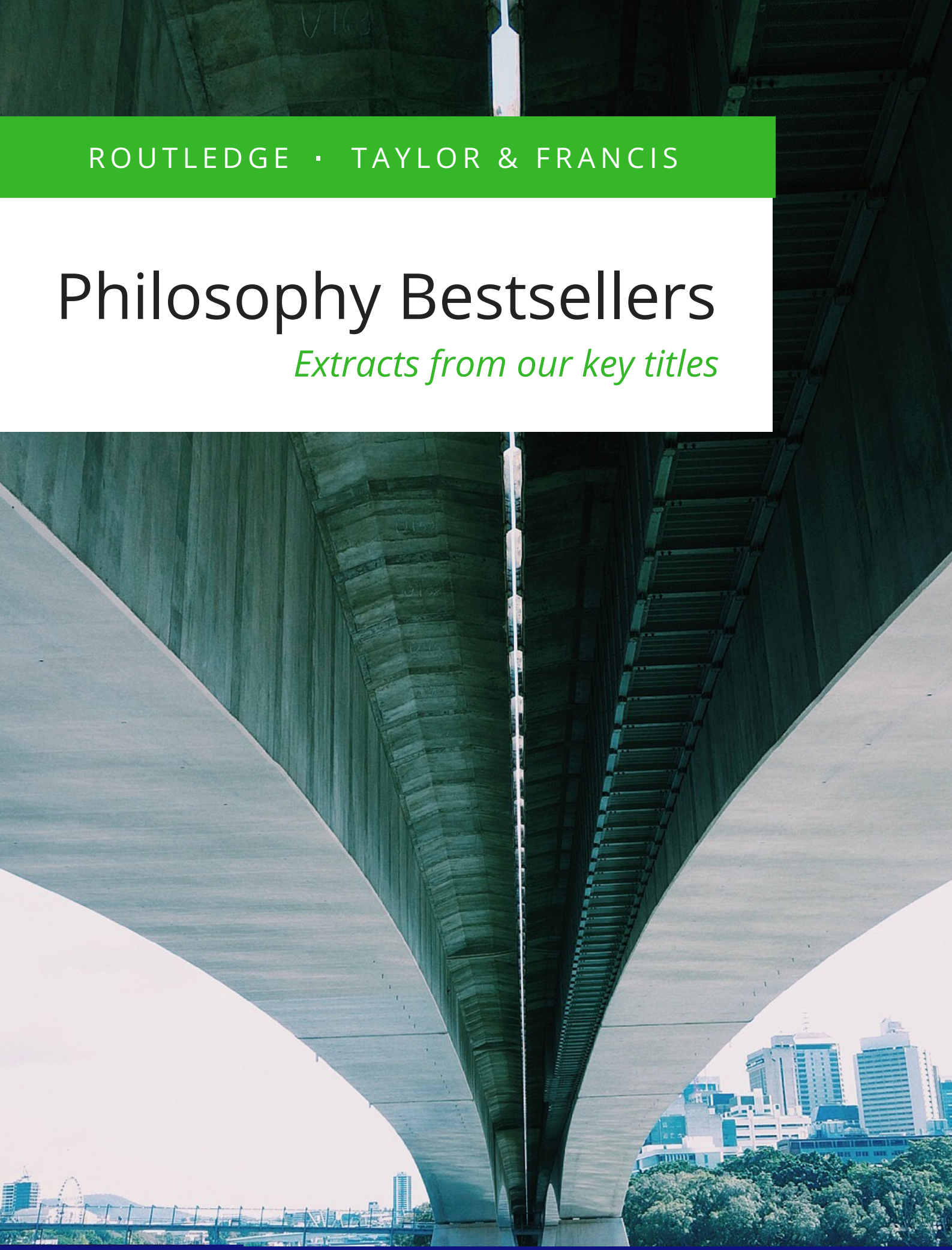
ROUTLEDGE · TAYLOR & FRANCIS

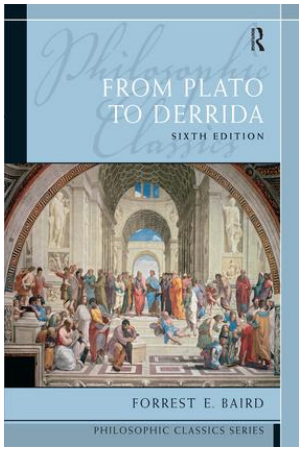
Philosophy Bestsellers

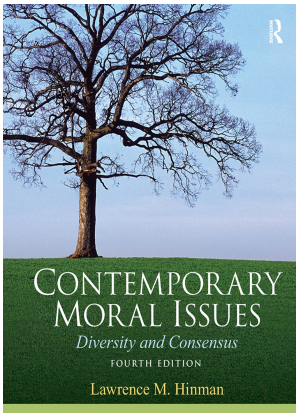
Extracts from our key titles

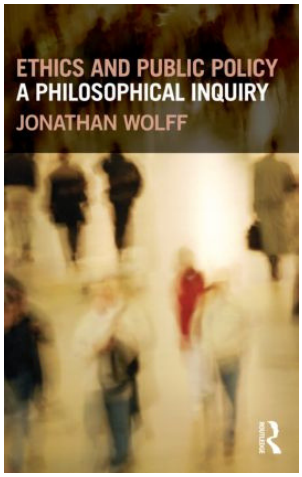


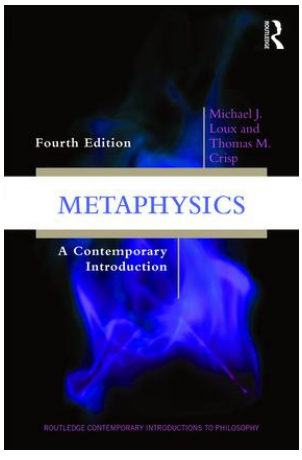
Routledge
Taylor & Francis Group

www.routledge.com/philosophy









Contents

Cloning and Reproductive Technologies

The Narratives

READING

Katy Duke

"Belgian Loophole Allows Swiss Parents a 'Savior' Baby"

In this article, published in the British medical journal *The Lancet* in 2006, the author describes how a Swiss couple travelled to Belgium to get the tests necessary for selecting an embryo that could provide their sick six-year-old son with life-saving umbilical cord blood stem cells. In Switzerland, such pre-implantation diagnosis is illegal, but it is permitted in Belgium.

As You Read, Consider This:

1. What role does pre-implantation genetic diagnosis (PGD) play in the creation of "savior babies?"
2. Why would this technique not be useful for the creation of "savior babies" to help friends?
3. This case describes a savior baby, conceived to save the life of a six-year-old sibling. How would it affect your moral assessment of the situation if the baby was conceived in order to save the mother's life? The father's life?

A successful bone-marrow transplant involving a baby born to save the life of a sick sibling has sparked a heated debate on medical ethics in Switzerland.

Swiss 1-year-old Elodie H, not fully named for legal reasons, was born from an embryo selected from a group by pre-implantation genetic diagnosis (PGD) as a compatible donor for her big brother Noah.

6-year-old Noah suffers from chronic granulomatous disease, which compromises the immune system, cutting life expectancy in half.

Parents Beatrice and Yves H from Geneva decided to use PGD to conceive what has been dubbed a "savior sibling," whose stem-cell-rich umbilical cord blood could save Noah's life after all attempts to find a matching bone marrow donor have failed.

Karen Sermon, from the Centre for Medical Genetics in Brussels, where the procedure was eventually done, says Noah's parents had to come to Belgium to save their son because PGD is currently banned in Switzerland.

She explains: "In Belgium there is nothing written down saying it is legal but the current law on embryo research which bans a number of procedures—such as reproductive cloning—does not ban PGD, so therefore we are allowed to perform the procedure."

The current blurred legal situation in Belgium is mirrored in several countries across Europe where PGD legislation is still being written, so the procedure is often neither officially sanctioned nor illegal.

Only Denmark, Spain, France, Norway, and Sweden have specific laws allowing PGD, though it is not illegal in Belgium, Finland, Greece, the Netherlands, and the UK. It is banned in Switzerland, Germany, Austria, Italy, and Ireland.

In Germany, which is considered to have restrictive research laws, a debate over whether or not PGD is already banned under the 1991 Embryo Protection Law is underway. However, despite a lack of consensus on the legal status, it is not currently practiced.

This lack of legal clarity in Europe is partly because PGD is a comparatively new science. It was originally developed to prevent the transmission of serious genetic disorders by screening embryos during *in vitro* fertilization (IVF). Newer uses of PGD include detection of mutations for susceptibility to cancer and for late-onset disorders such as Alzheimer's disease. It can also be used to check for histocompatibility to create a baby who is not only clear of a disease affecting an older sibling but is also a viable match as a donor.

Katy Duke, "Belgian loophole allows Swiss parents a "savior" baby," *The Lancet* www.thelancet.com Vol 368 July 29, 2006.

Due to the novelty and technical difficulty of the procedure, Elodie is only the 12th baby in the world to be successfully conceived for the purposes of saving the life of a sibling.

However, because of Noah's age and Elodie's small size at birth there was not enough blood in Elodie's umbilical cord to provide the stem cells needed so the siblings underwent a bone-marrow transplant.

Sermon says PGD is normally refused for couples where the ill child is already too old to be treated with cord blood cells, to save the savior sibling the pain of undergoing a surgical intervention. "We expected that cord blood cells alone could save Noah," she says, "but at Elodie's birth too few stem cells were found in the umbilical cord blood. But I would also like to stress that bone marrow transplants between existing siblings is also common practice."

Elodie's bone-marrow transplant was successfully completed in Zurich Children's Hospital earlier this year.

Reinhard Seger, who did the operation, says Noah's immune system is gradually building up and only the number of his T lymphocytes is still insufficient. For the past 2 months, Noah has been able to live at home without the need for total isolation and the continuous antibiotic treatments he used to need.

But opponents of the procedure say despite the obvious benefits, creating a baby so that it can undergo invasive surgery that is of no help to it individually is unethical. In the UK, the Human Fertilsation and Embryology Authority (HFEA), the regulatory body that grants licenses for any potential use of PGD, says imposing risks without benefits in exchange is not permitted. It has a set of stringent criteria that must be met before it will grant a license for a savior sibling to be born.

Joyce Harper, from the European Society of Human Reproduction and Embryology, says getting a PGD license in the UK is complicated, but the laws are considerably more liberal than most other countries.

She says: "Britain is highly regulated but at least it is possible to do PGD. In some countries the use is very limited. I feel that PGD should only be used to prevent disease—and in a way this is still preventing disease. I do not understand countries where an abortion of a normal fetus is legal, but the use of PGD to prevent a sick child is illegal."

In Noah and Elodie's home country of Switzerland, abortion was legalized in 2002 and recently hospitals, living rooms, and government buildings have turned into debating rooms discussing whether the same should happen for PGD. The media has devoted so much time to the topic the young pair are now recognizable faces across the country.

Christoph Rehmann-Sutter from the Swiss National Advisory Commission on Biomedical Ethics says he believes the Swiss ban is morally correct, despite the emotive argument for Elodie's birth. "With their presentation of Elodie and Noah, the media have given a face to the abstract concept of PGD. One can identify with the suffering of the boy and pain of the parents and through that one can understand how such a plan must appear morally justifiable."

Looking at it from such a perspective it is obvious that it is impossible for parents to say no to such a chance and it is also clear that they will love the 'savior child' for herself and accept Elodie completely as a daughter and a full member of the family.

"But next to this individual perspective there is also a societal perspective—one cannot base laws on one happy result."

Rehmann-Sutter believes PGD is necessary to screen for serious genetic diseases and is preferable to late-term abortion after a disease is uncovered by amniocentesis. But, he says, he worries that using it for cases like Noah's opens up a Pandora's box for the future. "We must consider the development of a moral pressure on parents. Do we want a society where one feels impelled for therapeutic purposes to have a new child every time?

"And if PGD is used to benefit anyone other than the embryo itself, it will be difficult to draw the line at cases like Elodie and Noah. Why should the child only help a sibling? Why not the mother, the father, a close family friend? Why only with bone marrow? Why not a piece of the liver? A kidney? How much is the new life expected to give up?"

Many other opponents to the procedure cite worries of a slippery slope, a quiet slide into positive eugenics, where parents choose babies with certain physical or character traits based on analysis of embryonic genes.

But Mohammed Taranissi, the Director of the Assisted Reproduction and Gynaecology Centre, a prominent clinic that uses PGD to create savior siblings in the UK, thinks the "slippery slope" worry is illogical and that the old arguments against savior siblings are losing ground as people learn the medical facts behind the procedure.

Taranissi says: "Once you explain the truth of what is happening most people accept it, as most objections are based on a misunderstanding of the facts. These are not 'designer babies'—we cannot manipulate them to

have certain characteristics like choose their hair color, it's simply not technically possible and won't be in our lifetime, if ever. And that's not even slightly what we are trying to do here. This uses normal and fully accepted IVF procedures to help young children with medical conditions."

"And the idea that it's a slippery slope and soon babies could be born to help sick parents or family or friends is illogical as the procedure only works when treating a young child. Basically there are only enough stem cells in cord blood to treat a certain size of child—which is why we are always pressuring the HFEA to speed up their evaluation procedure as often an entire year can be wasted in preparation. During this time, the sick child keeps growing and there is less of a chance to help him or her."

He adds: "It is ethically better to create a so-called 'savior sibling' than to use an existing sibling. This way all the necessary materials can be taken from the umbilical cord—something that is going to be thrown away anyway and avoids intrusive bone-marrow transplants that are the only option when a donor sibling is identified after birth.

"In the end the only argument is if your child was sick, would you do it? And I think we all know the answer to that."

Panel: **Criteria for the award of "savior sibling" licenses in the UK:**

- The condition of the affected child must be serious or life threatening.
- The embryos themselves must be at risk from the condition by which the existing child is affected.
- All other possibilities of treatment and sources of tissue for the affected child should have been explored.
- The technique should not be available where the intended recipient is a parent.
- Only cord blood should be taken.
- Appropriate counseling is a requirement.
- Families are encouraged to participate in follow-up studies, and clinics are required to report the treatment cycles and outcomes.
- Embryos may not be genetically modified to provide a tissue match.

Journal/Discussion Questions

1. The article seems to presume that everyone would make the same choice as this Swiss couple. Do you think this is true? Would you make this same choice?
2. When confronted with legal restrictions on possible medical treatments and procedures, it is increasingly common for people to go to other countries to obtain such procedures. This practice, sometimes called "medical tourism," allows them to circumvent their own country's laws. Should such practices be allowed? If not, in what way could they be restricted?

Eva Feder Kittay with Leo Kittay

On the Expressivity and Ethics Selective Abortion for Disability: Conversations with My Son[1]

About the Authors: Eva Feder Kittay is a philosopher of language, who has written extensively about metaphor and meaning. The context of the present dialogue emerges in the course of the writing, but suffice it to say that this is a dialogue between Professor Kittay and her son, Leo, who graduated from Princeton with a degree in philosophy. Sesha Kittay is Professor Kittay's daughter and Leo Kittay's sister.

About the Article: In this article, Professor Kittay and her son Leo discuss some of the profound ethical issues that surround children with disabilities. This article was occasioned by Professor Kittay's invitation to participate in a conference on the use of selective abortions in the context of prenatal testing. Professor Kittay speaks of Sesha as a mother, and Leo does so as a brother. Both address the range of questions arising out of their family experience as philosophers, as family members, and as sensitive, thoughtful human beings. Although this is primarily a discussion focused on the issue of disability, it sheds much light on larger questions about the meaning of abortion within the context of individual lives, especially the context of lives and families. It also contains an illuminating discussion of what such choices "mean" in a social context—a topic that Professor Kittay refers to as the expressivity of such choices.

As You Read, Consider This:

1. Kittay is concerned here with the "expressivity" of the choice of selective abortion for disability, that is, she is concerned with examining the messages such an act conveys to others, including family members. Is there a difference between what she feels such an act would express and what her son feels it would express? Discuss.
2. In one of his letters, Leo comments that the family seems to be more like a club than a family. Explain what he means by this. Why is this significant in regard to children with disabilities? Is it significant for those without disabilities? Discuss.

My daughter, Sesha, now twenty-seven years old, lives at home with us. It is sometimes easiest to describe her in the negative, what she is not and *does not do,* for these are the well-defined capacities: she doesn't talk, she walks only with assistance, she is not fully toilet trained, she can't feed herself, and so on. But what she is is so much more. She is a beautiful young woman with a winning smile, an affectionate nature, and a love for music, water, food, and the joys of physical affection. I had never before written of her or our relationship and had not used my knowledge of living with a disabled person directly in the service of my professional writing.

I was about to undertake the first of such writings when I was invited by The Hastings Center to participate in a project on prenatal testing for genetic disability. I had wanted my first forays into writing philosophically about my daughter to be about her and what her life means and has meant to me. I had to be persuaded to join the project, for it meant that instead I would have to reflect on the hypothetical of her nonexistence, and worse still, of the hypothetical of having had to choose whether or not she was to come into the world.

During the course of the project, I was asked to consider whether selective abortion for disability "sends a message" that devalues the life of the disabled. When some initial discussions on the question revealed differences between me and my twenty-one-year-old son, Leo, who has an undergraduate degree in philosophy, I chose to write a chapter by conducting a dialogue with my son. We carried on our conversation through e-mail over a period of a few months while my son was working at a ski resort. This article records our dialogue.

From My Diary[2]

I want to get some thoughts on paper before the intensity of this, the first of the four Hastings meetings, evaporates.

10/22. Reflecting now on one participant's memory of when her pediatrician told her that he didn't know if her underweight baby would be all right, and her recalling this as the most terrible moment in her life, I thought what I would answer had someone asked me, "was the moment you learned that Sesha was retarded the most terrible moment in your life." I would have answered, "No." The most terrible moment in my life was when I thought Sesha would die. The next most terrible moment was when my mother insisted (or tried to insist) that Sesha be institutionalized and that I give her up.

I was asked if my mother has changed her attitude toward Sesha, I said, "Yes," and said that had happened because she initially thought that keeping Sesha would ruin my life. She's seen that it hasn't ruined my life.

10/23. Wednesday morning I awoke feeling sore internally, somewhat nauseated, somewhat as if I were recovering from a physical torture. Tuesday evening, as I tried to cram the articles on psychological experiments on metaphors and idioms into my head for the upcoming class I teach, a dam burst and the floodgates came

undone. I sobbed deep, deep sobs from the interior of my soul. I cried, I cried for Sesha. I felt the hurt for her impairments, for the profound limits of the life she could experience, for the multiple aspects of life she could never know or even know that she couldn't know. I wept for Sesha—not for me, not for Jeffrey [my husband], not for Leo, but for her, her sweetness, her limitation, and the pain of knowing what a small aspect of human life she could inhabit. She, my daughter—the child I had brought into the world and the child I had raised and worked to nourish and protect. It is a hurt that doesn't dare to be felt, almost all the time, and it is a hurt that cannot be felt in her sunny presence. But it is there and at moments like post-Hastings it floods in.

Now what is this mysterious pain? Mysterious because who is hurt? I don't think Sesha is aware of her limitation. It is not like the sorrow for another's pain, because pain is felt by its bearer. So do I cry for myself and my expectations of the child I wanted to raise? That is not what those sobs were about—I know. In debriefing Jeffrey and Leo on Tuesday, I spoke of the question posed at the conference: Whether aborting after learning that the fetus is impaired sends the message that a disabled life is not valuable. I asked Leo: If I had aborted a fetus based on disability, would it have sent the same message to him as would the message he'd receive if we had institutionalized Sesha. He said, "No." It wouldn't have sent the same message, but he did think that the message of an abortion would have been that the disabled shouldn't exist. I asked him, "Even in the face of Sesha and our life with her?" He answered that it still would, although it wouldn't be as strong as the message would be if I had aborted an impaired fetus in the circumstance that Sesha was not part of our life. This surprised both me and Jeffrey. But it is information I must take seriously.

Dear Leo/Dear Mom

EFK's Letter #1

Dearest Leo,

I've been asked to address the question of expressivity of a woman's decision to abort a fetus that has been diagnosed with a disability following amniocentesis or other prenatal testing. The question of expressivity is the question of whether such a decision signals the devaluation of the life of a person with disabilities. I thought I would send you excerpts of the letter I sent to Erik Parens when he first invited me to participate in The Hastings Center project. Tell me what you think.

Much love,

Mom

Dear Dr. Parens:

I have a severely retarded daughter who also has cerebral palsy. As much as I value my daughter—she together with my well son, constitute the single greatest joy in my life—I do not agree with the negative appraisal of prenatal testing that you say has been articulated by some members of the disability community.[3] I believe that our society does not provide the conditions that make raising and caring for a severely handicapped child, while otherwise living a full and fulfilling life, possible for most parents, and I am skeptical about the possibilities of any society reaching such an ideal state in the foreseeable future. To undertake to care for a child with severe disabilities has been a difficult and painful course, and yet to abandon such a child to the care of strangers was and continues to be, for both me and my husband, unthinkable. We have garnered tremendous joy and learned more than one can imagine from our daughter, and yet the decision to have a child with such severe and multiple handicaps is not one I could easily endorse. I think it is terribly cruel to burden a couple with the responsibility for a severely handicapped child when prenatal testing can determine in advance the condition of the fetus. Furthermore, as a feminist, I must underscore that the responsibilities normally fall to the mother, as fathers not uncommonly abandon the family with the advent of a severely disabled child, and in most instances the mother provides the daily care.

