lab and other students in it. Joseph's



best interests are served by publishing first, and in as recognizable a journal as possible, or minimally co-publishing with Sandra. Hence Dr. Menendez' plan is not permissible.

Case 2. Brian, a second year graduate student, is choosing his dissertation advisor. He wants very much to investigate a mathematical model of group selection, and approaches Dr. Eienenheit with a suggestion to this effect. Though there are other theoretical population geneticists in the department, Dr. Eienenheit has much the best placement record, and Brian very much wishes to pursue an academic career in mathematical biology after defending. Dr. Eienenheit takes a dim view of Brian's proposal, on the grounds that even if it is successful, it is professional suicide to work seriously on group selection before tenure. He instead suggests an alternative project focusing on the integration of co-variation between demographic structure and fitness into population models. While this is fundamentally different from Brian's favored project, it will allow him to develop some of the modeling skills and mathematical expertise necessary for addressing the group selection questions that grip Brian, so he will be well prepared to take them on after he secures an academic position and earns tenure. Is it permissible for Dr. Eienenheit to refuse to direct Brian's preferred project?

Analysis. Yes. There are several considerations here. First, Brian is not yet a member of Dr. Eienenheit's research group, and so there is no power imbalance in virtue of which Brian is incompetent to make an autonomous decision. Since there are other scientists capable of directing Brian's preferred project, he has real choices. If Brian is not competent to make this decision, then, that can only be because Brian does not really recognize the consequences of his decision with respect to his future career options. So, if Brian's decision here is not autonomous that is because he is not competent to make the decision, and it is exactly in such cases that paternalistic beneficence is warranted, in which case Dr. Eienenheit is acting in a permissible way. On the other hand, if Brian's decision here *is* autonomous, Dr. Eienenheit's refusal does not fail to respect that autonomy, since Brian has real alternatives. Either way, Dr. Eienenheit is within his rights to refuse.

#### Study Problems.

Case 1. Well into the second year of work on her dissertation project, Amy and her advisor, Dr. Snuff, discover that the experimental design, aimed to test a theory of covalent bonding, is fundamentally flawed. Dr. Snuff recommends that Amy begin anew, with a better experimental design. Amy very much wants to finish as quickly as possible for personal reasons. She thinks that her current results, though negative, will provide the basis for a defensible thesis, and that with doctorate in hand she will be able to secure some kind of research post-doc. Her post-doctorate work, she reasons, will be the crucial factor in determining whether or not she can find an academic position, so it is enough now just to have finished the dissertation. Dr. Snuff disagrees: the work so far might make for a dissertation which is of marginally passable quality, but it will certainly not enable Amy to secure a post-doc position in which she can hope to do the kind of work which will enable her to pursue a full-time academic career. He insists that Amy

undertake a new line of experimental work. Is Dr. Snuff's insistence here permissible, or should he allow Amy to defend her thesis using the work she has already done?

Case 2. Ivan left Dr. Pavlovski's research group in his fourth year to take a very well paid job in industry. He had at that point collected 3 years worth of data as part of a five year longitudinal study central to Dr. Pavlovski's research. Dr. Pavlovski then assigned Svetlana to finish the study. Though Pavlovski and Ivan parted on very bad terms, Pavlovski offers to include Ivan as an author on the papers Svetlana and Pavlovski produce when the study is finished. Ivan, however, turns the offer down, and in fact claims that the first three years of data are his, and withholds permission for their use by Pavlovski in any respect. Svetlana and Pavlovski decide to publish anyway, crediting Ivan's contribution in a footnote. Is this permissible?

Case 3. Hartry has developed a statistical algorithm for analyzing time-series data as part of his dissertation work. The algorithm itself is unpublished, and in the processes of being patented. Dr. Ngo, Hartry's supervisor, and no other member of the research group, appear on the patent, since Hartry's work on the algorithm itself was done while he was supported by a University Dissertation Fellowship, before Dr. Ngo agreed to direct Hartry's dissertation work. However, various members of the group have copies of the algorithm, coded in C++. Jason has used the algorithm to analyze his data, generated by a randomized study of the performance of matched students from school districts which have a voucher system for private schools and those which don't. The analysis suggests that students from school districts with vouchers outperform students from districts which do not, and that this is true *whether or not* the students attend public or private secondary schools. Before publishing, Jason seeks Hartry's permission to use the algorithm. Hartry is morally opposed to voucher systems, and denies permission. Jason asks Dr. Ngo's advice about whether or not he may publish using the algorithm in any case. What should Dr. Ngo say?