On the other hand, prenatal testing does not eliminate the tragedy of a child who is severely impaired, for I can also envision the agony of making a decision either to abort or not to abort. Rather, the testing shifts the tragic moment and the decisions to be made. But if the decision is made to go to term, even in the face of the impairment, at least there is a conscious choice, made with the possibility of a truly informed understanding of what such a decision requires of the future parents. Moreover, the availability of the technology to avert some of the consequences of genetic disabilities does not absolve the society at large of mitigating the difficulties of raising and caring for disabled persons. In any case, such is the line that I would take based on my own personal reflections.

Sincerely,

Eva Kittay

Leo's Letter #1

Mom,

If we are to take the position that giving birth to a retarded child should be a choice in years to come, that is to say, that all parents will have their fetuses tested, and that the only fetuses to reach full gestation will be those whose parents have expressly chosen to raise them, then we must also be sure that some other changes are made. To begin with, it must be made public that raising a retarded child is equally, albeit differently, fulfilling than raising a normal child. If it ever feels more fulfilling, Mom, it is probably because we just expect it to be less so. Without such increased exposure to those different joys, sheer ignorance will cause the retarded population to become extinct. Anyone with the option would decide to abort a disabled fetus because they would not be able to imagine that the incredible burdens of raising a retarded child could be outweighed by the joys.

Why is this bad? Actually this is a hard argument to make. Social Darwinists might say that this is fine. This is a sort of "preemptive" survival of the fittest argument. However, a survival of the fittest argument is applied to fetuses or children only with difficulty. All babies are weak, and they tax parents and society. It might seem that we would all do better for ourselves in a world without the dependent young, but we all know such a world would be short-lived, if not absurd. Even Social Darwinists must take into account the dependents. They could argue that eventually "normal" children will grow big and strong. But, while many disabled children do not become strong and independent, some "normal" children do not either. This leads me to my next point. The argument you're making draws a major line between normal and retarded children, based on the difficulty of bringing them up. But beware the slippery slope, Mom. Are not all children a burden? If, someday, we could determine that a fetus will develop into a hyperactive child, or into one with recurring ear infections, will these children's births also have to be expressly willed? Children are a burden. But it is incredibly important to keep making them and tolerating them. No, Mom? No human child is fit for survival without the help of elders. To start drawing the line about how much help they should need is extremely problematic. Some groups of children will start vanishing. And we do not even want a Single species of animal to disappear. They are all intrinsically valuable. How do we show others how wonderful it can be to raise a retarded child, and how important and valuable her existence is? It is difficult! Especially if fewer retarded persons are being born.

What kind of message does aborting the retarded send to would-be siblings? I can only guess at this, Mom, because Sesha was born. But here are two different messages I could conceive of receiving:

1. The love my parents have for me is a condition of my being mentally and physically sound, not just of being a child of theirs. Rephrasing this: The only reason my parents want me is that I'm relatively smart and fit.
2. My parents chose me and therefore must really care about me. Again rephrasing: My parents wouldn't just love any child they might have, they love me because I possess the desirable properties or characteristics that make me who I am.

What I am trying to say is that the family starts to seem more like a club, and less like a family. In a club the members are selected based on one characteristic or another. This leads one to believe that if, for some reason, that characteristic is no longer attributable to the individual, or if anyone in the club comes to believe that this characteristic never applied, the membership in the group and the "love" that results can vanish. If a fraternity guy stops playing football well, he might be afraid he would not be wanted in the fraternity anymore. He was aware, after all, that his ability to play football allowed him entry into the club. If a child believes his membership in the family is contingent on not being retarded or otherwise disabled, he might at first value his place in it more highly because it was earned. (This goes hand in hand with the belief that those who are retarded or disabled are worth less. I think immediately of children who use the word "retard" on those they wish to insult and how this insult always seems to be underscored by the desire of the one doing the insulting to differentiate him- or herself from those who are retarded.) But the positive feeling that love has been earned can subside, and the child might instead feel a constant pressure to prove himself to be worthy of his place in the family. He will not view his family's love for him as unconditional love.

I hope these thoughts are helpful.

Love, Leo

EFK's Letter #2

Dearest Leo,

You raise many points, each of which is crucial and each of which I want to explore. My assignment was to consider the expressivity of prenatal testing with respect to disabilities generally, but you speak primarily of retardation. I will move from the one case to the next with some fluidity, although I will try to address the larger perspective of disability.

Parenthetically, let me say that in reading your letter, I realized why I have tended to speak of Sesha as "handicapped" rather than "disabled," a designation for which I was called to task at the first meeting of The Hastings Center project on prenatal testing. Sesha's disabilities are so severe that in speaking of my child as disabled I think that I will be failing to communicate the particular condition that is Sesha's. That is, to speak of her as disabled puts her in the same category with persons with relatively mild disabilities, disabilities that do not prevent them from leading very independent and productive lives. By whatever our standards for independence and productivity, Sesha doesn't now and never will meet those. I feel more comfortable speaking of Sesha as "seriously disabled," although someone like Stephen Hawking is, by any standards, seriously disabled, and again by any standards is productive, though not independent. I recognize, of course, that "independence" too is a slippery notion because in some very important respects no one is independent.

The Americans with Disabilities Act speaks of a disability as "a physical or mental impairment that substantially limits one or more of the major life activities." Disability activists speak of handicaps as the *consequence* of a disability, where environment limits an individual with a given disability.[4] Many of these handicaps can be overcome with social interventions and modifications of the environment. It takes a social and political will to structure the environment so that it responds to the needs of those who are disabled. But Sesha's condition is such that most of her disabilities remain handicaps even with environmental modifications. In Sesha's case handicap and disability are nearly coincident. I think the distinction is nonetheless a very important one and speaks to the importance of how we express ourselves with respect to disability. I want to make it clear at the outset, that nothing I have to say is meant to deny the importance of how we speak with respect to the disabled. I think that Michael Berube is correct when he speaks of how representations of the disabled figure in their treatment and life prospects.[5]

Let me now summarize what I think are your main points. I will then address them.

First, you speak of the need to expose people to the joys and fulfillment of raising a child who is retarded (or severely disabled). Second, you address the possibility that with advances in prenatal testing all cases of retardation (and other serious disabilities) will be eliminated. You then ask us to consider what would be lost if we no longer had persons with mental retardation. Third, you point to the problem of arguing from the difficulties and burdens imposed in raising a seriously disabled child, and the slippery-slope problems connected to such a position. Finally, you address the message that the sibling gets if the family chooses to abort a fetus diagnosed with a disability.

I will start with the second point, the speculation that with advances in prenatal testing all cases of retardation (and other serious disabilities) will be eliminated. Most cases of retardation are not genetically based. Most cases of retardation result from something going wrong during the pregnancy itself or immediately after birth. This was probably the case with Sesha. Such cases could not be picked up in prenatal testing, which depends on examining genetic material. Even if all retardation or other disability were picked up prenatally, there would still be problems that occur during birth and immediately after birth. Then, of course, there is disease and trauma that leave children (and adults) disabled (and sometimes mentally retarded)—some of the children in in Sesha's early intervention program, for example, were casualties of car accidents or gunshot wounds. However, certain populations, such as those with Down syndrome and spina bifida, are likely to be diminished by selective abortion following prenatal testing.

Among fetuses diagnosed with Down syndrome upon results from amniocentesis, it has been said that 90 percent are aborted.[6] However, that figure has to be looked at more carefully. Prenatal testing, although increasingly available, is not available to large numbers of women—and even when it is available, not all women avail themselves of it. So when we get the 90 percent figure, we get the percentage of women who have taken the test, often having already concluded that they would abort if the results are that the child has a severe disability. The literature indicates that the reasons for aborting—whether or not they are well informed—are various and complicated. But we can discuss that later. My point now is only that I do not believe we will ever see a world without persons with disability, without serious disability, without mental retardation, or even without persons with Down syndrome or spina bifida—although there may well be fewer of the latter individuals, and we can certainly talk about the extent to which this would be undesirable.

Now you go on to ask what would be lost if it became the case that populations of significantly retarded persons and others with serious disabilities would be eliminated or significantly diminished through selective abortion. Let us confine this question to the case we know well, mental retardation. And qualify that case to include not all mental retardation, just all mental retardation that occurred before or even at the moment of birth. Well, I agree that the world would be a poorer place without persons with Down syndrome or other sources of retardation, without people like Jamie Berube, or Sesha.[7] Our household has been immeasurably enriched by Sesha. People like Jamie, Sesha, or Abbie [our neighbor's little girl], force us to think much more profoundly about what it is to be human, what our obligations are to others, why we have these obligations, what the source of human joy and human sorrow is. I haven't begun to plumb the depths of these questions with respect to Sesha, but they are my measure of the truth, and the value, of all philosophical theories. If they cannot include Sesha in their universe, they are at best incomplete, at worst faulty. And that is not because Sesha is so different from us, or even because she is so much like us, but that at the very core, we are so much like her.[8] We understand so much more about who we are and what moves us, when we see what moves Sesha. I understand so much more of what it is to be a parent and love a child like you, when I know what it is to love Sesha. (But, of course, there are also limits to that. If I kissed you as much as I kissed Sesha, you would have been gone from this house much sooner!)

And yes, Berube is right. Sesha's value, like Jamie's value, is not in what they teach us. They are of value in and of themselves, in the same way that you are of value in and of yourself. Perhaps the world is always dimin-

ished when that which has value in and of itself, intrinsic value, is lost to the world. Perhaps Wittgenstein was wrong when he wrote, "The world is the totality of facts."[9] Perhaps he ought to have said, "The world is the totality of intrinsic value." Because our world—our lived world—does not consist of facts, but of our understanding of facts and the value those facts have in our life.

Now, if I choose to abort a fetus that would grow into a child with disabilities, have I diminished the world? That is a painfully difficult question. Yet I cannot see that it is necessarily a very different question from the question: "If I abort have I diminished the world?" For to abort any fetus will be to abort a being that would have intrinsic value in this world. Perhaps I deviate from some feminists in thinking that a decision to abort is itself a difficult one; often, psychologically painful and ethically problematic (not necessarily wrong, but not easily right.) Yet I do not deviate from the feminist position in believing that the moral choice must be the woman's to make. I remain convinced that the same must be said, though perhaps with more poignancy (and you will be right to ask why), when the phrase "with disability" is inserted. Let me end this letter for now. Perhaps you want to reply before I move on to the other points. Hope the skiing continues to be wonderful.

All my love, Mom

Leo's Letter #2

In response, Mom:

I think that your summary doesn't highlight my point that it is only through exposure to actual retarded people that anyone can really appreciate how much they contribute to our lives, at least with the most severely "disabled." (Here I don't include Stephen Hawking because we can measure at least some of his value on the same scale that we measure the abled.) Not all disabled people are wonderful, though, and we would not want to put together an argument that is based on that premise. (Berube makes this point, his most memorable in my opinion. The story of *My Left Foot* also contains wonderful examples of this, when the disabled protagonist is often less than charming.) We could argue that severely disabled persons are pivotal to our world because they too add to it. Yet evil people, boring people, everyone adds to a culture in some way, and probably in some positive way to boot. No one tries to suggest that we attempt to abort these groups. You are right though, Mom, when you say we will never have the choice, or at least not in the conceivable future, of whether or not to *allow* disabled people to exist. So I don't know if this is an interesting line of argument at all. Two possibilities might result if there were fewer retarded persons. The first is that fewer people will know what joys can come from being around someone like Sesha. More ignorance, and this is never a good thing. But here is another way of looking at it: a smaller population of this minority would be less threatening socially, politically, and economically (like the single African American child in an all-white school, or the sole Jewish family in a town of Christians). Sometimes it is easier for a minority to prosper under these conditions But I tend to buy the first result more readily. The second feels too artificial and, in circumstances like these, the danger of stereotyping remains substantial. You point out that I would want to consider why you say that aborting a disabled child is more poignant. I think it is because it feels, for a moment, to be an easier question than that of aborting a normal child. We fear that it will not get the same weight as the other question; that our system of values weighs the death, or (sorry) the lack of life, of one as less meaningful than the other. I think the way we will have to argue this point is to equate the value of the disabled with that of the normal. Maybe this is obvious to you, but it is not to me. Just as you are arguing that there is no difference between aborting a normal fetus and aborting a fetus with a disability, so we have to say that Sesha is not special—she is equal. Yes, she takes more money, more time, more patience maybe, but these needs should only be an adjustment in the mental figuring of the would-be parents. This is and must be separate from any appraisal of the child's worth.

Love, Leo

EFK Letter #3

Dearest Leo,

I want to respond to the point that people need exposure to disabled persons if they are to understand that the value of disabled people is, as you say, equal to the value of those not impaired. But I want to start by addressing the equally serious matter of the slippery slope argument. You write that the argument I'm making draws a major line between normal and retarded (and other seriously disabled) children, "in the sense of them being difficult to raise." And while you acknowledge that this may be the case, you argue that we have to watch out for slippery slope arguments that would have us ask if, whenever detectable, other conditions which make a child more difficult to raise would make such conditions eligible for selective abortion as well. So if a child has a condition which would, for example, lead to recurring ear infections, that would be a condition for which parents may choose to abort. And you ask, "Are not all children a burden?"

A slippery slope problem is always hard. One doesn't even need to move into science fiction to face some of these dilemmas, since, although most genetic disorders detectable by prenatal testing today are potentially severely disabling, the tests cannot tell us how severe these disorders will be. In the language of genetics, test results cannot tell us the degree of expressivity of the genetic anomaly. Some conditions, such as Turner's syndrome, can result in a life that is little different from the life of persons without this disorder, except that the individual cannot bear a child.

Then again, women are already free to abort a fetus irrespective of any manifestations of disability. Some disability rights activists, who consider themselves to be feminists and pro-choice, argue against selective abortion on the grounds that there is an important distinction to be made between "aborting *any* fetus" and "aborting *this* fetus." That is, they maintain it is one thing to determine that you do not want to have a child, or to have a child at this time, or even with this man (that is, abortion *simpliciter),* and another to say that you do not want to have this particular child because it manifests such and such a trait (that is to say, selective abortion).[10] Perhaps this is the argument you would like to endorse?

Well then, here is my question. Why do women choose not to have a child? Or not to have a child at a particular time? Or not to have the child of a particular man? Well, for many reasons. But whatever the reasons (unless the decision is that they do it for ideological reasons, for example, "I do not believe it is right to increase the population of an already overcrowded world, but I will adopt and raise a child already born" or "I will not raise cannon fodder for a war state," or "I think this is too evil a world into which to bring new life"), the reason to abort involves some decision not to assume the burden , yes, *burden* of raising a child, now or under the current conditions of the woman's life. For yes, my dearest son, children are a burden.

Children, however, even in terrible times, under terrible conditions, are also a source of the deepest joy and satisfaction imaginable. Even under slavery, many women had their babies and raised them in spite of the near certainty that these children would be slaves, as abject as they themselves. Harriet Jacobs was a young slave woman who wrote of an old slave woman who chided her for shedding tears over her children: "Good old soul! She had gone through the world childless . . . No sweet little voices had called her mother; she had never pressed her own infants to her heart, with the feeling that even in fetters there was something to live for."[11]

We can add that even where a child is as profoundly disabled as Sesha, there is so much to treasure. Does my assertion that "I think it is terribly cruel to burden a couple...with the responsibility for a severely disabled child when prenatal testing can determine in advance the condition of the fetus" belie the value of a child like Sesha, a child who requires very extensive resources, material and emotional, to survive and thrive, whose care is so burdensome, even as it has such special rewards? Children are a burden, but we each engage in

numerous decisions about when and how to assume the burden, responsibility, and privilege of raising a child. We choose a mate or defer or decide against marriage (or cohabitation); we engage in or avoid sexual intercourse; we refuse or use contraception; we choose whether or not to take the pregnancy to term, when abortion is a choice; we commit to raising the child, or have family members raise the child temporarily or permanently, or give it up for adoption, and so forth.

Women have been thwarted in making such choices at various points along the way. As in all matters, we make choices but circumstances foil us and we are faced with unanticipated consequences of our actions or the actions of others. At each fork in the road, we have to decide. And when the matter is the care of a child, well or sick, able-bodied or disabled, we have to think if and how we can assume that burden and if the sacrifices required, at this time in our life, under these life circumstances, and given our current estimate of what our capacities are, what resources we can muster (remember how Berube quotes Janet declaring to him, "We can do this"), and what this child will require to survive and thrive. Rayna Rapp, an anthropologist studying women who have refused amniocentesis or who sought (or submitted to) it, and then based on a fetal diagnosis decided to abort, writes of the different decisions women make.[12] They are based on the women's perceptions and understandings, both of their circumstances and of the kind and extent of the disability.

The choices are enormously complex. An unmarried woman in her late thirties, whose pregnancy is "an accident" but who is delighted to be pregnant, chooses not to have amniocentesis because she knows that she will not have another opportunity to have a child. She knows she can welcome the child, whether or not there is a disability, as long as she has the support of the church she once left, the Seventh Day Adventists. Another woman, in her forties and with two sons and a daughter, chooses to abort a fetus diagnosed with Down syndrome because she is concerned about having a child with a disability at her advanced age. She fears that she will not live long enough to care for the child as the child ages. She is further concerned that such a major and unending responsibility will fall to her daughter alone. Another family, which includes a cousin with Down syndrome, in learning that the child will have a disability that may result in the child's being "slow," but outwardly normal in physique, decides to bring the pregnancy to term. They would have aborted if the child had Down syndrome because they were witness to the exertions on the part of their family in caring for the physical aspect of the disability of their cousin.

Many of these decisions are inflected by experience of race and the history of racial oppression. One African American family, whose fetus was diagnosed with Down syndrome, was told of farm communities where adults with Down are cared for and where they can participate in farm work. The father's response was, "Sounds too much like slavery to me." They decided to abort. Many urban white families, in contrast, find the thought of a rural life for their Down syndrome children a comforting notion.

So yes, all children are a burden, and maybe you are quite right to say that to argue for the permissibility of abortion when the fetus is diagnosed with a severe disability on the grounds that a disabled child presents greater burdens is untenable, since the question of where we draw the line is an inevitable and unavoidable one. Perhaps the best rejoinder (if there is one) is to say that because having a child, any child, is a great burden and a great responsibility, our obligation as a society and as prospective parents is to go into that great adventure with our eyes open and with as much forethought as we can muster about whether we can assume that burden in a responsible way. Because a disabled child poses special burdens and responsibilities, a mother and a family must know that it is a challenge that they are prepared and willing to meet, when, that is, foreknowledge of an impairment is an option.

In fact, judging from the accounts that Rayna Rapp has accumulated, it is just such thinking that does, in fact, predominate. These thoughts and these facts have a great bearing on the question of the expressivity of selective abortion, which I would like to sort out in a future letter. But for now, I want to mention two things.

First, the stigmatization of the disabled. Doesn't that play a big part in the decision making of families and in their thinking of what they can and cannot handle (and so, argue those opposed to selective abortion, permitting such abortions only reinforces the very stigmatization that is a causal factor in the decision). If so, the question you raise, whether stigmatization increases or decreases with a greater or lesser population of the disabled, is pertinent.

Second, a very important part of the decision-making around testing and abortion has to do with the resources that the society itself makes available. For many women, the idea of raising a child with disabilities is weighed against her own ambitions, the ambitions she has for her other children, the prospects for her disabled child when she can no longer care for her. Perhaps it is more appropriate to question how the larger society values or devalues the disabled life-by looking at the resources it withholds or devotes to children with disabilities and their families-than to impute a disregard for the value of the life of disabled persons to the pregnant woman who tests for and aborts a fetus with impairments. I want to talk more about this social dimension later. Must go now. Call us tonight, and let us know if you are going to remain in Taos.

Much love as always, Mom

EFK's Letter #4

Dearest Leo,

I am now going to try to respond to the first point, which you have been pressing throughout: The need to expose people to the joys and fulfillment of raising a child who is retarded (or severely disabled). And you want to add, rightly, that not all who are disabled are wonderful nor that anyone who is disabled needs to be wonderful or sweet or whatever positive attribute we want to put in, in order to be valued. Again, I have no argument. I also have no argument with the need to expose all of us to more persons who are disabled, whatever the disability. That educating ourselves and others about differences in abilities, in the rewards of raising a severely disabled child, is crucial if women are to make a well-informed and genuine choice. Perhaps it is especially important to become aware of those who are severely cognitively impaired, and of their presence as being crucial to enriching all of our lives. Increasing such awareness is vitally important if we want those who have had no intimate contact with disability to open their hearts and devote resources to improving the lives of the disabled and their families.

I think few things are as difficult for humans to face as disabilities they themselves do not have; few "differences"—not race, not gender, not sexuality—are as threatening to a person's notion of self. Most characteristics that put us in a relatively privileged position are ones that it is difficult for us to imaginatively transmute. A man won't turn into the devalued woman; the white into the devalued black; the Christian into the devalued Jew. But the able-bodied can in fact turn into the devalued disabled at the next turn in the road. You would think therefore that prejudice against the disabled would be contained, confined, because, after all, at any time "I" could turn into "them." But, instead, such a possibility only increases the prejudice, the avoidance, and the stigma.

I don't know how to get past this, except to show people our love for Sesha, to recognize the difficulty others have with Sesha and simply, by our example, help them past this. But does this mean I have gotten past all my prejudices concerning the disabled? No. The first thing is to recognize them, know where they come from, and then relate to the person and not the disability, except as you can be of service or learn from the person who has had to engage in struggles you yourself have not faced. I think that all the kids in your high school who watched your friend's sister participate in high school performances and athletics will have more understanding about what it means to have the Down syndrome that marked her as "different." They will not automatically respond to a pregnancy with, "If there's Down, we'll abort." But I also am sure that among special education teachers, who have a deeper knowledge of what retardation and severe disability mean, there

will be those who decide that, while they value everyone of their students, they themselves cannot take on the challenge and responsibility of raising a child with a severe disability. In fact, among the women in one of Rapp's studies, there were two women who were special education teachers who chose to abort.

Rayna Rapp cites another woman, on the other hand, who upon receiving the diagnosis visited a group home for the mentally retarded and chose to bring the pregnancy to term. Knowledge is crucial. I agree. And the time to get informed is not just when you are facing the decision. We need to be active in integrating persons with disabilities into every aspect of life, to seeing that our society devotes the resources that can facilitate such integration and facilitate the lives of disabled persons and their families. Only when this is the case will people have the exposure to children such as Sesha and Jamie Berube that will permit prospective parents to truly understand what is involved in raising a child with severe impairments.

However, when that is the case, raising a child like Sesha will also be different. Raising a child with developmental delays and deficits today is so very different than it was when Sesha was born. What was available to Jamie was not available, or only becoming so, when Sesha was born. "Early intervention" was an entirely new concept then. New York City sidewalks didn't have a cut in their curbs that made using a wheelchair so much less cumbersome—an improvement that helps not only the disabled but also every parent who has an infant or toddler in a carriage and every shopper with a shopping cart.

Still, we live in a society without guaranteed health care for every child, much less every adult. How would Janet and Michael Berube have paid for Jamie's care if they didn't have generous insurance plans through their employment? During Sesha's recent back operation, her surgeon alone cost $25,000, paid for through the generous health care plan my job provides. One professor I know who has a severely disabled child has an ongoing battle with his university because they set a one million-dollar limit on her medical insurance, and in the time he has been employed there (their daughter is now an adult), they have already exceeded the limit! So our society has done little to provide for even as basic a need as health care for the disabled. While this is also a difficulty for families with unimpaired children, for families with a disabled child, where medical emergencies are so much more frequent, having to consider cost can be devastating. The story that Berube tells of Jamie's early years is about par for the course—for some it's better, for some it's worse.

Then there is the question of the daily care of the severely disabled. There are now some respite programs that provide care for a disabled child so that a parent may have some time away from her disabled child, but these are woefully inadequate, as are the facilities for the severely retarded once they "age out" of the mandated school programs. If we want to speak of acts that are expressive of the devaluation of the life of the disabled, then to direct our attention to selective abortion is to direct us away from acts that are most egregiously expressive of this devaluation. The devaluation of the disabled life is expressed over and over again in the failure of our society to provide adequately for the disabled and their families. A woman who decides that she must make what is an excruciatingly difficult choice to abort (see Rapp's account of her own decision) may not be expressing that devaluation except in a secondary sense. She acts thus because she is faced with Hobson's choice—this or not at all: To raise a child with disabilities with only minimal social support (this) or to abort (not this child at all). It is an act motivated in part, at least, by the difficulties created in a society that fails to accord full humanity and citizenship to the severely disabled. This isn't to say that in a utopian society not a single fetus will be aborted because of disability. Society can make available certain material resources, but individual emotional resources will vary from family to family. Perhaps some people shouldn't be parents at all, and some shouldn't be parents to disabled children, at least when that situation can be foreseen. Some parents cannot love unconditionally. I have heard parents say that their love for a child was diminished because the child wasn't as smart as they wanted their child to be. How sad for that child, I think. How much more devastating for a child not to get the love and the special love that she needs to sustain the illnesses, the pain, the loneliness that so often accompany a disability.

People who come into our house say Sesha is lucky to have parents who love her so much. And our standard response is that we are lucky to have Sesha whom we can love so much. But, in truth, they are right. As lovable as Sesha is, not every family may have allowed themselves to find out how wonderful she is. It's hard to imagine since she touches your soul so, but I just know it's true. To be able to love her so, to find it hard to imagine that anyone couldn't love her so, is to be touched by a bit of grace, and it has been our good fortune to be granted that grace. But what would her life be like if she didn't have people to love her as we do?

That, my dear, is the most painful thought—the thought of what happens to her when we are no longer around. No, these are things no one has any right to tell a family—no one has a right to say to a family: You must take this on and if you don't you are immoral, you don't value a life that is disabled.

Finally, I am ready to address some of your concerns as a sibling about the expressivity of the act of abortion in the case of disability. I'll write this tonight and tomorrow and e-mail you tomorrow night.

Love, Mom

EFK's Letter #5

Dearest Leo,

You ask, "What kind of message does aborting the retarded send to would-be siblings?" And you say that there are two possibilities to consider. The first is a negative message, that parental love is conditioned on "soundness" and accomplishment, or as you put it: "The love my parents have for me is a condition of my being mentally and physically sound, not just of being a child of theirs." And you provide an alternative formulation of this idea, that "the only reason my parents want me is that I'm relatively smart and fit." I see your alternative formulation as one that has to do not with the infant when born (a time when we cannot assess intelligence or athletic ability, but only good health and absence of anomalies) but has rather to do with the child's realization of the potential that good health and soundness make available.

The other possible message seems at first more positive, but contains a hidden explosive that can shatter a child's sense of well-being. This is a very disturbing message that we need to explore. You suggest that the message received might go something like this: "My parents chose me and therefore must really care about me." Or, "My parents wouldn't just love any child they might have, they love me because I possess the desirable properties or characteristics that make me who I am." But this seemingly positive message becomes just another statement of a conditioned love. For then, as you say, "the family starts to seem more like a club, and less like a family," in which the members are selected based on some desirable features. But if a person starts to fall short of the desirable characteristics, she knows that she is no longer welcome in the club. In a family, this would lead a child to feel "that if I don't toe the line and exhibit the desirable characteristics, I'll no longer be valued." You continue: "If a child believes his membership in the family is contingent on not being retarded or otherwise disabled, he might at first value his place in it more highly because it was *earned*. . . . But the positive feeling that love has been earned can subside, and the child might instead feel a constant pressure to prove himself to be worthy of his place in the family. He will not view his family's love for him as unconditional love." We need to address these two possibilities separately.

First, however, we need to think a bit about what it means to send a message. You are asking about the kind of message the act of selective abortion based on disability sends to the sibling. Opponents of this sort of selective abortion ask, "What kind of message does it send to society about the value of the life of persons with disabilities?" Many opponents of selective abortion (see, for example, Saxton[13, 14, 15, 16]) claim that something is communicated in the decision to abort selectively for disability that we say we want a child, but we

do not want *this* child. They ask us to consider the claim that aborting that life sends the message that a disabled life is not one worth living; very much the way feminists have claimed that selective abortion for gender, which is generally a choice against having a girl, is a statement devaluing the life of females. They ask us furthermore to consider the impact of this sort of message on those who are female, in the one instance, and disabled, in the other instance. In focusing on the message that the sibling receives, your point is more specific, yet some of the considerations are the same. Some others, I'll try to show, are different.

The first thing we have to consider is whether the act of selective abortion is a "*saying!*" Is it an act of communication at all? What are we committing ourselves to when we claim that it is? Can we base an ethical evaluation of the act of selective abortion on the claim that the message sent is a devaluation of the life of those possessing the properties that determined the choice to abort? . . . If selective abortion is an act of communication in which the message arrives in a degraded form, then I fail to see how it can provide grounds for any ethical judgments or moral (much less legal) prescriptions.

First, in the case of selective abortion, we identify the addresser as the woman who decides to abort the fetus and the addressee as society in general (alternatively, the disabled community). Now, it is not clear that a contact is ever established between addressee and addresser. A woman rarely says: "Listen up, world. I am having an abortion based on a diagnosis of fetal abnormality and I am about to tell you why I choose to abort a fetus with such an abnormality." (There are, of course, exceptions.[17]) When we learn that someone had such an abortion we may not be in a position to query that decision, and the woman may not be in a position to query the addressee about the correctness of the message "received."

Second, there is no established code by which to decipher the "meaning" of such an act. That is, there is no established code or convention or practice to which both addressee and addresser can appeal when determining the meaning of that act. When I tell you, "It is raining," you (if you speak English) know how to understand that statement, and I know that you know. There is a common code that allows us to communicate a statement such as that. But such codes are not always available. If, given the conventions of foot apparel, I wear one green sock and one blue sock, you don't know how to interpret that action. Perhaps I dressed before dawn in the dim light and failed to discern the colors of my socks. Perhaps I lacked a clean pair of matching socks. Perhaps I was engaging in a flight of fancy. Or costumed myself for a play. Or dressed according to a pre-established code, thereby signaling to a comrade the start of a revolution.

The failure to discern a univocal—or indeed any—meaning of the act of selective abortion partakes of the ambiguity of all those actions which fail to be situated in practices that have an agreed upon meaning. (The same may be said about abortion itself, an act whose meaning remains contested.) I may already spend my life caring for persons with disabilities. I may have decided to adopt a retarded child once I have health insurance. I may feel that I can take on the care of a healthy child now, but a disabled child only at a later time in my life. Or I may think that even a limp makes life not worth living. There is no established code by which you can interpret my action and so understand what my abortion means to me, nor by which I can discern what my action might mean to you.

Third, we have to consider the extent to which the context influences the act of communication. If I am in a drought-stricken area, the statement "It's started to rain" carries an emotional charge very different from one carried by the same utterance in a flood-torn area. Context will affect the cognitive meaning or emotive charge of an utterance. The newspaper headline announcing the crash of Hemingway's plane, when the writer was assumed dead, ran: "Hemingway Lost in Africa." When it was learned that he was still alive but missing, the headline remained, but with a different meaning.[18] The less developed the code, the more ambiguity the code itself permits, the more the context will determine meaning. Because codes concerning acts of abortion and selective abortion are so underdeveloped and so contested, context is virtually, though not

entirely, determinative. In the case of selective abortion, the context includes both the particulars of the individual lives affected by the decision and the larger social setting in which the decision is made. Most contestations over this new technology and the decisions people make with it are struggles over how to understand and determine the context that, in turn, contributes to the meaning of the act. For many in the disability community, the context is one in which disability is stigmatized and persons with disabilities are devalued. That context, they argue, inevitably means that we interpret the act of selective abortion as another sign of the devaluation of a disabled life. That is to say, they believe society in general (the addressee) interprets the "utterance" of selective abortion in the context of the stigmatization of disability and that the message sent is that a child with disabilities will not be welcomed into a family. Therefore, they argue that the act of communication that results is that the disabled life is not worth living.

In response, we have to ask, "If we alter the context, is it the same message or not?" If we utter, "It's raining," to folks in a sodden Seattle and make them more miserable, is it the utterance (and what that utterance refers to, the fact that it is raining) or the context in which it is uttered that is the source of the misery? If we issue the same utterance to drought-stricken East Africa, are we uttering a message that makes people miserable? No.

If we are concerned with the devaluation of the life of the disabled (and that is something we should each be concerned about, regardless of whether we ourselves are disabled or have a disabled family member or if disability has never personally touched our lives), then we need to fix on, and fix, the context, not the utterance. Still, you might argue, to abort fetuses with disabilities is itself to further devalue the disabled. But that can't be the reply, because that is exactly what is at issue. What I will grant is that it is reasonable to infer that if many persons choose to abort fetuses with a particular characteristic, it is fair to make a hypothesis that those characteristics are devalued. But that is again, at best, a conclusion hypothesized about the causal factors that lead to the abortion and not a message that is sent out by the abortion. Only further questioning of actual motives can establish whether this woman aborted because she devalues disabled life. But even so, we should not confuse a message sent with a causal determinant of an action.

Now you may want to respond, "Look, Mom, it is only through exposure to actual retarded people that anyone can really appreciate how much they contribute." With selective abortion, "fewer people will know what joys can come from being around someone like Sesha. More ignorance, and this is never a good thing."

But if we could fix those conditions *in utero*, if we could have Sesha without the retardation, would we balk, even for one moment? And maybe, probably, Sesha wouldn't have the incredible sweetness she now has, a sweetness that is perhaps, in part, the result of her not encountering conditions that most of us encounter—an innocence of intentional evil, of senseless nastiness and stupidity that humans are capable of, of corrosive ambition, of frustrated dreams, of biting competition, and so on. What of it? Would we hesitate one moment to exchange her for a Sesha with all her mental faculties intact. Although every day I lay eyes on her, Sesha melts my heart with the purity of her joy, her laughter, I would not hesitate. Truly, I wouldn't hesitate.

Sesha's condition isn't just a difference, only it is that too. Sesha's condition is an impairment. If I can contemplate a Sesha without her impairments, or another child in her place, does this mean that I think that a disabled life is not worth living—that Sesha's life is not worth living? Absolutely not.

My life is worth living. Nonetheless there are conditions, ones that those I love would have rejoiced in, under which I would not have been born. Had my mother left Poland before the war and been spared the horrors of Auschwitz, I would not have been born. Does this mean that I cannot wish with all my heart that she had married the man from Toronto who had betrothed her and sent her the papers to leave Poland before the war? The fact that she didn't, of course, says nothing about the value she placed on my life. She could know nothing of what that life would have been like. But if she had had a crystal ball, and had foreseen it all—yet

had no idea of what awaited her if she went off with the beau from Toronto—foreseen both surviving Auschwitz and me and would still have chosen Toronto, could I blame her? Could I say she devalued my life—could I blame her for not choosing this child? But these are fantasies, and she could never know.

We can know no more of the life we do not conceive or the life we choose to abort. Would I have aborted Sesha if I had known of her condition? I don't know. It might depend on the level of attachment I felt at the time I learned that the fetus had some problem. If it was already my child in my heart and mind, I may not have. I may have thought just as I did once Sesha was born—our own version of Janet Lyon's "We can handle this."[19] But maybe that is not what I would have said. Maybe I would have investigated further, learned something of the lives of the retarded. Our decision (because *both* Dad and I would make this decision) may well have depended on where our investigation led us: to a home like the Berubes' or to a day treatment program like the one Sesha is currently in? These considerations reinforce some of the powerful points Berube makes with respect to the representation of the disabled. But it is also a confirmation of what I have wanted to underscore—namely the importance of the commitment of the society in general to the disabled.

So maybe we would have decided that there is joy enough for us here, and that we can make a good life for ourselves and our child. Maybe, and it is hard to think of it, I would have aborted. And we would never know Sesha. And that loss seems unimaginable. But I may have given birth to another child, whose nonexistence would seem equally unfathomable, and I would have wondered about the child I aborted. I would have stopped every time I saw a mentally retarded child or adult and wondered, with tears in my eyes. Just as now, with tears in my eyes, I think about the young woman of twenty-seven who might be a graduate student like my wonderful graduate students, or be thinking about marriage, or be out on the ski slopes with you. In each case there is a loss. It is a human tragedy.

No one can judge the choices of another in these cases based on what is at best a degraded form of communication. No one can make a moral evaluation based on this incomplete communicative situation. There is no singular utterance enunciated through a clear channel in an accepted code, in a nonambiguous context. It is a moral wrong to utter the word "nigger" in speaking of or to a person of African descent. It is a moral wrong to produce degrading and demeaning portrayals of women as sex toys for men. It is a moral wrong to reduce services for the disabled poor (doubly wrong). All these send vile messages that some people do not possess the value that others possess. But to selectively abort because the fetus I carry is likely to develop into a child with profound disabilities does not send any clear and unambiguous message. And the morality of that choice must be weighed in the conscience of the woman who makes that choice. She alone can know just what her act meant and if it was carried out as a consequence of moral sloth and uncaring, or through a responsible choice.

Now, at last, we get to your point about the message that the sibling receives. First, let's consider this situation of communication with the six factors that Jakobson delineates. The situation here is quite different than that of an undefined audience, "society in general." Why? First, because in this case one can establish that channels of communication are open, and second, because one can adopt a code by which to interpret this utterance. Furthermore, we can delimit the context, or at least specify the relevant contextual features. Put more simply, one can discuss the matter through an exchange, not unlike the exchange you and I are now having. I wonder, however, if such exchanges do take place. Were I an empirical scientist, I would like to conduct a study in which to ask this question. But I think it is a parental duty to explain to one's child why one makes, or why one has made, such a decision. Otherwise, all sorts of misinterpreted, unintended, or garbled messages are an inevitable outcome.

Let's take the first scenario you envision. The sibling assumes that the parent's love for him is conditioned on his sound mind and body and is concerned that if anything should happen to him that would cause him to be disabled, the parent would want to discard him, as, once before, she had aborted his disabled unborn

sibling. The first distinction I would want to put into place is the distinction between a born child, and the commitment a parent has to a child that has been born, and an unborn fetus. I believe that the concern you raise is a concern a child might raise in the case of any abortion. If the reason for the abortion is "I can't afford another one," a child may wonder, "What if we have less money than we have now? Will my parents want to be rid of me, too?" If the reason is, "I have my hands full with the children I have now," the child may wonder, "If I'm more trouble than I am now, will my parents want to be rid of me, too?" I think you see how it can go.

Remember, without discussion, with only the act of the abortion itself, we have not a true act of communication but a very degraded one. Once we have the distinction between the commitment to the born child and the tentative commitment to the unborn fetus, we are able to develop other features of the code and the context. We can make the case that the decision to abort was in significant measure a question of the parent's commitment to children already born, or to the other children the parents were likely to have. If there is a decision to bring the fetus to term, there is also considerable parental input that is demanded. A child may view any sibling as a rival, but a sibling that requires the additional attention a disabled child does may raise the level of resentment and jealousy. It is the job of the parent to open the channels of communication, to explain the decision (or the fate, as the case may be), and to integrate the normal and disabled siblings into one cohesive, caring family.

Too often we think that the message is obvious. We needn't check with our addressee if the intended message has been received. Your dad and I thought that it was obvious that our love for Sesha would give you the clear message that we love our children, unconditionally, irrespective of achievement. You might have gotten the unconditional part, but I'm not so sure about the "irrespective of achievement" part. Instead, you thought the message was that you had to compensate for the fact that Sesha would never have accomplishments, as those are normally tallied. Because we presumed the message was clear, we never made the effort to be certain that it was being received. (How dangerous to make moral judgments based on such bad communication channels.) I recall how as a four-year-old you mistook our affectionate responses to Sesha as a sign that we loved you less. We had to explain to you that Sesha understands only kisses and hugs—it is our sole means of communicating with her—whereas we could play and talk with you. Again, how careful we have to be in explaining our messages.

Now let's move to the second case: the sibling who first bathes in the love garnered for his particular characteristics and then comes to fear that such love is too unstable—that the family based on such love is more like a club than like a family. Here too one can invoke the distinction between commitment to the unborn and commitment to the born. But here I think the important point is that a family must not be like a club, whose membership is based on a set of desirable features. That is not how a family nurtures. We need a place where love is unconditional, where our mistakes are forgiven, where our imperfections are accepted and even cherished. We need such a place if we are to be emotionally whole. If the "message" that selective abortion for disability sends is that a disabled child is of less value, then it cuts into the sanctity of such a space and is corrosive. So here we have to be very clear. But once again, clarity comes from how we treat those with disabilities and not with a family's (and especially a woman's) decision to bring a fetus, any fetus, or this fetus to term. If we treat persons with disabilities with care and respect; if we attend to need when we see it and listen to the voices of those who wish to speak; if we treat all persons as moral equals, irrespective of ability or accomplishment; and if a household reflects this in all that it undertakes, then no child should think that it is valued merely for having certain desirable traits. If a child comes into a household where these values predominate, then the child comes into a home that welcomes her for the person she is, not for the traits that she bears. And if the message isn't getting through, then it's time to clear the channels of noise.

I love both you and Sesha with all my heart. Mom.

Leo's Letter #3

Mom,

Yes, the lines of communication must be open. And this is incredibly difficult. As open and honest as our family is, only in my twenty-first year have you and Dad and I discussed at any length many of the more painful, difficult aspects of having Sesha in the family. I have not even allowed a healthy dialogue to take place in my own head about Sesha until recently. Tremendous issues of anger and guilt have been lurking within me regarding Sesha, and coming to grips with them has been a big part of my post-college soul searching. You said that the act of aborting a disabled fetus will convey a harmful message of conditional love to the sibling unless the following condition is met: "If we treat persons with disability with care and respect; if we attend to need when we see it and listen to the voices of those who wish to speak; if we treat all persons as moral equals, irrespective of ability or accomplishment; and if a household respects this in all that it undertakes, then no child should think that it is valued only for having certain desirable traits."

There is only one problem, Mom. No child is consistently under the impression that the above condition is the case. In fact, no person for that matter thinks that his or her family is always treating him or her in such a way all the time. Even a family as wondeful as ours, *n'est-ce pas?* This passage does help me answer one thought that has been plaguing me throughout our discussions. Even though you did not abort Sesha, I remember experiencing every feeling that we have discussed a would-be sibling goes through as a result of a selective abortion. Just because you had Sesha and raised both of us honestly, better than I can imagine, I still managed to feel quite frequently and strongly throughout childhood, and even during many of my most formative moments, that Dad's and your love for me is a condition of my physical and mental abilities. Without these, I often felt, on some level, that I would not command your love and respect.

But when you break down the manner in which these messages get communicated in the case of an abortion, it helps me to see how this message could have been communicated so counterintuitively in the case of the elected birth, Sesha. It was in those moments in my upbringing when I felt treated as more than equal, when I got more attention than Sesha, or alternatively when I did not feel treated with the same care and respect as Sesha, that my young mind sometimes interpreted this nonequal treatment in terms of the inequalities and not the equalities. I thought I must be getting more attention than her because I can do more, or that I was getting less because she needed more. I think to some extent this phenomenon exists between all siblings, even between a child and a parent's career, between a child and the other spouse, whenever a parent's energies have to be distributed fairly. Anytime a child feels his status change, he is constantly searching for the cause of the change. Only a completely open line of communication continually sending a message of equally high value to all can truly do away with a mixed message. So, yes, Mom, I think you have hit on the secret of how not to send the wrong message to one's children when one decides to abort. I think it also happens to be a secret of parenting in general.

This leads me to my final thought. Let me say I do fear that allowing abortion based on prenatal screening will result in many abortions that are decided more quickly and based on less information than is ideal. Some women will even elect to have an abortion because they think less of disabled people, or because they want their children to be perfect. But, and this is my thought, parents make lousy decisions all the time. Some spend their money irresponsibly, some raise their kids to think they're worthless, others raise their kids to think they're worth more than everyone else, some beat their children. While sometimes I think it would be great to make laws that put a stop to such behavior, I know that in general that would not be a good idea. To insist that parents have children they are not thrilled about doesn't strike me as the best way to give children a great start in life. After all, the great burden of deciding whether or not to abort the child is small compared to the burden of raising the child. And if someone is not going to handle the decision responsibly, I would hate to see how they would handle the child. Might they become thrilled before the nine months are up? Might having the child shatter their prejudice against the disabled? Yes, but it also could take longer. And what messages would be sent out meanwhile?

Love, Leo

Notes

1. This chapter was prepared for The Hastings Center project, "Prenatal Genetic Testing for Genetic Disability," funded by the National Center for Human Genome Research of the National Institutes of Health, grant 1-R01- HG116801A2. It is a revised version of a chapter in *Norms and Values: Essays on the Work* if *Virginia Held,* ed. Joram G. Haber and Mark S. Half ton (Lanham, Md.: Rowman and Littlefield, 1998).
2. These diary entries were written prior to the correspondence that follows them.
3. Erik Parens, personal correspondence to Eva Feder Kittay, Briarcliff Manor, N.Y., May 21, 1996.
4. Adrienne Asch, "Reproductive Technology and Disability" in *Reproductive Laws for the 1990's,* ed. Sherill Cohen and Nadine Taub (Clifton, N.J.: Humana Press, 1988), pp. 69–124.
5. Michael Berube, *Life as We Know It* (New York: Pantheon Books, 1996).
6. Berube, *Life as We Know It,* p. 76.
7. Berube, *Life as We Know It.*
8. See Elizabeth Spelman, *Inessential Woman: Problems of Exclusion in Feminist Thought* (New York: Beacon Press, 1988), who speaks of the practice of saying about some people who are viewed as Other "but they are just like us" as "boomerang perception," one in which we only can see the other as "just like us," and never see ourselves as "just like them."
9. Ludwig Wittgenstein , *Tractatus Logico-Philosophicus* (London: Routledge & Kegan Paul, 1921).
10. Marsha Saxton, "Why Members of the Disability Community Oppose Prenatal Screening and Selective Abortion," in this volume; and Adrienne Asch and Gail Geller, "Feminism, Bioethics, and Genetics," in *Feminism and Bioethics: Beyond Reproduction,* ed. Susan Wolf (New York: Oxford University Press, 1996), pp. 318–50.
11. Harriet Jacobs, "Incidents in the Life of a Slave Girl: Written By Herself," in *The Classic Slave Narratives,* ed. Henry Louis Gates (1861; reprint, New York: Penguin, 1987), pp. 333–513, at p. 427.
12. Rayna Rapp, "The Ethics of Choice: After My Amniocentesis, Mike and I Faced the Toughest Decision of Our Lives," *Ms. Magazine* April (1984): 97–100; "The Power of 'Positive' Diagnosis: Medical and Maternal Discourses in Amniocentesis," in *Childbirth in America: Anthropological Perspectives,* ed. Karen Michaelson (South Hadley, Mass.: Bergin & Garvey, 1988), pp. 103–16; "Women's Responses to Prenatal Diagnosis: A Sociocultural Perspective on Diversity," in *Women and Prenatal Testing: Facing the Challenges of Genetic Technology,* ed. Karen H. Rothenberg and Elizabeth J. Thomson (Columbus: Ohio State University Press, 1994), pp. 219–33; "Refusing Prenatal Diagnosis: The Uneven Meanings of Bioscience in a Multicultural World," in *Science, Technology & Human Values* 23 (1998): 45–71; and "Risky Business: Genetic Counseling in a Shifting World," in *Articulating Hidden Histories,* ed. Rayna Rapp and Jane Schneider (Berkeley: University of California Press, 1995), pp. 173–89.
13. Marsha Saxton, "Disability Rights and Selective Abortion," in *Abortion Wars: A Half Century of Struggle, 1950-2000,* ed. Rickie Solinger (Berkeley: University of California Press, 1997), pp. 374–95.
[*Authors Note:* Citations to notes 14–16 fall within the omitted portion of this text.]
17. Rapp, "The Ethics of Choice."
18. Donald Davidson, "What Metaphors Mean," in *Inquiries into Truth and lnterpretation,* ed. Donald Davidson (Oxford: Oxford University Press, 1984); and Eva F. Kittay, *Metaphor: Its Cognitive Force and Linguistic Structure* (Oxford: Oxford University Press, 1987).
19. Berube, *Life as We Know It.*

1 Perception

Sensing, believing, and knowing

As I look at the green field before me, I might believe not only that there is a green field there but also that I *see* one. And I do see one. I visually perceive it. Both beliefs, the belief that there is a green field there, and the self-referential belief that *I* see one, are grounded, causally, justificationally, and epistemically, in my perceptual experience. They are produced by that experience, justified by it, and constitute knowledge in virtue of it.

The same sort of thing holds for the other senses. Consider touch. I not only believe, through touch (as well as sight), that there is a glass here, I also feel its cold surface. Both beliefs—that there is a glass here and that it is cold—are grounded in my tactual experience. I could believe these things on the basis of someone's testimony. My beliefs would then have a quite different status. For instance, my belief that there is a glass here would not be a *perceptual belief*, but only a belief *about a perceptible*, that is, a perceivable object, the kind of thing that can be seen, touched, heard, smelled, or tasted. Through testimony we have beliefs about perceptibles we have never seen or experienced in any way.

My concern is not with the hodgepodge of beliefs that are simply about perceptibles, but with perception and perceptual beliefs. Perceptual beliefs are not simply beliefs about perceptibles; they are *beliefs grounded in perception*. We classify beliefs as perceptual by the nature of their roots, not by the color of their foliage; by their grounds, not their type of content. Those roots may be visual, auditory, and so forth for each perceptual mode. But vision and visual beliefs are an excellent basis for discussing perception, and I will concentrate on them and mention the other senses only when it adds clarity.

Perception is a source of knowledge and justification mainly by virtue of yielding beliefs that *constitute* knowledge or are justified. But we cannot hope to understand perceptual knowledge and justification simply by exploring those beliefs. We must also understand what perception is and *how* it yields beliefs. We can then begin to understand how it yields knowledge and justification or—sometimes—fails to yield them.

The elements and basic kinds of perception

There are apparently at least four elements in perception: (1) the perceiver, me; (2) the object, the field I see; (3) the sensory experience, say my visual experience of colors and shapes; and (4) the relation between the object and the subject, commonly taken to be a causal relation by which the object produces the sensory experience in the perceiver. To see the field is apparently to have a certain sensory experience as a result of the impact of the field on our vision.

Some accounts of perception add to the four items on this list; others subtract from it. To understand perception we must consider both kinds of account and how these elements are to be conceived in relation to one another. But first, it is essential to explore examples of perception.

There are several quite different ways to speak of perception. Each corresponds to a different way of perceptually responding to experience. We often speak simply of what people perceive, for instance see. We also speak of what they perceive the object to be, and we commonly talk of facts they know through perception, such as that the grass is long. Visual perception most readily illustrates this, so let us start there.

I see, hence perceive, the green field. Second, speaking in a less familiar way, I see it *to be* rectangular. Thus, I might say that I know it looks irregular from the nearby hill, but from the air you can see it to be perfectly rectangular. Third, I see *that it is* rectangular. Perception is common to all three cases. Seeing, which is a paradigm perception, is central in each.

The first case is one of *simple perception*, perception taken by itself (here, visual perception). I see the field, and this experience is the visual parallel of hearing a bird (an auditory experience), touching a glass (a tactual experience), smelling roses (an olfactory experience), and tasting mint (a gustatory experience). If the first case is simply *perceiving of* some object, the second is a case of *perceiving to be*, as it is seeing something to be so: I do not just see the field, as when I drive by at high speed and do not even realize what is in my peripheral vision; rather, I see the field to be rectangular. The third case is one of *perceiving that*; it is seeing that a particular thing is so, namely that the field is rectangular.

These cases represent three kinds, or *modes*, of perception. Perception of the simplest kind (or in the simplest mode), such as seeing, occurs in all three; but, especially because of their relation to knowledge and justified belief, they are significantly different. We can best understand these three kinds (or modes) of perception if we first focus on their relation to belief.

Perceptual belief

The last two cases—perceiving *that*, and perceiving *to be*—are different from the first—perceiving *of*—in implying corresponding kinds of beliefs: seeing that the field is rectangular implies believing that it is, and seeing it

to be green implies believing it to be green. If we consider how both kinds of beliefs—beliefs *that* something is so and beliefs *of* (hence *about*) something—are related to perception, we can begin to understand how perception occurs in all three cases, the simple and the more complex. In my second and third examples of perception, visual perception (seeing) issues in beliefs that are grounded in seeing and can thereby constitute visual knowledge, such as knowing that the field is green.[1]

In our example of simple perception, my just seeing the field provides a basis for both kinds of beliefs. It does this even if, because my mind is entirely occupied with what I am hearing on the radio as I glance over the field, no belief about the field actually arises in me. The visual experience is, in this instance, like a foundation that has nothing built on it but is ready to support a structure. If, for example, someone were to ask if the field has shrubbery, then given the lilacs prominent in one place, I might immediately form the belief that it does and assent. This belief is visually grounded; it comes *from* my seeing the field though it did not initially come *with* it. When visual experiences do produce beliefs, as they usually do, what kinds of beliefs are these, and how are they specifically perceptual?

Many of my beliefs arising through perception correspond to perception *that*, say to seeing that the lilacs are blooming. I believe that the field is lighter green toward its borders, that it is rectangular in shape, and that it has many ruts. But I may also have various beliefs about it that are of the second kind: they correspond to perception *to be*, for instance to seeing something to be a certain color. Thus, I believe the field to be green, to be rectangular, and so on. The difference between these two kinds of belief is significant. As we shall shortly see, it corresponds first of all to two distinct ways in which we are related to the objects we perceive and, second, to two different ways of assessing the truth of what, on the basis of our perceptions, we believe.

The first kind of belief just described is the kind people usually think of when they consider beliefs: it is called *propositional*, as it is generally considered a case of believing a proposition—say, *that* the field is rectangular. The belief is thus true or false depending on whether the proposition in question—here *that the field is rectangular*—is true or false. In holding the belief, moreover, in some way I think of what I see as a field which is rectangular: in believing that the field is rectangular, I *conceive* what I take to be rectangular *as* a field.

The second kind of belief might be called *objectual*: it is a belief regarding an object, say the field, with which the belief is actually connected. This is an object *of* (or about) which I believe something, say that it is rectangular. If I believe the field to be rectangular, there really is such an object, and I have a certain relation to it. A special feature of this relation is that there is no particular proposition I must believe about the field. To see that there is no particular proposition, notice that in holding this objectual belief I need not think of what I see *as* a field. I might mistakenly take it to be (for instance) a lawn or a grasslike artificial turf, yet still believe it to be rectangular. I might

think of it just in terms of what I believe it to be and not in terms of anything else.

Thus, although there is *some* property I must take the field to have—corresponding to what I believe it to *be*—there is no other particular way I must think of it. With objectual belief, then, there is no particular notion, no specific conceptual "handle," that must yield the subject of any proposition I believe about the object: I do not have to believe that the *field* is green, that the *grass* is green, or any such thing. Perception leaves us vast latitude as to what we learn from it. People differ greatly in the beliefs they form about the very same things they see.[2]

The concept of objectual perception, then, is very permissive about what one believes about the object perceived. This is one reason why it leaves so much space for imagination and learning—a space often filled by the formation of propositional beliefs, each capturing a different aspect of what is perceived, say that the field is richly green, that it is windblown, and that it ends at a treeline.

A different example may bring these points out further. After seeing a distant flare and coming to believe, of something blurry and far away, that it glowed, one might ask, 'What on Earth was it that glowed?' Before we can believe the proposition that a *flare* glowed, we may have to think about where we are, the movement and fading of the glow, and so forth. The objectual belief is a guide by which we may arrive at propositional beliefs and propositional knowledge.

Perception, conception, and belief

The same kind of example can be used to illustrate how belief depends on our conceptual resources in a way that perception does not. Suppose I had grown up in the desert and somehow failed to acquire the concept of a field. I could nonetheless see the green field, and from a purely visual point of view it might *look* the same to me as it does now. I could also believe, regarding the field I see—and perhaps conceive as sand artificially covered with something green—that it is rectangular. But I could not believe that the *field* is rectangular. This propositional belief as it were portrays what I see *as* a field in a way that requires my having a concept of one.

There is a connection here between thought and language (or at least conceptualization). If I believe (think) that the field is rectangular, or even simply have the thought that it is, I should be able to *say* that it is and to know what I am talking about. But if I had no concept of a field, then in saying this I would not know what I am talking about.[3] Similarly, a two year old, say, Susie, who has no notion of a tachistoscope, can, upon seeing one and hearing its fan, believe it to be making noise; but she cannot believe specifically that the tachistoscope is making noise. Her propositional belief, if any, would be, say, that the thing on the table is making noise. Since this is true, what she believes is true and she may know this truth, but she need not know much

about the object this truth concerns: in a way, she does not know what it is she has this true belief *about*.

The general lesson here is important. A basic mode of learning about objects is to find out truths about them in this elementary way: we get a handle on them through perceptually discriminating some of their properties; we form objectual (and other) beliefs about them from different perspectives; and (often) we finally reach an adequate concept of what they are. From the properties I believe the flare in the distance to have, I finally figure out that it is a flare that has them. This suggests that there is at least one respect in which our knowledge of (perceptible) *properties* is more basic than our knowledge of the substances that have them; but whether that is so is a question I cannot pursue here.

Unlike propositional beliefs, objectual beliefs have a significant degree of indefiniteness in virtue of which it can be misleading simply to call them true or false; they are accurate or inaccurate, depending on whether what one believes of the object (such as that it is rectangular) is or is not *true of* it. Recall Susie. If she attributes noise-making to the tachistoscope, she truly believes, *of* it, that it is making noise. She is, then, *right about it*. But this holds even if she has no specific concept of what it is that is making the noise. If we say unqualifiedly that her belief about it is true, we invite the question 'What belief?' and the expectation that the answer will specify a particular proposition, say that the tachistoscope is making noise. But it need not, and we might be unable to find any proposition that she does believe about it. She can be right about something without knowing or even having any conception of what kind of thing it is that she is right about.

Knowledge is often partial in this way. Still, once we get the kind of epistemic handle on something that objectual belief can provide, we can usually use that to learn more about it.[4] Suppose I see a dog's tail projecting from under a bed and do not recognize it as such. If I believe it to be a slender furry thing, I have a place to start in finding out what else it is. I will, moreover, be disposed to form such beliefs as that there is a slender furry thing there. I will also have justification for them. But I need not form them, particularly if my attention quickly turns elsewhere.

Propositional and objectual perception

Corresponding to the two kinds of beliefs I have described are two ways of talking about perception. I see *that* the field is rectangular. This is (visual) *propositional perception*: perceiving *that*. I also see it *to be* rectangular. This is (visual) *objectual perception*: perceiving *to be*. The same distinction apparently applies to hearing and touch. Perhaps, for example, I can hear that a piano is out of tune by hearing its sour notes, as opposed to hearing the tuner say it needs tuning. As for taste and smell, we speak as if they yielded only simple perception: we talk of smelling mint in the iced tea, but not of smelling that it is minty or smelling it to be minty. Such talk is, however,

intelligible on the model of seeing that something is so and seeing it to be so. We may thus take the distinction between perceiving *that* and perceiving *to be* to apply in principle to all the senses.

It is useful to think of perceptual beliefs as *embedded* in the corresponding propositional or objectual perception, roughly in the sense that they are integrally tied to perceiving of that kind and derive their character and perhaps their authority from their perceptual grounding. Take propositional belief first. My belief that the field is rectangular is embedded in my seeing that it is, and Susie's believing the tachistoscope to be making noise is embedded in her hearing it to be doing so. In each case, the belief is an element in perception of the corresponding kind. These kinds of perception might therefore be called *cognitive*, since belief is a cognitive attitude: roughly the kind having a proposition (something true or false) as its object.[5] The object of the belief that the field is rectangular is the specific proposition that the field is rectangular, which is true or false.

Now consider objectual perceptual beliefs. If believing the tachistoscope *to be* making noise has a propositional object, that object may be plausibly taken to be some proposition or other to the effect that it is making noise, which (though left unspecified by the ascription of the belief) is also true or false. But some objectual perceptions may also be plausibly conceived as simply attributions of a perceptible property to the thing perceived; here the embedded objectual belief is *true of* the object rather than simply true. A tiny, prelingual child might see the liquid offered to it to be milk yet not believe (or disbelieve) the proposition that it is milk. In this respect, belief is unlike attitudes of approval or admiration or indignation, which are evaluated not as true or false but rather as, say, appropriate or inappropriate.[6]

Both propositional and objectual beliefs are grounded in simple perception. If I do not see a thing at all, I do not see that it has any particular property and I do not see it to be anything. Depending on whether perceptual beliefs are propositional or objectual, they may differ in the kind of knowledge they give us. Propositional perception yields knowledge both of *what* it is that we perceive and of some *property* of it, for instance of the *field's* being *rectangular*. Objectual perception may, in special cases, give us knowledge only of a property of what we perceive, say of its being green, when we do not know what it is or have any belief as to what it is.

In objectual perception, we are, to be sure, in a good position to come to know *something* or other about the object, say that it is a green expanse. Objectual perception may thus give us information not only about objects of which we have a definite conception, such as home furnishings, but also about utterly unfamiliar objects of which we have at most a very general conception, say 'that noisy thing'. This is important. We could not learn as readily from perception if it gave us information only about objects we conceive in the specific ways in which we conceive most of the familiar things we see, hear, touch, taste, and smell.[7]

Seeing and believing

Both propositional and objectual perceptual beliefs are commonly grounded in perception in a way that apparently connects us with the outside world and assures their truth. For instance, my visual belief that the field is rectangular is so grounded in my seeing the field that I veridically (truly) see that it is rectangular; my tactually believing the glass to be cold is so grounded in my feeling it that I veridically feel it to be cold. Let us explore the relation between perception and belief.

Perceptually embedded beliefs

Must beliefs grounded in seeing be true? Admittedly, I might visually (or tactually) believe that something is rectangular under conditions poor for judging it. Compare viewing a straight stick half submerged in water (it will look bent). My visually grounded belief might then be mistaken. But such a mistaken belief is not *embedded* in propositional perception that the stick is bent—that proposition is false and hence is *not* something one sees is so (or to be so). The belief is merely *produced* by some element in the simple perception of the stick: I see the stick in the water, and the operation of reflected light causes me to have the illusion of a bent stick. I thus do not see that the stick is bent: my genuine perception is of it, but not of its curvature. Seeing that curvature or seeing that the stick is bent would entail that it *is* bent, which is false. If the stick is not bent, I cannot see that it is.

As this suggests, there is something special about both perceiving *that* and perceiving *to be*. They are *veridical experiences*, that is, they imply truth. Specifically, if I see that the field is rectangular, or even just see it to be rectangular, then it truly is rectangular. Thus, when I simply see the rectangularity of the field, if I acquire the corresponding embedded perceptual beliefs—if I believe that it is rectangular when I see that it is, or believe it to be rectangular when I see it to be—then I am correct in so believing.

Perceiving *that* and perceiving *to be*, then, imply (truly) believing something about the object perceived—and so are *factive*. Does simple perception, perception *of* something, which is required for either of these more complex kinds of perception, also imply true belief? Very commonly, simple perception does imply truly believing something about the object perceived. If I hear a car go by, I commonly believe a car is passing. But could I not hear it, but be so occupied with my reading that I form no belief about it? Let us explore this.

Perception as a source of potential beliefs

As is suggested by the case of perception overshadowed by preoccupation with reading, there is reason to doubt that simple perceiving *must* produce any belief at all. Moreover, it commonly does not produce beliefs even of

what *would* be readily believed if the question arose. Suppose I am looking appreciatively at a beautiful rug. Must I believe that it is not producing yellow smoke, plain though this fact is? I think not; there seems to be a natural economy of nature—perhaps explainable on an evolutionary basis—that prevents our minds from being cluttered with the innumerable beliefs we would have if we formed one for each fact we can see to be the case.

This line of thought may seem to fly in the face of the adage that seeing is believing. But properly understood, that may apply just to propositional or objectual seeing. In those cases, perception plainly does entail beliefs. Seeing that golf ball-size hail is falling *is* (in the sense that it entails) believing it.[8] This fact, however, is not only perceptible; it is striking.

In any event, could I see the field and believe nothing regarding it? Must I not see it to be something or other, say green? And if so, would I not believe, of it, *something* that is true of it, even if only that it is a green object some distance away? Consider a different example.

Imagine that we are talking excitedly and a bird flies quickly across my path. Could I see it, yet form no beliefs about it? There may be no clearly correct answer. For one thing, although there is much we *can* confidently say about seeing and believing, 'seeing' and 'believing' are, like most philosophically interesting terms, not precise. They have an element of vagueness. No standard dictionary definition or authoritative statement can be expected either to tell us precisely what they mean or, especially, to settle every question about when they do and do not apply.[9] Still, we should be wary of concluding that vagueness makes any significant philosophical question unanswerable. How, then, should we answer the question whether seeing entails believing?

A negative response might be supported as follows. Suppose I merely see the bird but pay no *attention* to it because I am utterly intent on our conversation. Why must I form any belief about the bird? Granted, if someone later asks if I saw a blue bird, I may assent, thereby indicating a belief that the bird *was* blue. But this belief is not perceptual: it is about a perceptible and indeed has visual content, but it is not grounded in seeing. Moreover, it may have been formed only when I recalled my visual experience of the bird. Recalling that experience in such a context may *produce* a belief *about* the thing I saw even if my original experience *of* the thing did not. For plainly a recollected sensory experience can produce beliefs about the object that caused it, especially when I have reason to gain information about that object. Perhaps one notices something in one's recollected image of the bird, an image merely recorded in the original experience, but one formed no belief about the bird. Granted, perception must produce a sensory experience, such as an image, and granted such an image—and even a recollection of it—is raw material for beliefs; it does not follow that perception must produce beliefs.

It might be objected that genuinely seeing an object must produce beliefs, even if we are not *conscious* of its doing so. How else can perception guide our behavior, as it does when, on seeing a log in our path, we step over it?

One answer is that not everything we see, including the bird that flies by as I concentrate on something else, demands or even evokes a cognitive response, particularly one entailing belief-formation. If I am cataloguing local birds, the situation is different. But when an unobtrusive object we see—as opposed to one blocking our path—has no particular relation to what we are doing, perhaps our visual impressions of it are simply a *basis* for forming beliefs about it should the situation call for it, and it need not produce any belief if our concerns and the direction of our attention give the object no significance.

Despite the complexity I am pointing to in the relation between seeing and believing, clearly we may hold what is epistemologically most important here. Suppose I can see a bird without believing anything about (or of) it. Still, when I do see one, I *can* see it to be something or other, and my perceptual circumstances are such that I might readily both come to believe something about it *and* see that to be true of it. Imagine that someone suddenly interrupts a conversation to say, 'Look at that bird!' If I see it, I am in a position to form some belief about it, if only that it is swift, though I need not actually form any belief about it, at least not one I am conscious of.

To see these points more concretely, imagine I am alone and see the bird in the distance for just a second, mistakenly taking it to be a speck of ash. If there is not too much color distortion, I may still both know and justifiedly believe it to be dark. Granted, I would misdescribe it, and I might falsely believe that it is a speck of ash. But I could still know something about it, and I might point the bird out under the misleading but true description, 'that dark thing'. The bird *is* the thing I point at; and I can see, know, and justifiedly believe that there is a dark thing there.

My perception of the bird, then, gives me a ready *basis* for some knowledge and justification, even if the perception occurs in a way that does not cause me to believe that there is, say, a *bird* before me and so does not give me *actual* knowledge of it. Seeing *is* virtual believing, or at least potential believing. A similar point holds for simple perception in the other senses, though some, such as smell, are in general less richly informative than sight.[10]

The perceptual hierarchy

Our discussion seems to show that simple perceiving need not produce belief, and objectual perceiving need not always yield propositional perceiving. Still, this third kind of perception is clearly not possible without the first and, I think, the second as well. I certainly cannot see *that* the bird is anything if I do not see it at all; and I must also see it in order to see it *to be* something, say a speck of blue. Thus, simple perceiving is fundamental: it is required for objectual and propositional perceiving, yet does not clearly entail either. If, for instance, you do not perceive in the simple mode, say see a blue speck, you do not perceive in the other two modes either, say see a speck to be blue or see that it is blue. And as objectual perceiving seems possible without

propositional perceiving, but not conversely, the former seems basic relative to the latter.

Simple, objectual, and propositional perception

We have, then, a perceptual hierarchy: propositional perceiving depends on objectual perceiving, which in turn depends on simple perceiving. Simple perceiving is basic, and it commonly yields, even if it need not always yield, objectual perceiving, which, in turn, commonly yields, even if it need not always yield, propositional perceiving. Simple perceiving, such as just seeing a green field, may apparently occur without either of the other two kinds, but seeing something *to be* anything at all, such as rectangular, requires seeing it; and seeing *that* it is something in particular, say green, requires both seeing it to be something and, of course, seeing it.

Thus, even if simple perception does not always produce at least one true belief, it characteristically does position us to form any number of true beliefs. It gives us cognitive *access* to perceptual information, perhaps even *records* that information in some sense, whether or not we register the information conceptually by forming perceptual beliefs of either kind.

The informational character of perception

As this suggests, perception by its very nature is *informational*; it might even be understood as equivalent to a kind—a sensory kind—of receipt of information about the object perceived.[11] The point here is that not all perceptually given information is *propositional* or even conceptualized. This is why we do not receive or store all of it in the contents of our beliefs. Perceptual content—conceived as the content of a simple perception—is at least in part determined by the properties we are sensorily conscious of in having that experience; it is not equivalent to the content of the perceptual belief(s) that experience may produce.

Some of the information perception yields is imagistic. Indeed, we may think of all the senses as capable of yielding images or, for the non-visual senses, at least of yielding the non-visual counterparts of images—*percepts*, to use a technical term for such elements in perceptual experience occurring in any sensory mode, whether visual or auditory or of some other kind. It is in these sensory impressions that the bulk of perceptual information apparently resides. This point explains the plausibility of the idea that a picture is worth a thousand words—which is not to deny that, for some purposes, some words are worth a thousand pictures. A single report of smoke may avert a catastrophe; a single promise may alter a million lives.

It is in part because perception is so richly informative that it normally gives us not only imagistic information but also situational justification. Even if I could be so lost in conversation that I form no belief about the passing bird, I am, as I see it pass, normally justified in believing something about it,

concerning its perceptible properties, for instance that it glides.[12] There may perhaps be nothing highly specific that I am justified in believing about it, say that it is a cardinal or that its wingspan is ten inches, but if I really see it, as opposed to its merely causing in me a visual impression too indistinct to qualify me as seeing it, then there is something or other that I may justifiably believe about it.

When we have a clear perception of something, it is even easier to have perceptual justification for believing a proposition about it without actually believing it. Just by taking stock of the size of the field in clear view before me, I am justified in believing that it has more than 289 blades of grass; but I do not ordinarily believe—or disbelieve—any such thing about grassy fields I see. It was only when I sought a philosophical example about perception and belief, and then arbitrarily chose the proposition that the field has more than 289 blades of grass, that I came to believe this proposition. Again, I was justified in believing the proposition before I actually did believe it.

Perceptual justification and perceptual knowledge

What is it that explains why seeing the bird or the field justifies us in believing something about what we see, that is, gives us situational justification for such a belief? And does the same thing explain why seeing something enables us to know various facts about it?

Seeing and seeing as

One possible answer is that if we see something at all, say a bird, we see it *as* something, for instance black or large or swift, and we are justified in believing it to be what we see it as being. The idea is that all seeing and perhaps all perceiving is *aspectual perception* of a kind that confers justification. We see things by seeing their properties or aspects, for instance their colors or their front sides, and we are justified in taking them to have the properties or aspects we see them as having.

Let us not go too fast. Consider two points, one concerning the nature of seeing *as*, the other its relation to justification.

First, might not the sort of distinction we have observed between situational and belief justification apply to seeing itself? Specifically, might not my seeing the bird imply that I am only in a *position* to see it *as* something, and not that I *do* see it as something? It is true that when we see something, we see it *by* seeing some property or aspect of it; but it does not follow that we see it as *having* this property or aspect. I might see a van Gogh painting by its colors, shapes, and distinctive brush strokes, but not see it as having them because my visual experience is dominated by the painting as a whole. Someone might reply that if I see it by those properties, I am disposed to believe it has them and so must see it as having them; but this disposition

implies at most a readiness to see it as having them. There may, to be sure, be a sense in which if we see something aright, for example see a van Gogh with recognition of it as his, then we must see it as what we recognize it to be.

Seeing *as* can also be a matter of conceptualization—roughly, *conceiving as*. But this is different from perceptual seeing as. The distinction between perceptual seeing *as* and perceptual seeing *by* remains. Seeing *by* is causal and discriminative but not necessarily ascriptive or, especially, conceptual. Seeing *as,* though also causal, is often ascriptive and commonly conceptual. We see faces by seeing (for example) the distinctive shape of the eyes and mouth, but need not ascribe those to those we see or conceptualize these properties. But if we see a painting as blurry, we commonly ascribe that property to it and may conceptualize the painting as blurry.

Second, suppose that seeing the bird did imply (visually) seeing it *as* something. Clearly, this need not be something one is justified in believing it to be (and perhaps it need not be something one *does* believe it to be). Charles, our biased birdwatcher, might erroneously see a plainly black bird as blue, simply because he so loves birds of blue color and so dislikes black birds that (as he himself knows) his vision plays tricks on him when he is bird-watching. He might then not be justified in believing that the bird is blue.

Assume for the sake of argument that seeing implies seeing *as* and that typically, seeing as implies at least objectually believing something or other about the thing seen. Still, seeing an object as having a certain property—say, a stick in the water as bent—does not entail that it has the property. Nor does it always give one (overall) situational justification for believing it to have that property.

Perceptual content

It is natural to think of perception as in some way *representational*. If we see things by seeing their properties, for instance, then our perceptual experience in some way represents the object as having them. If perceiving entailed believing, we could perhaps take it to have the same content of the entailed belief(s). But (simple) perception apparently does not entail believing, so this conception of its content is mistaken. For propositional and objectual perception, however, we might plausibly say something like this: the content of my perception that *p* includes both the proposition that *p* (hence also the content of that proposition) and also the content of my objectual perception of the thing in question; that content includes the properties I perceive the thing to have.

If we seek a broad notion of perceptual content for simple perception, we might say that all the properties represented in a perceptual experience constitute its content. Then, for greater specificity, we might call the totality of perceptually represented properties the *property content*. These include properties an object is seen *as* having.[13] They apparently also determine "what it

is like" to perceive the object, say a squirrel in a tree. In seeing it, one's visual field is determined mainly by the grey, the distinctive furry shape, and the arboreal background.

For propositional and objectual perception, we might call the property-ascriptive propositions that the perceiver perceptually believes on the basis of the perceptual experience their *doxastic propositional content.* If we want to capture all the propositions that one might justifiedly believe (and know) on the basis of the perception, we might speak of its *total propositional content*. This would include such propositions as that the squirrel is crouching, has a nut in its mouth, is in sunlight, and many more that need not be believed as a result of simply seeing the animal.[14]

Seeing as and perceptual grounds of justification

Whether or not seeing always implies seeing *as*, it does have property content and normally puts one in a position to form at least one justified belief about the object seen. Suppose I see the bird so briefly and distractedly that I do not see it as anything in particular; still, my visual impression of it has some feature or other by which I am justified in believing something of the bird, if only that it is a moving thing. Even Charles would be justified in believing something like this. His tendency to see black birds as blue is irrelevant to his perception of movement and does not affect his justification for believing such moving objects to be in motion.

Suppose, however, that for hours Charles had been hallucinating all manner of unreal things, and he knows this. Then he might not be justified in taking the bird he sees to be *anything* real, even though it is real. For as a rational person in this position he should see that if his belief is true, it may well be true only in the way a lucky guess is. Thus, the best conclusion here—and I suggest that this is an important justification principle concerning perception—is that normally, seeing an object gives one situational justification for believing something or other about it.

More broadly, it is very plausible to hold that *the evidence of the senses*—including above all the sensory experiences characteristic of perception—normally provides justification for beliefs with content appropriate to that evidence. If your experience is of a green expanse, you are justified in believing there is something green before you; if it is of something cool in your hand, you are justified in believing there is something cool in your hand; and so on.

One might also say something slightly different, in a terminology that is from some points of view preferable: seeing an object (always) gives one *prima facie justification* for believing something or other about it. Prima facie justification is roughly justification that prevails unless *defeated.* The two main kinds of defeater are such overriding factors as a strong justification for believing something to the contrary and such undermining (or undercutting) factors as my knowledge that I have been hallucinating and at present cannot

trust my senses. Overriders defeat prima facie justification by justifying an incompatible proposition instead; undermining defeaters simply prevent the would-be justification from succeeding. If I see a green field, I have a justification for believing it to be green; but I may not be justified, *overall*, in believing this if credible friends give me compelling reason to believe that despite appearances the field is entirely covered by blue grass, or that I am not seeing a field at all but hallucinating one.[15] In the former case, my justification is defeated by my acquiring better justification for a contrary proposition; in the latter, my visual justification is reduced below the threshold of success. If it is not eliminated, it is too weak to license saying I am justified in believing the proposition.

If seeing is typical of perception in (normally) putting us in a position to form at least one justified belief about the object seen, then perception in general normally gives us at least situational justification. This is roughly justification *for holding* a belief of the proposition for which we have the justification. As our examples show, however, it does not follow that every perceptual belief is justified. Far from it. Some perceptual beliefs, such as perceptual beliefs that are evidentially undermined by one's having formed similar beliefs based on hallucinations, are not. As with the biased bird-watcher, belief can be grounded in perception under conditions that prevent its being justified by that grounding.

Nevertheless, there is a simple principle of justification we can see to be plausible despite all these complexities: normally, a visual belief that is embedded in seeing that something is so or in seeing it *to be* so is justified (and it is always prima facie justified). If we see that an object has a property (say, that a field is rectangular) and, in virtue of seeing that it has that property (say, is rectangular), believe that it does, then (normally) we justifiedly believe that it does. Call this *the visual justification principle*, since it applies to cases of belief based on seeing that what is believed is true (or seeing it to be true).

I say *normally* (and that the justification is prima facie) because even here one's justification can be defeated. Thus, Charles might see that a bird is blue and believe on this basis that it is, yet realize that all morning he has been seeing black birds as dark blue and thus mistaking the black ones for the blue ones. Until he verifies his first impression, then, he does not justifiedly believe that the bird is blue, even though it in fact is. (We could say that he has *some* justification for believing this, yet better justification for not believing it; but to simplify matters I am ignoring degrees of justification.) He does indeed see a bird and may justifiedly believe that, but his belief that the bird is blue is not justified.

Suppose, on the other hand, that Charles has no idea that he has been hallucinating. Then, even when he does hallucinate a blue bird, he may be justified in believing that there is a blue bird before him. This suggests a related principle of justification, one that applies to visual experience whether it is a case of seeing or merely of visual hallucination: *When, on the basis of an apparently normal visual experience (such as the sort we have in seeing a bird*

nearby), one believes something of the kind the experience seems to show (for instance that the bird is blue), normally this belief is justified. Call this the *visual experience principle*, since it applies to cases in which one has a belief based on visual experience even if not an experience of actually seeing (the veridical kind). The visual principle takes us from seeing (vision) to justification; the visual experience principle takes us from visual experience—conceived as apparent seeing—to justification. The latter is wider: it indicates that visual experience can justify a huge range of beliefs, not just a belief to the effect that an object in fact has a property one sees it to have.

Similar principles can be formulated for all of the other senses, though the formulations will not be as natural. If, for example, you hear a note to be flat and on that basis believe that it is flat, normally your belief is justified. It is grounded in a veridical perception in which you have discriminated the flatness you believe the note has. And suppose, by contrast, that in what clearly seem to be everyday circumstances you have an utterly normal-seeming auditory hallucination of a flat note. If that experience makes it seem clear that you are hearing a flat note, then if you believe on the basis of the experience that this is a flat note, normally your belief would be justified. You have no reason to suspect hallucination, and the justification of your belief that the note is flat piggybacks, as it were, on the principle that normally applies to veridical beliefs.[16]

Seeing as a ground of perceptual knowledge

Some of what holds for the justification of perceptual beliefs also applies to perceptual knowledge. Seeing the green field, for instance, normally yields knowledge about the field as well as justified belief about it. This suggests another visual principle, a visual knowledge principle. It might be called an *epistemic principle*, since it states a condition for the visual generation of *knowledge*: *At least normally, if we see that a thing (such as a field) has a property (say is rectangular), we (visually) know that it has it*. A parallel principle holds for objectual seeing: At least normally, if I see something to have a property (say to be rectangular), I know it to have the property.

There are, however, special circumstances that explain why these epistemic principles may have to be restricted to "normal" cases. It may be possible to see that something is so, believe on that basis that it is, and yet not know that it is. Charles's case seems to show this. For if, in the kind of circumstances he is in, he often takes a black bird to be blue, then even if he sees that a certain blue bird is blue and, on that basis, believes it is blue, he apparently does not know that it is.[17] He might as well have been wrong, one wants to say; he is just lucky that this time his belief is true and he was not hallucinating. As he has no reason to think he has been hallucinating, and does not realize he has been, one cannot fault him for holding the belief that the bird is blue or regard the belief as inappropriate to his situation. Still, knowledge apparently needs better grounding than is provided by his blameless good fortune. This

kind of case has led some philosophers to maintain that when we know that something is so, our being right is not *accidental*.

There is an important difference here between knowledge and justification. Take knowledge first. If Charles is making errors like this, then even if he has no idea that he is and no reason to suspect he is, he does not know that the bird he believes to be blue is blue. But even if he has no idea that he is making errors, or any reason to suspect he is, he *may* still justifiedly believe that the bird is blue. The main difference between knowledge and justification here may be this: he can have a true belief that does not constitute knowledge because there is something wrong for which he is in no way criticizable (his errors might arise from a handicap which he has no reason to suspect, such as sudden color blindness); but he cannot have a true yet unjustified belief without being in some way criticizable. The standards for knowledge, one might say, permit fewer unsuspected weaknesses in discriminating the truth than those for justification, if the standards for knowledge permit any at all.

This difference between knowledge and justification must be reflected in the kinds of principles that indicate how justification, as opposed to knowledge, is generated. Justification principles need not imply that the relevant basis of a belief's justification assures its truth; but since a false belief cannot constitute knowledge, epistemic principles (knowledge principles) cannot capture elements that generate knowledge unless they rule out factors that might produce a false belief. A ground of knowledge must, in *some* way, suffice for the truth of the proposition known; a ground of justification must, in some way, *count toward* the truth of the proposition one is justified in believing, but need not rule out its falsehood.

On the basis of what we see, hear, feel, smell, and taste, we have a great many beliefs, propositional and objectual. There is apparently no good reason to doubt that these perceptual beliefs are commonly justified or that, quite often, they are true and constitute knowledge. But to see that perception is a basis of justification and knowledge is to go only part way toward understanding what perception, justification, and knowledge are. Here the main question is what constitutes perception, philosophically speaking. Until we have a good understanding of what it is, we cannot see in detail how perception grounds belief, justification, and knowledge. These problems cannot be fully resolved in this book, but we can achieve partial resolutions. I want to discuss (further) what perception is first and, later, to illustrate in new ways how it grounds what it does. The next chapter, then—also concentrating on vision—will start by considering some of the major theories of the nature of perception.

Notes

1 Perceiving *of*, perceiving *to be*, and perceiving *that* may also be called perception of, perception to be, and perception that, respectively; but the

second expression is not common, and in that case at least, the *-ing* form usually better expresses what is intended.

2 A related way to see the difference between objectual and propositional beliefs is this. If I believe something to have a property, say a British Airways plane to be a Boeing 777, then this same belief can be ascribed to me using any correct description of that plane, say, as the most traveled plane in the British Airways fleet: to say I believe BA's most traveled plane to be a 777 is to ascribe the same belief to me. This holds even if I do not believe it meets that description—and it can hold even when I cannot understand the description, as a child who believes a tachistoscope to be making noise cannot understand 'tachistoscope'. By contrast, if I have a propositional belief, say that the United Airlines plane on the runway is the most traveled in its fleet, this ascription cannot be truly made using just any correct description of that plane, say the plane on which a baby was delivered on Christmas Day, 2001. I may have no inkling of that fact—or may mistakenly think it holds for a BA plane. A rough way to put part of the point here is to say that propositional beliefs about things are about them *under a description or name*, and objectual beliefs about things are not (even if the believer could describe them in terms of a property they are believed to have, such as being noisy). It is in part because we need not conceptualize things—as by thinking of them under a description—in order to have objectual beliefs about them that those beliefs are apparently more basic than propositional ones.

3 In terminology common in epistemology, objectual belief is *de re*—of the thing—whereas propositional belief is *de dicto*—of the proposition—and I am similarly distinguishing between objectual and propositional perception. The objectual cases, unlike the propositional ones, require no particular concept of the thing perceived. To be sure, those who do have the concept of a field and know that I believe it to be rectangular may say, 'He believes the field is rectangular', meaning that I believe it *to be* rectangular. English idiom is often permissive in this way, and in everyday life nothing need turn on the difference. Moreover, some philosophers have held that a thing, such as a field, can be a constituent in a proposition—in which case it might be considered a kind of content of a belief of that proposition—and this might provide a basis for saying that the two belief ascriptions may be properly interchangeable. I am ignoring that controversial and uncommon conception of a proposition. For detailed discussion of the extent to which perception is conceptual and of how it yields perceptual beliefs, see Michael Pendelbury, 'Sensibility and Understanding in Perceptual Judgments', *South African Journal of Philosophy* 18, 4 (1999), 356–69.

4 It may be best to leave open here that Susie could, at least for a moment, believe (in an admittedly weak sense of the term), of a tachistoscope, that it is making noise, yet not believe any proposition about it: she *attributes* noise-making to it, yet does not conceptualize it in the way required for

having a propositional belief about it, the kind of belief expressed in a complete declarative sentence such as 'The thing on the table is making noise'. She would then have no propositional belief about the instrument, the kind of belief that should unqualifiedly be called true (or false), such as that the tachistoscope is making noise. On this approach, what I am calling objectual belief is (or often is) better called *property attribution*. It is an attribution *to* the thing in question because of the kind of causal role that thing plays in grounding the attribution; and if it is not strictly speaking a belief, it does imply a disposition to form one, such as that the thing on the table is making noise.

5 Specifically, these are *doxastic* attitudes (from the Greek *doxa*, for 'belief'). A fear can be propositional and thereby cognitive, but it need not entail believing the proposition one fears is so, for example that the man approaching one will attack. Some might consider objectual awareness, say awareness of perfect symmetry, cognitive, at least when the person has the concept of relevant property. By contrast, desires, the paradigm *conative* attitudes, should not, I think, be taken to have propositional objects (e.g. 'to swim' in 'my desire to swim' does not express a truth or falsehood).

6 Perceptions that embody beliefs in the ways illustrated are also called *epistemic*, since the embedded belief is commonly considered to constitute knowledge. Their connection with knowledge is pursued in this chapter and others.

7 The distinction between simple and propositional perceiving and other distinctions drawn in this chapter are not always observed. At one point W.V. Quine says:

> think of "*x* perceives *y*" rather in the image of "*x* perceives that *p*". We say "Tom perceives the bowl" because in emphasizing Tom's situation we fancy ourselves volunteering the observation sentence "Bowl" rather than "Surface of a bowl," "Front half of a bowl," "Bowl and background," and so on. When we ask "What did he perceive?" we are content with an answer of the form "He perceived that *p*".
>
> (*Pursuit of Truth*, revised edn [Cambridge, MA: Harvard University Press, 1992], p. 65)

Notice that because seeing that (say) there is a bowl in front of one obviously entails seeing a bowl, it is no surprise that we are content with a report of the propositional perception even if we wanted to know only what object was seen: we get what we sought and more. It does not follow that simple seeing *is* or even entails propositional seeing. It is also worth noting that Quine is apparently thinking only of seeing here; for the other four senses, there is less plausibility in maintaining what he does.

8 The adage could not be taken to refer to simple seeing, for what we simply

see, say a glass or leaf or field, is not the sort of thing that can be believed (to be true or false). To be sure, seeing something, especially something as striking as golf ball-size hail, produces a *disposition to believe* certain propositions, say that this is a dangerous storm. But, by what seems an economy of nature, there are many things we are disposed to believe but do not. I have defended these points in detail in 'Dispositional Beliefs and Dispositions to Believe', *Noûs* 28 (1994), 419–34.

9 This applies even to full-scale philosophical dictionaries written by teams of experts, though such a work can provide concise statements of much valuable information. See, for example, the entries on blind sight and perception in Robert Audi (ed.), *The Cambridge Dictionary of Philosophy* (Cambridge: Cambridge University Press, 1995, 1999).

10 In the light of what has been said in this chapter so far we can accommodate much of what is plausible in the common view that, as D.M. Armstrong puts it:

> [perception] is an acquiring of knowledge or belief about our physical environment (including our own body). It is a flow of information. In some cases it may be something less than the acquiring of knowledge or belief, as in the cases where perceptions are entirely discounted or where their content has been confidently anticipated.
>
> (*Belief, Truth and Knowledge* [Cambridge: Cambridge University Press, 1973], p. 22)

First, I can agree that perception entails acquisition of *information*; the point is that *not all our information is possessed as the content of a belief.* Second, Armstrong himself notes an important way in which perception might fail to produce belief: it is "discounted," as, for example, where one is sure one is hallucinating and so resolutely refuses to accept any of the relevant propositions.

11 This is the kind of view developed in detail by Fred Dretske. See esp. *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1981).

12 The notion of normality here is not statistical; it implies that what is not normal calls for explanation. In the world as we know it, exceptions to the normality generalizations I propose seem at least rare; but the point is not that statistical one, but to bring out that the very concepts in question, such as those of seeing and knowing, have a connection in virtue of which explanation is called for if what is normally the case does not occur.

13 A property that something is seen as having need not be a property it actually has; but here seeing *as* is phenomenal, not doxastic. Roughly, the perceptual content represents what the object is like if it in fact has the properties it is seen as having.

14 A detailed discussion of the representationality of perception and the kind of content it has is provided by Fred Dretske in *Naturalizing the Mind* (Cambridge, MA: MIT Press, 1997). He deals with the sense in which perceptual content is external. If, loosely speaking, we call the perceived object the *objectual content* then simple perception obviously has a kind of external content; but as the object is "in" the experience, it might be considered a kind of content, as indeed it may for propositional and objectual perceptions as well. With this idea in mind, it is clear how the perceptually believed propositions themselves may also be conceived as having external content. I have discussed internal and external content in relation to such examples in 'Internalism and Externalism in Epistemology and Semantics', in Mark Timmons, John Greco, and Alfred R. Mele (eds.), *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi* (Oxford: Oxford University Press, 2007). (This responds to a challenge from Timothy Williamson, 'On Being Justified in One's Head', ibid., 106–122)

15 In speaking of justification that prevails, and of overall justification, I have in mind the kind appropriate to a rational person's believing the proposition in question, construed as roughly the kind such that when we believe a true proposition with that kind of justification then (apart from the kinds of case discussed in Chapter 10 that show how justified true beliefs *need not* constitute knowledge) we know it.

16 There are complexities I cannot go into, such as how one's competence figures. I am imagining here someone competent to tell whether a note is flat (hence someone not tone deaf): in general, if we are not competent to tell whether a kind of thing has a property or not, an experience in which it seems to have it may not justify us in believing it does. There is also the question of *what* the belief is about when the "object" is hallucinatory, a problem discussed shortly. Still other problems raised by this justification principle are discussed in Chapter 11 in connection with the controversy between internalism and externalism.

17 If, as is arguable, seeing that it is blue entails knowing that it is, then he does *not* see that it is, though he sees its blue color. But this entailment claim is far from self-evident. Suppose he clearly sees a blue bird and believes it is blue, but does not know that it is because of his frequent hallucinations. A moment before, he hallucinated such a bird; a moment later, he will again; and he realizes his senses have been playing such tricks on him. Still, he cannot help believing this bird is blue and believes that *on the basis of* clearly seeing it and its color in normal light. Might we say that he sees that the bird is blue, but does not know this? We cannot say that he "can't believe his own eyes," because he does; but if, in the normal way, they show him the truth and he thereby believes it, might he not see it through them?

1

Scientific experiments on animals

INTRODUCTION

How should human beings treat non-human animals? This question is often debated under the heading of 'animal rights' or, rather differently, 'animal liberation' (you might want to liberate animals even if, strictly thinking, you don't think animals can have rights). This may well be, from the perspective of the future, the defining question of our age. Will future human beings look back at contemporary practices of eating meat and using animals for scientific experiments with the horror we have for earlier practices of slavery? Indeed on some views what we do to animals is far worse than what was at least routinely done to slaves.

It is unlikely that we can come to an accurate view about what future generations will think of us. But we can try to come to a view about the correct approach to the ethical question of the treatment of animals. My main task here, however, is not to argue for any particular answer to that question, although I will towards the end of this paper set out some tentative conclusions. Rather I will attempt to argue that one standard way of approaching the moral question of our treatment of non-human animals is unhelpful, and an alternative framework is much more promising both philosophically and for policy debates. My discussion will focus on the use of animals in scientific research, and I will say very little about other practices such as eating animals or hunting them for sport.

THE USE OF ANIMALS IN SCIENTIFIC EXPERIMENTS

Before getting started on philosophical discussion, it is worth looking at some of the details about the use of animals in scientific research. My discussion will focus on animal experimentation in the UK, where, it is sometimes said, the regulations are the most restrictive in the world. Nevertheless, similar considerations also apply elsewhere. In the UK the main legislation is the Animals (Scientific Procedures) Act 1986. Each year the Home Office provides a set of statistics concerning the licences granted for the year. It is not always necessary to obtain a licence, but as the Home Office explains, 'Under this Act any scientific procedure carried out on any living vertebrate animal, or one species of octopus (*Octopus vulgaris*), which is likely to cause that animal pain, suffering, distress or lasting harm is a regulated procedure requiring licence authority' (Home Office 2009, 3). No licence has been granted in the UK for experiments on the great apes (chimpanzees, gorillas and orang-utans) since the current Act of Parliament has been in force.

A licence will only be granted if, in the opinion of the authority, the benefits of the research outweigh the harms, and experimenting on animals is the only feasible way of obtaining the information sought. This is not intended to rule out 'basic' scientific research with no obvious, immediate application, but it must at least be plausible that the experiment will contribute to the scientific enterprise, with possible eventual benefits to human or to animal welfare. Experiments, or other licensed procedures, are divided into four classifications, mild, moderate, severe and unclassified. Unclassified are those where the animal suffers no pain, as, for example, in experiments where it remains unconscious throughout the procedure and never regains consciousness (we will look at such cases in more detail later). Mild, moderate and severe refer to the degree of pain or suffering involved, although how a particular procedure is classified is generally a matter of judgement and experience, as coming up with an operational definition of the boundaries is probably an impossible

task. Relatively few licences are granted for severe procedures, but the majority of licences are for moderate ones.

The sheer numbers of animals involved, however, may come as a surprise. In 2008 licences were granted for 3.7 million procedures, up from about 3.2 million the previous year, but a long way down from the peak of above 5 million in the 1970s. The very great majority of animals used are mice, rats and fish, although together dogs, cats and non-human primates numbered over 11,000. Pigs, turkeys and other farm animals were also used in experiments relating to veterinary medicine. Although many types of experiments are carried out, animals are used particularly for drug discovery and testing. At an early stage animals are used to attempt to establish the effects of particular substances, normally a chemical compound. These compounds are likely either to have been manufactured in the lab or derived from a natural source, often the rain forest, or even the sea bed. Perhaps it is no surprise, given evolution, that nature seems to be a wonderful source of compounds with health-protective properties. Once a desirable effect is detected, and firmly established, the next stage is to test the compound to see if it is safe, or whether it has adverse side effects, prior to testing the substance on human beings.

The scientific use of animals has a long history, especially in dissection. Indeed in the seventeenth century the philosopher Descartes reveals himself in his writings to be an enthusiastic devotee of animal dissection, in order to further his understanding of human anatomy (Descartes 1985 [1637]). It may well have been that Descartes even performed vivisection: experiments on living animals. Vivisection became more common in the eighteenth and nineteenth centuries, and began to spark considerable protest and disquiet. Experimentation on the mass scale we now see began only in the twentieth century with the use of animals anaesthetized with ether and chloroform. The use of animals in research has always been accompanied by protest, although it has been stronger and more vocal, and, indeed, even violently active at certain times (NCB 2005). And the extent to which animal experimentation

takes place is not always made explicit. For example, when you put money in the collecting tin for heart or cancer research there is a high chance that the money you give will pay for experiments on animals. Indeed those who support antivivisection charities may well, unwittingly, in their support of other charities, be funding exactly the work they want to ban. In collecting evidence for the Nuffield Council we found that among the strongest supporters of animal experiments were societies and charities desperately seeking a cure for a severe medical conditions.

MORAL PHILOSOPHY AND POLICY DEBATES: ANIMAL EXPERIMENTATION

In the Introduction to this book I mentioned some of the difficulties in trying to influence public policy on the topic of animal experimentation by means of philosophical argument. Given that philosophers have such radical disagreements among themselves, and their views often have implications that are very far from current practices, philosophical discussions, on their own, are likely to be treated as fairly marginal to practical debates in policy. But I do not for a second want to diminish the effect that philosophers can have on changing the intellectual climate. Peter Singer's arguments for 'animal liberation', for example, have had a huge influence on how these questions are considered and discussed, and as a result of the efforts of Singer and others very significant changes have been made. Many of the types of experiments once taken for granted in the 1960s and 70s, which inflicted great harm and suffering merely to satisfy the curiosity of the researchers, are now outlawed. But one can hardly argue that animals are now liberated, or that the world is on its way there, even if some of the worst abuses have been eliminated. This sets the question of how philosophers can have greater influence, or even any influence at all, in practical areas of policy.

There is also a second background issue that needs to be brought out before going further. In public debates about the ethics of

animal research, two distinct but intermingled questions need to be separated. The first is the scientific question of whether experimenting on animals is a useful way of finding out about human beings: do the animal models 'work'? Some critics say they fail: if chocolate had been safety-tested on beagles it would never have reached the market, so it is said. Apparently a beagle could die if it ate a whole box of dark chocolates. Others take a more nuanced view. One researcher said to me, 'I know an awful lot about pain in rats. I don't know how much I know about pain in human beings'. Hence there is a scientific debate to be had about the efficacy of the science of animal experiments.

It is possible, however, to believe that a well-defined animal experiment can teach us a great deal, yet still be morally wrong. After all, we can imagine numerous experiments on humans that would yield very useful information – the Nazi scientists on trial at Nuremberg did some of them – yet most of us have a strong view that it would be wrong to do this type of experiment on unwilling subjects, however much that would increase our knowledge. Indeed some people think it is wrong to use the information gained in the Nazi experiments, even if significant benefit would come of it (Moe 1984). But the main point is that even if animal experiments work, this doesn't settle the moral question of whether they should be permitted. Conversely, however, if it is shown that the experiments do not work then, for all but the most extreme view, that would be enough to show that the experiments are not only scientifically flawed, but also morally wrong. Hence arguments about the efficacy of experiments can, on their own, only show that experimenting is wrong; they cannot show it is right. Different types of empirical evidence – for example, about the capacity to feel pain – may have greater bearing, of course.

THE STANDARD APPROACH: DEFINING THE MORAL COMMUNITY

One popular way of trying to advance the debate on the morally proper treatment of animals is to try to define what it is about

humans being that makes us 'members of the moral community' and to explore whether this – whatever it is – is also true of at least some animals. Now, there is one obvious proposal that would settle the matter immediately: the critical morally relevant property is 'being a human being' and this would explain why it is that all and only human beings are members of the moral community. Such a view resonates with the often-heard expression that it is 'obvious that human beings are more important than animals'. However the form of this claim is suspiciously like the claims once heard that it is 'obvious that men are more important than women' or that 'whites are more important than blacks'. Rather than statements of the moral obvious they are now, of course, taken to be statements of sexism and racism, and the term 'speciesism' has been coined to make a similar point in the current context (Ryder 1975; Singer 1995). In effect, the challenge is to find why being a human being is so important. Is there a morally significant property that human beings have, and at least some animals do not, which would then justify drawing the bounds of the moral community in such a way that it leaves out those animals we eat, hunt or experiment upon, or in other words treat in ways we would never treat human beings? On this view, the property 'human being' is not sufficient: membership in a species has no moral weight in itself.

We need then, to look for some underlying property to explain why human beings are morally special. To jump ahead, some possible candidates offered by moral philosophers are sentience, autonomy, possession of a conception of the good, capability to flourish, sociability and possession of a life. These are all properties typically held by human beings, and to varying degrees, by animals. Our question, then, is whether any of them provides a criterion for membership of the moral community. An immediate difficulty was pointed out by John Rawls. On the face of it many of these properties come in degrees, but it seems that, as far as the moral community is concerned, you are either a member of it, or not. Hence, Rawls argued, we need what he called a 'range property': one such that

either you have it or you do not. Rawls' example was whether or not a point on a plane was 'inside the circle'. Of course one point could be closer to the centre of a circle than another, but this is not the same as being 'more inside' the circle. Either the point is inside the circle or it is outside (ignoring those points that hit the line). Similarly, it seems, in the current context of trying to draw the boundaries of the moral community we need a property that is either had, or is not had (Rawls 1971, 508).

The first suggestion on the list was 'sentience', to be understood as the capacity to suffer or feel pleasure and pain. Possibly this could be a range property. Of course some creatures may have a capacity to feel differing ranges or intensities of pain and pleasure, but it is not unreasonable to suppose that an entity either has a capacity or fails to do so. As Jeremy Bentham put it, 'the question is not whether they can talk or reason but whether they can suffer' (Bentham 1996 [1781], 283). Yet it is well known that there are problems with this approach in that it gives a rather uncomfortable answer to the question of the boundaries of the moral community. On this view more or less any creature with a nervous system is a member of the moral community. Indeed, there is an interesting echo of this thought in the UK regulations mentioned above. As we saw, a licence is needed to experiment on all vertebrates and the common octopus. Presumably the justification for this is that we know that such creatures have a very clear capability to suffer. However, one obvious, and rather troubling, consequence of the position that a capacity to suffer puts a creature into the moral community is that it would seem to leave a small number of human beings out: those with a seriously malfunctioning nervous system or those in a permanent coma (although perhaps these individuals can be regarded as suffering in other ways).

Some will be very happy to accept sentience as the basis of entry into the moral community, but we should be aware of the very radical consequences of doing so: that there is no moral privilege to human status. This, of course, will be welcomed by many who

object to our current treatment of animals. However, the further implications of such a view are not so clear. Often it is assumed that it entails that animals have rights, on the basis that if human beings have rights and there is no moral distinction between human beings and other sentient animals then such animals must have rights too. Yet Bentham, who, as we saw, is a defender of the view that sentience is what matters, equally famously denied that human beings had rights in any substantial sense (Bentham 1987 [1796]). For Bentham the consequence of drawing the bounds of the moral community in terms of sentience is that other animals have, not rights, but equal weight in the utilitarian calculus with human beings. For this reason, perhaps, Peter Singer, a contemporary utilitarian, named his book *Animal Liberation* rather than *Animal Rights*.

We will return later to the question of rights versus utilitarian aggregation. In the meantime, we should look at a second approach to drawing the bounds of the moral community, which draws more on the Kantian tradition in moral thought. It comes in a number of variants, but all take as the qualifying property for the moral community something like autonomy, will or freedom, which either is, or is based on, some sort of higher-level cognitive functioning, possibly involving the ability to reflect on the thinker's own thoughts. Accordingly it draws the bounds of the moral community much more tightly than the sentience approach, leaving out almost all, if not all, non-human animals. Perhaps a case can be made for great apes and dolphins, but not much more.

While many may be pleased to draw the line in a way that allows us to continue to eat and experiment on animals there are two well-known immediate consequences that should give us pause. First, those creatures that do not have higher-level cognitive functioning are therefore excluded from the moral community. It appears to follow from this that they are, therefore, owed no more concern than inanimate objects. Animals, then, could be treated just as we treat plants or mineral ores, and so on, which is to say without regard

for their own welfare or interests in any respect. This is a notorious consequence of the Kantian view. Kant's own response was that we should treat animals well out of a concern for ourselves, so as not to demean our own moral status (Kant 1997, 212). But this seems to get things exactly the wrong way round. If it were not in some way wrong to treat animals badly it is hard to see why it would be demeaning of our own humanity to treat them so.

The other obvious problem is that, just as with the sentience approach, some human beings would also be left out of account. In this case, though, the problem is much more serious. Babies, adults with severe learning difficulties and those suffering dementia would also be excluded. Babies could, perhaps, be rescued on the basis of potential moral status, but the other categories are much more problematic.

Now other properties have been proposed as possible bases for membership of the moral community, such as sociability or possession of a life, but rather than go through them one by one, we can note it seems unlikely that any of them is capable of generating the 'common-sense' position that we owe moral concern to (many) non-human animals, but we need not treat (all of these) animals the way we treat human beings. Even if we are justified in killing animals for food, few would think that we need show no concern about how they are kept or treated. However, on the approach we are considering, if we think in terms of a 'range' property, then, to put it crudely, you are either in or out, and so the common-sense position that we owe something to animals but not the same as we owe to human beings is unsupportable. On the 'moral-community' approach we should either treat animals as we do human beings, or we have no moral duties to them at all. The fact that the philosophical debate is so polarized in this way is the crux of the matter of why philosophers' views match up so poorly with current policy and regulations, which appears more complex in structure than is typically offered by philosophers. This, of course, cannot be offered as an argument that public policy is right and

philosophical theories wrong. But before we can make progress we should at least try to understand the moral assumptions behind the common-sense view.

AN ALTERNATIVE APPROACH: MORALLY RELEVANT PROPERTIES

The obvious alternative is to deny that we need a range property, but that membership of the moral community is a matter of more or less. Perhaps we need to find some sort of 'sliding-scale' property to explain why we should treat some creatures, such as human beings, in a different way to others, such as mice, which in turn should be treated in a different way to ants. There is, I think, something to this idea, but I think the way it is stated is misleading. First, the notion of 'moral community' is unhelpful and should be abandoned. It suggests a cut-off point: as we said, either you are in or you are out. But once a sliding scale or continuum is adopted there seems no reason for even thinking there is a boundary line that we need to police in some way.

Second, the assumption that we should explain the grounding of moral duties on a single 'sliding-scale' property is probably the wrong way of trying to generate a more complex account. We can see this by reviewing, once again, the properties that we have already mentioned: sentience, autonomy or higher cognitive functioning, possession of a conception of the good, capability to flourish, sociability and possession of a life. Now, some of these could be turned into sliding-scale properties, it is true. But to do so and pick one as the essential property would appear to imply that we should treat the others as being of no moral relevance at all. But this seems hard to justify. Let us consider sentience as an example. Once we know that a creature is capable of feeling pain, how could we not feel morally obliged to take that into account in working out how to treat it? I am not suggesting that this alone justifies an absolute prohibition on causing pain, but rather it would be

inhumanely callous simply to ignore the fact that a creature can feel pain, even if we find reasons to justify experimenting on it.

Having agreed that we must take pain into account, it would seem very strange to conclude that this is all we must do; that we have found the relevant property and that is that. Some animals are capable of higher cognitive capacities, or will by instinct live in groups. Once more it would simply seem wrong to ignore such facts about creatures when deciding how to act towards them. And once more I am not supposing that we should argue from the premise that an animal has higher cognitive capabilities or that it lives in groups to the conclusion that we must treat it in the same way as we treat human beings. Rather the conclusion is much less precise: that we should take this fact about the animal into account when working out how to treat it. This could, of course, mean treating it with as much respect as we do human beings, or, more modestly, housing it in particular types of environment, or ensuring that it has particular forms of stimulation. The argument then, is that we can match concerns about particular types of treatment with particular properties of the creature. The fact that a creature can feel pain is relevant only to the forms of treatment that threaten to cause it pain. The fact that a creature is sociable by nature is relevant only to those issues that bring sociability into play, perhaps in how it is kept or how other creatures might suffer at its absence or distress.

In other words, rather than setting the terms of membership of the moral community and then supposing that membership brings with it full moral concern, we can approach things a different way. Rather we can say that a very wide range of objects in the world have morally relevant features. These objects include humans and animals, and might also include such things as plants, mountains and rivers, although that is not our direct concern here. But the point is that moral agents have a duty to take all morally relevant features into account in their treatment of those objects. A patchwork of morally relevant properties generates a patchwork of potentially

problematic forms of treatment. Rather than suggesting that there is a single 'sliding-scale' property, we can observe that if creatures differ in their morally relevant features, something resembling a sliding scale will be generated. It will not be a smooth graduation, but may involve differences in principle between different types of animals. For example, great apes and dolphins may well be thought to have more substantial morally relevant properties than, say, dogs and rabbits, which would then correlate with how they are now treated as a matter of policy in the UK and a few other countries. But a creature could be 'higher' than another in one respect, but 'lower' in another. The point is not to generate a biological league table, but to ensure that in our treatment of animals we take all their morally relevant properties seriously.

Which features are morally relevant is a matter on which there can be some debate, although, with one very important exception to which I shall return, these are likely to involve rather small-scale controversies. *How* these features should be taken into account is, though, going to be much more a matter of disagreement. This, in fact, is where the action is, and where we can model and understand the real issues driving debates on the ethics of our treatment of animals.

UNDERSTANDING PHILOSOPHICAL DISAGREEMENT

The burden of the argument so far is really to suggest that the standard philosophical debate about animal ethics has, in a sense, painted itself into an impossible corner. By framing the question in terms of possession of what property provides a creature with a passport to the moral community, philosophers have saddled themselves with literally unbelievable consequences. If the condition is sentience then humans are absolute moral equals with all animals with a nervous system. If the condition is higher cognitive functioning, then there are no moral constraints on the permitted treatment of many animals, at least not for the sake of those animals (as distinct

from the human-centred objection that we can demean ourselves in acting in such ways). My suggestion is that, rather, a whole range of properties are morally relevant and should all be taken into account. Yet this does not explain how they should be taken into account, and here, as I have said, we hit the heart of current debates, even though those who take part in the debates may not see things in these terms.

To see the force of the question of how properties can be taken account of in different ways, consider, for example, how some have taken inspiration from the work of Peter Singer, arguing that as animals can feel pain just as human beings can, they should be treated with the same moral concern. Singer's book, called *Animal Liberation*, is sometimes said to be 'the bible of the animal rights movement'. But as noted above, Singer's book is not called *'Animal Rights'*. To say that human beings and animals should be treated the same way is not yet to say what that treatment should be. One could read Singer and conclude that we should start eating other human beings or perform highly invasive experiments upon them. Those who think that Singer has provided a defence of animal rights appear to argue that if a creature is capable of feeling pain then there is an absolute moral requirement not to inflict pain, or allow the infliction of pain, on that creature, perhaps unless it is for its own good.

This is a very strong conclusion. Even if all animals are equal, it is not true that all pains are equal. We sometimes permit the infliction of avoidable pain, or at least discomfort, on human beings for the sake of the greater good. Forms of crowd control are often intensely uncomfortable and in some circumstances, say, standing in a confined space, can be painful. Yet we would sometimes forcibly require this of people if there is an emergency and the police need to clear a space by penning people into a small area. But that aside, the main point is that while Singer and animal rights activists can both agree that the capacity to feel pain is a morally important property, and how we treat a creature should take into account its

possible pain, they do not agree about the moral consequences of the possession of this capacity. The animal rights theorist argues that possession of this property is a sufficient reason to justify an absolute prohibition on action that causes pain, while Singer, as a consequentialist must take a different view. Pain, of course, must be weighed in the consequentialist scales, but there is no reason why it should not regularly be outweighed by other factors.

What matters, then, is how each morally relevant feature is to be taken into account, and the main candidates are whether they generate absolute prohibitions on possible forms of treatment, or whether they are simply properties to be put in the balance to be measured against other factors, such as the prevention of greater pain. Clearly, treating a morally relevant property as generating an absolute (or even near-absolute) prohibition is very close to the idea that the possessor of that property has a right not to have it violated. Whether all theorists will want to draw this conclusion will depend on their theory of rights, but that need not detain us here. By contrast, treating a feature as something to be taken into account, rather than as generating an absolute prohibition, seems to be friendly to the idea that, though weighed in the balance, it can be overturned by other considerations. For example, other things being equal we should not cause mild anxiety in any creature, but when weighed against the possible benefits of a major medical breakthrough, it could be that the consequentialist calculation comes out in favour of permitting the anxiety. Of course, the calculation could also go the other way; much depends on the probabilities of harm and benefit, and the particular weights given to the different factors, which, when we get down to the real details, could differ between theorists even if they accept the same 'high-level' theory.

To illustrate the different ways in which moral factors can be taken into account it is worth exploring the moral underpinnings of the current UK regulations for experiments upon animals, but also broadening the discussion to include experiments on human subjects. As in many other countries, the regulatory regimes

apparently rest on a combination of assumptions. The rule for experiments on human subjects is that no one should be subject to any sort of potentially harmful intervention without their consent. The respect given to the idea of 'informed consent' in medical practice, medical research and other research involving human beings, shows how important this idea is in common currency. It is now, for example, virtually impossible to obtain ethical approval for any experiment on human beings that involves deception, thus making highly problematic a whole range of studies in social psychology and elsewhere. For example, it would no longer be possible to conduct the famous Milgram 'obedience to authority' experiments in which the subjects were tricked into believing they were giving other human beings severe, even fatal, electric shocks (Milgram 1974). Many of the experiments on which the contemporary discipline of social psychology is now founded could not have been done under present regulations; hence informed consent has not always been given the weight it now has. Perhaps in the past deception was considered acceptable if the scientific objectives were deemed sufficiently important. Now, though, informed consent is at the heart of medical ethics and research involving human beings.

Should such regard be extended to creatures presumed to be of high cognitive power, but who do not have human language in which to express their wishes? As we saw, in the UK no licence will be granted to conduct invasive scientific experiments on great apes, for example, although not all countries have followed the UK's lead. But other primates, such as monkeys, and mice, rats, rabbits, dogs, fish and other animals are typically treated as if they do not have the capacity to give or withhold consent, although it certainly appears that dissent can be expressed by most animals, by running away if they get the chance. However, we can sum up the moral assumption in question as: those creatures who are capable of autonomous thought should be given the right to determine whether or not they are subjected to invasive treatment (where 'invasion' is

understood widely, and mere deception would be treated as a form of invasive treatment). Those creatures of high cognitive capacity that do not explicitly consent are presumed to dissent.

The capacity to feel pain and suffering, however, is treated as a rather different matter. On the face of it current UK regulations do not allow severe, prolonged pain to any creature, either as part of a scientific experiment or as a by-product of raising or slaughtering animals for food or other purposes. Hence wherever possible, experiments are conducted with anaesthetized animals, and farm animals are stunned before slaughter. Where pain is impossible to avoid as part of the experiment, as in the obvious example of testing painkillers, scientists do whatever they can to reduce the duration and intensity of pain to the lowest degree possible. The UK regulations seem to make a distinction between severe and prolonged pain and suffering, on the one hand, and milder forms, perhaps of the sort that would normally be part of the daily or weekly experience of any sentient creature. Severe and prolonged pain creates something close to an absolute prohibition, or at least is given very high weight in the consequentialist balancing. Animals have a 'near right', we might say, not to suffer prolonged severe pain as a consequence of human treatment. We see this concern run right through laws and regulations regarding animals, such as the types of traps hunters are allowed to use to catch wild animals. Short-term mild pain, mild suffering or mild distress is treated as an undoubted harm, but one that can be outweighed by other factors. In between is a large grey area – prolonged mild pain, short severe pain, moderate pain of any duration – on which decisions will also need to be made.

Let us move next to the morally relevant feature of 'having a good'. Here the idea is that there are forms of treatment that are good or bad for an animal in the sense of furthering or impeding its flourishing, given the type of animal it is. The most obvious way in which such issues are taken into account as a matter of current practice is by means of the conditions in which animals are

housed. Where possible some semblance of 'natural conditions' will be attempted. For example, foraging creatures will often be placed in environments to allow them to root around in wood shavings or similar material. Animals that live in social groups are often housed together. Once more we see an illustration of how morally relevant features of any animal can be used to determine how that animal should be treated, although the conditions of scientific research and farming mean that the constraint is interpreted in different ways. In the case of scientific research all aspects need to be controlled, and so simulated forms of expressing natural behaviour are likely to be offered, if anything is. This, of course, adds to the cost of scientific research, but in a rather minor way, compared to the immense cost of many experiments and procedures. Furthermore scientists often argue that contented, non-anxious animals make for better scientific subjects, and so in purely scientific terms the money is well spent. In the case of farming, marginal costs are more important as they must be passed on to customers in a highly competitive market, although product differentiation through 'humane' forms of farming is also common, as in the case of 'free-range' eggs. However there is less need to regulate all aspects of a farm animal's life, and so forms of behaviour that allow normal functioning are sometimes possible, although, of course, by no means the norm. But it is rather troubling that in the case of farming we often allow minor cost considerations to outweigh a creature's ability to live a life that is natural to the type of creature it is.

The final feature I shall discuss, which is not to presume that the list is complete, is 'possession of a life'. Here matters are somewhat more dramatic and potentially controversial. As things stand, while, as we have seen, pain and suffering is treated with great seriousness in the UK regulations, death of an animal, by contrast, is treated as a 'humane end point'. A research scientist, a farmer or slaughterhouse worker has to learn to accept the death of an animal as morally unproblematic, provided it happens the right way. The experiences that happen within a life, such as pain and

suffering, are one thing; the life is quite another, and is treated as of little, if any, moment.

Take, for example, the experimental procedure known as 'anaesthesia without recovery', which is used to try to determine the immediate effects of an experimental substance (often referred to as a 'compound') on an animal, usually a rabbit, mouse or rat. In one version of this type of procedure, the animal is fitted with two catheters, one to deliver the anaesthetic, the other to deliver the compound under test. First, the animal is rendered unconscious by means of the anaesthetic. Next the animal is subjected to a series of radical surgical incisions to expose its internal organs. Its skin is peeled back and then it is pinned out on the laboratory bench to keep it immobile and its internal organs exposed. Various probes are then placed on its organs and blood vessels, in order to monitor their state during the experiment. The compound is then introduced and precise observations can be made concerning its effects on blood flow, metabolism or whatever else is under investigation. The anaesthetic is topped up as necessary to keep the animal unconscious and hence not in pain. When the experiment is concluded, after a matter perhaps of some hours, the dose of anaesthetic is increased until the animal dies: a 'humane end point', without pain.

Such experiments are not uncommon, although being labour-intensive they are comparatively expensive. But, of course, they are far from the only method by which animals are killed for human purposes, whether within science, agriculture or recreation. It appears that current regulations in all these areas assume that an animal's life in itself has no, or very little, value, again as distinct from the experiences which happen within a life, which can have both positive and negative value. (Note that it is assumed that the preservation of a species is of great value, but this is not to say that such value is somehow 'spread out' among all the members of that species.)

Now, if human beings and other animals all have lives, yet human life is treated with reverence while animal life is treated as

of no value, then, prima facie, there is an unjustified partiality in favour of human beings, unless some relevant points of differentiation can be made to explain the difference. In response it should be noted that some philosophers argue that for human beings it is not true that life, itself, has value for the person whose life it is. Of course, family, friends, even admirers, can find value in the life of that other person – in their company, support or achievements, and so on. But none of this entails that the life itself has value for its possessor.

Still, it seems very odd to deny that life has value for human beings, especially in the face of the evidence that the overwhelming majority of human beings take steps to prolong their lives as long as they can. But perhaps this can be understood in terms of the hopes, desires and fears of the agent concerning his or her future. Perhaps, then, it is the continuation of these plans and projects that have value rather than the life itself. Someone with no plans, no sources of enjoyment, no family and no friends, and no prospect of any of these things, may see little or no value in the continuation of his or her life. And, indeed, although it may sound strange to say such a thing, the thesis that life has no value in itself is not an unfamiliar one. For example it is sometimes said that to believe the alternative – that life does have value – entails a duty to create as much life as possible, which seems an unappealing doctrine. Whether this really does follow appears to rest on some fairly strong assumptions concerning the relationship between value and duty. But be that as it may, we can draw some conclusions. First, current regulations concerning animal use assume that life, if it is of value at all, has weak value which is easily outweighed by other factors. Second, such an assumption does not necessarily show that there is an unjustified partiality in our practices in favour of human beings, for it is possible that the apparent extra value given to human life derives from other recognized sources of value, such as the value of the plans of the person or their desires or their place in a wider social networks, including mutual relationships of care, and so on.

THE PROBLEMATIC STATUS OF HUMAN TREATMENT OF ANIMALS

So far, though, I have done no more than describe how to understand disagreement about human use of animals, and to consider the moral assumptions most likely to underlie contemporary regulations. Roughly the position is this. First, creatures with high cognitive capacity have strong rights against interference. Second, sentient creatures have a 'near right' not to be subjected to pain that is both severe and prolonged, while more moderate pain and suffering are a matter of concern and are taken note of within a consequentialist calculation. Finally, sociability and the possession of a good are also taken account of to some lesser degree, but little or no weight is given to life itself. I have not, however, said whether I think that any such view is defensible. For my own part, I simply do not know whether to accept that life has no value, whether it is the life of a human being, a non-human animal or a plant. Other things being equal, it is a worse world if a living creature or plant perishes, but this could be explained by a wide range of factors. If there is value to life itself, then there is something to regret in the death of any animal in scientific research or in farming. Equally, and this sounds less plausible, there is something to regret in the death of every plant, including those grown as an annual crop, although it is easier to see the loss in the death or destruction of, say, a large tree or well-established bush.

But leaving aside the case of plants, if there is value in animal life, then the moral assumptions underlying current regulations are questionable. As we have seen, there seems no recognition of the (possible) value of life. But perhaps even more problematic is the one-sided consequentialism of the regulations. Costs are weighed against benefits, and a judgement is made as to whether the benefits sufficiently outweigh the costs. Even if it is right that there is need for a balance between various factors, and that it can be acceptable to allow small and moderate pains, or confinement and restrictions on natural behaviour, for the sake of great possible

benefits, there is something deeply suspicious about the fact that the costs all fall on non-humans while the benefits rebound to human beings. Even when experiments are undertaken to devise pharmaceuticals for use on animals, the benefits are usually sought for human purposes, and the particular individual animals that suffer do not also benefit. Generally when suffering is systematically imposed on one group or individual for the benefit of another group or individual we consider it to be, at least, exploitative. It is hard, therefore, to escape the charge that human beings exploit animals in a way that is problematic.

A common response is to argue that the animals who suffer also benefit in a quite different way. After all, in the vast majority of cases the particular animals in question would not exist at all if it were not for human purposes. Most animals used by humans are bred specifically for such purposes and hence if life has a value this value is conferred on any animal brought into existence. In addition, or alternatively, the lives of many animals used by human beings may contain a significant balance of positive experience over negative to make that life worthwhile, and hence the charge of exploitation is somewhat mitigated (Scruton 2000).

While this defence may be appropriate in some cases, the fact that so few animals used by human beings are allowed to live to an age where they die a natural death should give us pause. Consider the way in which animals such as dogs and monkeys are used in safety tests. After a short exposure – a few weeks or months – to the compound under trial they will be 'euthanized' so that a post-mortem can take place to see whether their internal organs have been affected. These are animals otherwise in the prime of life (and this is why they have been chosen for the experiments). Consider also the example of anaesthesia without recovery described earlier. Think too, of food production. Sheep on the hills seem to have a rather utopian life. Yet the lamb we generally eat, as distinct from the adult sheep kept for their wool, or for breeding, are slaughtered when between a month and a year old: presumably when the

combination of their size and tenderness brings greatest economic rewards. A simple awareness of the facts can make many people, including myself, very sympathetic to the moral case for ending animal experiments and for vegetarianism.

Yet the peculiarity is that, for me at least, I do not find that such arguments, however intellectually compelling, of very strong motivational force. I still use medicines and household products that have been tested on animals. I do not protest against the animal experiments that take place in my university. I continue to eat meat, albeit rather guiltily. The moral philosopher R.M. Hare would respond to this combination of professed belief and action by arguing that my claims are insincere (1952). Hare argued that sincere moral belief is always expressed in action, but as my actions don't follow my claimed beliefs, any moral argument I make against harmful treatment of animals is necessarily insincere. This seems to me, however, dogmatic and uncompelling. Phenomenologically, it seems to me that moral argument hits hardest at the level of *conscience*, and whether it spurs *action* is a further issue. Partly, I think, we must take into account the consequences of acting on one's moral beliefs. Where it is costly, or even awkward or inconvenient, to act as your conscience prompts, individuals may find themselves acting in ways which they, themselves, at some level disapprove of. Consider, by way of analogy, the existence of slavery in the American south in the nineteenth century. I find it very hard to believe that every slave owner sincerely believed that there was nothing wrong with the practice of one human purchasing another and holding arbitrary power over him or her. No doubt many did think that somehow it was in the natural order of things, but surely some had their doubts? These 'guilty masters' may well have accepted the moral argument that no one should be a slave of another, yet did not seriously consider liberating his or her slaves, believing that there was no other way of surviving at an acceptable standard of living. By analogy many of us are unwilling to give up the benefits of what, in reflective moments, we take to be morally unacceptable

uses of animals, as doing so would make our lives more inconvenient and uncomfortable.

If the actions we choose to pursue are, we believe, morally unjustified, then we have a choice. We can live with the apparent hypocrisy, or change our way of life or adjust our moral beliefs. Politically or structurally, however, there is a further option: institutional or technical advance which allows us to pursue our ends without accepting behaviour we believe to be unjustified. Presumably slavery was easier to abolish when it became clear that it had become economically possible to remain in business without slaves. By analogy, if we can find ways of producing non-animal foods which are just as delicious and nutritious as meat, or ways of testing pharmaceuticals that do not involve animals, then we can continue to pursue the ends we seek without acting in ways that are morally troubling. In effect such an ambition is to solve the moral problem by avoiding it.

In the case of animal experimentation the leading suggestion in the direction of avoidance is the doctrine of the 'three Rs' proposed by Russell and Burch (1959). The three Rs are refinement, reduction and replacement. Refinement is the idea that experiments should be modified so that they are as little harmful as they could be to animals. Reduction, naturally enough, calls for a reduction in the number of animals used. Replacement is the idea that the knowledge sought by experimenting on animals might be achieved in some other way.

Refinement and reduction are generally welcomed by scientists. After all, if you could reduce the pain to animals, or the number of animals involved, without compromising the science, what could be the objection? Replacement is rather more complicated. In some cases it is a matter of conducting experiments on cells or tissues that have been cultivated in the lab – so-called *in vitro* experiments. These can yield useful knowledge but at the moment tend to be used only at an early stage of research. Another possibility is computer modelling, but this is in its infancy, and at the moment seems very

limited, as computer models will always be a simplification of the real world. But there are other methods of replacement. Consider, for example, experiments conducted in the 1960s in which infant monkeys were removed from their mothers' care to see how they responded. The scientists conducting some of these experiments claimed that they were hoping to gain insight into human depression (Kaufman and Rosenblum 1967). But it seems that such experiments are scientifically very strange, even independently of the ethical questions. For surely psychological and sociological studies would be a much more effective way of obtaining this sort of information. More generally, sometimes a research question might better be addressed by social science or statistical analysis than by animal experiments. This is likely to make experimental scientists nervous: nervous of redundancy. Replacement is the goal of those who oppose animal experiments, but, of course, it is hardest of the three Rs to crack.

PROGRESS IN PUBLIC POLICY

In the present context we can split moral attitudes to the human use of animals in scientific experiments roughly into three groups. First, there are those who think that harms to animals are, on balance, outweighed by the scientific and medical benefits such experimentation allows. Second, there are those who see irrefutable arguments on both sides, and conclude that this is a genuine moral dilemma with no clear solution. Third, there are those who feel that the moral considerations show that we are wrong to use animals in scientific research. Of these groups the first find their views most clearly reflected in current regulations. But still, partially in deference to the second and third groups, changes are regularly being made. In essence the regulators have taken their question to be how we can modify our treatment of animals so that more and more people find it less and less objectionable. In other words, the regulators are attempting to bring together what, in the Introduction,

we called, following Rawls, an overlapping consensus. Think of the changes made over recent decades: much more attention to animal welfare in farming and slaughtering, banning of animal testing for cosmetics and household products, banning of tests on great apes, and seeking out alternatives to animal experiments through test-tube and computer models. Although it would be wrong to say that such moves have fully satisfied those who argue against human use of animals, they must see them as steps in the right direction. At the same time those who have found their activities limited by such new regulations have not, in general, had to give up very much. In general they can adapt to the new situation. Hence there appears to be moral progress in public policy at relatively little cost, by a series of concessions around the edges. But equally clearly, in this case, there remains work to be done.

This is not to say that radical, discontinuous, change is impossible. Slavery was abolished. Neither must there always be a consensus behind change. Often change is highly contentious or unpopular. The banning of hunting with dogs is an example, although as we will explore later, changing law in absence of widespread agreement can create important problems of compliance. In the case of human treatment of animals in research and farming, for the moment the best we can hope for in the short to medium term is to make current practices less objectionable to more people.

CONCLUSION: LESSONS FOR PHILOSOPHY

This chapter has, I hope, illustrated in detail a point made briefly in the Introduction to this book. Approaching a problem in public policy by means of the methodology 'first choose your theory', as if you were signing up for some sort of crusade, could lead to interesting philosophical consequences but it is very unlikely to lead to a usable contribution to current policy debates. Of course as I hope I have also made clear, radical philosophical arguments are a vital part of the debate, and add to the stock of ideas that enrich discussion. But

on their own they will settle nothing. The methodology implicitly recommended here suggests that when thinking about a practical issue, we should start at the other end: not at the philosophical theories but current disagreements in the public policy area. We need to ask: what do people think they disagree about? And is that the best way of understanding their disagreement? Is there a better way? And if so, does that open up new avenues for making progress? It is a commonplace to say that philosophers can help clarify the terms of public debate. Philosophers are not the only people who can do this, of course, but it is part of our training to make distinctions, to follow arguments out to their conclusions, and to reconstruct relatively loose arguments in a more rigorous form. But to do this one has first to become immersed in the debate in which one wishes to intervene.

It is implied in the methodology I am suggesting that participants in public debates do not always fully comprehend or perfectly articulate what they disagree about. A simple slogan or principle, while helpful for campaigning, can have a distorting effect on argument. In my view, contemporary public policy debates about human treatment of animals are not, centrally, debates about whether animals have rights, or whether all animals are equal. They are not even debates about which properties of creatures are morally relevant. Rather they are debates about how morally relevant properties of animals should be taken into account in human treatment of them.

I have also made a further point about moral argument and human motivation. Here, I have to admit, I have made a claim that I have tried to illustrate, rather than demonstrate. The claim is that moral argument is much better at making people feel guilty about what they do, rather than changing their behaviour, and if this is true it has implications about the type of structural change that is likely to be needed to meet social objectives. We will return to questions about motivation, though, in later chapters.

CHAPTER 1
Introduction

1.1 Logic

Logic[1] is *the analysis and appraisal of arguments*. Here we'll examine reasoning on philosophical areas (like God, free will, and morality) and on other areas (like backpacking, water pollution, and football). Logic is a useful tool to clarify and evaluate reasoning, whether on deeper questions or on everyday topics.

Why study logic? First, logic builds our minds. Logic develops analytical skills essential in law, politics, journalism, education, medicine, business, science, math, computer science, and most other areas. The exercises in this book are designed to help us think more clearly (so people can better understand what we're saying) and logically (so we can better support our conclusions).

Second, logic deepens our understanding of **philosophy** – which can be defined as *reasoning about the ultimate questions of life*. Philosophers ask questions like "Why accept or reject free will?" or "Can one prove or disprove God's existence?" or "How can one justify a moral belief?" Logic gives tools to deal with such questions. If you've studied philosophy, you'll likely recognize some of the philosophical reasoning in this book. If you haven't studied philosophy, you'll find this book a good introduction to the subject. In either case, you'll get better at recognizing, understanding, and appraising philosophical reasoning.

Finally, logic can be fun. Logic will challenge your thinking in new ways and will likely fascinate you. Most people find logic enjoyable.

1.2 Valid arguments

I begin my basic logic course with a multiple-choice test. The test has ten problems; each gives information and asks what conclusion necessarily follows. The problems are fairly easy, but most students get about half wrong.[2]

[1] Key terms (like "**logic**") are introduced in bold. Learn each key term and its definition.

[2] Http://www.harryhiker.com/logic.htm has my pretest in an interactive format. I suggest that you try it. I developed this test to help a psychologist friend test the idea that males are more logical than females; both groups, of course, did equally well on the problems.

Here are two of the problems – with the right answers boxed:

If you overslept, you'll be late.
You aren't late.
Therefore:
(a) You did oversleep.
(b) You didn't oversleep.
(c) You're late.
(d) None of these follows.

If you overslept, you'll be late.
You didn't oversleep.
Therefore:
(a) You're late.
(b) You aren't late.
(c) You did oversleep.
(d) None of these follows.

While almost everyone gets the first problem right, many wrongly pick "(b)" for the second problem. Here "You aren't late" doesn't necessary follow, since you might be late for another reason; maybe your car didn't start.[1] The pretest shows that untrained logical intuitions are often unreliable. But logical intuitions can be developed; yours will likely improve as you work through this book. You'll also learn techniques for testing arguments.

In logic, an **argument** is a set of statements consisting of premises (supporting evidence) and a conclusion (based on this evidence). Arguments put reasoning into words. Here's an example ("∴" is for "therefore"):

Valid argument ➔
If you overslept, you'll be late.
You aren't late.
∴ You didn't oversleep.

An argument is **valid** if it would be contradictory (impossible) to have the premises all true and conclusion false. "Valid" doesn't say that the premises *are* true, but only that the conclusion *follows from* them: if the premises were all true, then the conclusion would have to be true. Here we implicitly assume that there's no shift in the meaning or reference of the terms; hence we must use "overslept," "late," and "you" the same way throughout the argument.[2]

Our argument is valid because of its *logical form*: how it arranges logical notions like "if-then" and content like "You overslept." We can display the form using words or symbols for logical notions and letters for content phrases:

If you overslept, you'll be late.	If A then B	**Valid**
You aren't late.	Not-B	
∴ You didn't oversleep.	∴ Not-A	

Our argument is valid because its *form* is correct. Replacing "A" and "B" with other content yields another valid argument of the same form:

[1] These two arguments were taken from Matthew Lipman's fifth-grade logic textbook: *Harry Stottlemeier's Discovery* (Caldwell, NJ: Universal Diversified Services, 1974).

[2] It's convenient to allow arguments with zero premises; such arguments (like "∴ x = x") are valid if and only if the conclusion is a necessary truth (couldn't have been false).

If you're in France, you're in Europe.	If A then B	**Valid**
You aren't in Europe.	Not-B	
∴ You aren't in France.	∴ Not-A	

Logic studies forms of reasoning. The content can deal with anything – backpacking, math, cooking, physics, ethics, or whatever. When you learn logic, you're learning tools of reasoning that can be applied to any subject.

Consider our **invalid** example:

If you overslept, you'll be late.	If A then B	**Invalid**
You didn't oversleep.	Not-A	
∴ You aren't late.	∴ Not-B	

Here the second premise denies the *first* part of the if-then; this makes it invalid. Intuitively, you might be late for some other reason – just as, in this similar argument, you might be in Europe because you're in Italy:

If you're in France, you're in Europe.	If A then B	**Invalid**
You aren't in France.	Not-A	
∴ You aren't in Europe.	∴ Not-B	

1.3 Sound arguments

Logicians distinguish *valid* arguments from *sound* arguments:

> An argument is **valid** if it would be contradictory to have the premises all true and conclusion false.
>
> An argument is **sound** if it's valid and every premise is true.

Calling an argument "valid" says nothing about whether its premises are true. But calling it "sound" says that it's valid (the conclusion follows from the premises) *and* has all premises true. Here's a *sound* argument:

Valid		If you're reading this, you aren't illiterate.
and true	➔	You're reading this.
premises		∴ You aren't illiterate.

When we try to prove a conclusion, we try to give a *sound* argument: valid and true premises. With these two things, we have a sound argument and our conclusion has to be true.

An argument could be unsound in either of two ways: (1) it might have a false premise or (2) its conclusion might not follow from the premises:

First premise false:	*Conclusion doesn't follow:*
All logicians are millionaires.	All millionaires eat well.
Gensler is a logician.	Gensler eats well.
∴ Gensler is a millionaire.	∴ Gensler is a millionaire.

When we criticize an opponent's argument, we try to show that it's *unsound*. We try to show that one of the premises is false or that the conclusion doesn't follow. If the argument has a false premise or is invalid, then our opponent hasn't proved the conclusion. But the conclusion still might be true – and our opponent might later discover a better argument for it. To show a view to be false, we must do more than just refute an argument for it; we must give an argument that shows the view to be false.

Besides asking whether premises are true, we can ask how certain they are, to ourselves or to others. We'd like our premises to be certain and obvious to everyone. We usually have to settle for less; our premises are often educated guesses or personal convictions. Our arguments are only as strong as their premises. This suggests a third strategy for criticizing an argument; we could try to show that one or more of the premises are very uncertain.

Here's another example of an argument. In fall 2008, before Barack Obama was elected US president, he was ahead in the polls. But some thought he'd be defeated by the "Bradley effect," whereby many whites *say* they'll vote for a black candidate but in fact don't. Barack's wife Michelle, in an interview with Larry King, argued that there wouldn't be a Bradley effect:

> Barack Obama is the Democratic nominee.
> If there's going to be a Bradley effect, then Barack wouldn't be the nominee [because the effect would have shown up in the primaries].
> ∴ There isn't going to be a Bradley effect.

Once she gives this argument, we can't just say "Well, my opinion is that there *will* be a Bradley effect." Instead, we have to respond to her reasoning. It's clearly valid – the conclusion follows from the premises. Are the premises true? The first premise was undeniable. To dispute the second premise, we'd have to argue that the Bradley effect would appear in the final election but not in the primaries. So this argument changes the discussion. (By the way, there was no Bradley effect when Obama was elected president a month later.)

Logic, while not itself resolving substantive issues, gives us intellectual tools to reason better about such issues. It can help us to be more aware of reasoning, to express reasoning clearly, to determine whether a conclusion follows from the premises, and to focus on key premises to defend or criticize.

Logicians call statements *true* or *false* (not *valid* or *invalid*). And they call arguments *valid* or *invalid* (not *true* or *false*). While this is conventional usage, it pains a logician's ears to hear "invalid statement" or "false argument."

Our arguments so far have been **deductive**. With **inductive** arguments, the conclusion is only claimed to follow with probability (not with necessity):

Deductively valid	*Inductively strong*
All who live in France live in Europe.	Most who live in France speak French.
Pierre lives in France.	Pierre lives in France.
∴ Pierre lives in Europe.	That's all we know about the matter.
	∴ Pierre speaks French (probably).

The first argument has a tight connection between premises and conclusion; it would be impossible for the premises to all be true but the conclusion false. The second has a looser premise–conclusion connection. Relative to the premises, the conclusion is only a good guess; it's likely true but could be false (perhaps Pierre is the son of the Polish ambassador and speaks no French).

1.4 The plan of this book

This book starts simply and doesn't presume any previous study of logic. Its four parts cover a range of topics, from basic to rather advanced:

- Chapters 2 to 5 cover syllogistic logic (an ancient branch of logic that focuses on "all," "no," and "some"), meaning and definitions, informal fallacies, and inductive reasoning.
- Chapters 6 to 9 cover classical symbolic logic, including propositional logic (about "if-then," "and," "or," and "not") and quantificational logic (which adds "all," "no," and "some"). Each chapter here builds on previous ones.
- Chapters 10 to 14 cover advanced symbolic systems of philosophical interest: modal logic (about "necessary" and "possible"), deontic logic (about "ought" and "permissible"), belief logic (about consistent believing and willing), and a formalized ethical theory (featuring the golden rule). Each chapter here presumes the previous symbolic ones (except that Chapter 10 depends only on 6 and 7, and Chapter 11 isn't required for 12 to 14).
- Chapters 15 to 18 cover metalogic (analyzing logical systems), history of logic, deviant logics, and philosophy of logic (further philosophical issues). These all assume Chapter 6.

Chapters 2–8 and 10 are for basic logic courses, while other chapters are more advanced. Since this book is so comprehensive, it has much more material than can be covered in one semester.

Logic requires careful reading, and sometimes rereading. Since most ideas build on previous ideas, you need to keep up with readings and problems. The companion LogiCola software (see Preface) is a great help.

1 The Problem of Universals I

Metaphysical Realism

Overview

The phenomenon of similarity or attribute agreement gives rise to the debate between realists and nominalists. Realists claim that where objects are similar or agree in attribute, there is some one thing that they share or have in common; nominalists deny this. Realists call these shared entities universals; they say that universals are entities that can be simultaneously exemplified by several different objects; and they claim that universals encompass the properties things possess, the relations into which they enter, and the kinds to which they belong.

Toward showing us that we must endorse the reality of universals, realists point to the phenomena of subject-predicate discourse and abstract reference. They claim that unless we posit universals as the referents of predicate expressions, we cannot explain how subject-predicate sentences can be true, and they argue that we can explain the truth of sentences incorporating abstract referring terms only if we take universals to be the things identified by the use of those terms.

Realists, however, frequently disagree about the generality of their accounts of predication and abstract reference. Some realists, for example, deny that their account of predication holds for sentences incorporating the term 'exemplifies.' Other realists insist that their account holds only for primitive or undefined predicates or abstract terms. Furthermore, some realists hold that there are universals corresponding only to predicates that are actually true of existing objects, whereas other realists believe that there are both exemplified and unexemplified properties, kinds, and relations.

Realism and Nominalism

The objects we talk and think about can be classified in all kinds of ways. We sort things by color, and we have red things, yellow things, and blue things. We sort them by shape, and we have triangular things, circular things, and square things. We sort them by kind, and we have elephants, oak trees, and paramecia. The kind of classification at work in these cases is an essential component in our experience of the world. There is little, if anything, that we can think or say,

little, if anything, that counts as experience, that does not involve groupings of these kinds. Although almost everyone will concede that some of our ways of classifying objects reflect our interests, goals, and values, few will deny that many of our ways of sorting things are fixed by the objects themselves.[1] It is not as if we just arbitrarily choose to call some things triangular, others circular, and still others square; they *are* triangular, circular, and square. Likewise, it is not a mere consequence of human thought or language that there are elephants, oak trees, and paramecia. They come that way, and our language and thought reflect these antecedently given facts about them.

There are, then, objective similarities among things. Prior to our classifying them in the ways we do, the familiar objects of the everyday world agree in their characteristics, features, or attributes. This is not a claim born of any metaphysical theory. It is, on the contrary, a prephilosophical truism, but one that has given rise to significant philosophical theorizing. Indeed, a question that goes back to the origins of metaphysics itself is whether there is any general explanation for the prephilosophical truism that things agree in attribute. Suppose it to be a fact that certain objects agree in attribute; they are all, say, yellow. Is there some fact more basic or fundamental than this fact, such that it is because and only because the more fundamental fact holds of these objects that they are all yellow? And if there is, is it possible to generalize from this case? That is, is there a very general type or form of fact such that, given any case of attribute agreement, that case obtains because and only because some fact of the relevant, very general type or form obtains?

An affirmative answer to this question is suggested in Plato's *Parmenides*, where we read that:

> there exist certain Forms of which these other things come to partake and so to be called after their names; by coming to partake of Likeness or Largeness or Beauty or Justice, they become like or large or beautiful or just.[2]

What is being proposed here is a general schema for explaining attribute agreement. The schema tells us that where a number of objects, *a . . . n*, agree in attribute, there is a thing, φ, and a relation, *R*, such that each of *a . . . n* bears *R* to φ, and the claim is that it is in virtue of standing in *R* to φ that *a . . . n* agree in attribute by being all beautiful, or just, or whatever. It turns out that many philosophers since Plato have found this schema attractive.[3] They have not always used Plato's language. Where he speaks of things partaking of a Form, they have said that things *instantiate*, *exhibit*, or *exemplify* a single property, quality, or attribute. Nonetheless, the form of explanation being recommended is precisely the one Plato proposes. Different things are qualified or characterized in some way by virtue of their all standing in a relationship to the quality or characteristic in question. Attribute agreement gets grounded in a characteristic or quality *common to* or *shared by* the agreeing objects.

Philosophers who endorse the Platonic schema have traditionally been called *metaphysical realists* or simply *realists*;[4] but while many philosophers have found

the realist's explanation of attribute agreement in terms of shared or common entities attractive, the form of explanation proposed by Plato has also had its critics. These critics have been known as *nominalists*. They argue that there are deep conceptual problems with the metaphysical machinery implied by the Platonic schema. Some nominalists take those problems to point to the need for a quite different theoretical explanation for attribute agreement, one making no reference to shared or common entities, whereas others take them to show that no theoretical account at all is required here, that the phenomenon of attribute agreement is a basic or fundamental fact not susceptible of further analysis. The debate between metaphysical realists and nominalists is perhaps the oldest sustained debate in metaphysics. Certainly the issues on which the debate has turned are as important as any in metaphysics. We need to become clear on these issues, and we will begin by attempting to delineate the main contours of the perspective labeled metaphysical realism.

The Ontology of Metaphysical Realism

Metaphysical realists want to insist that an adequate account of attribute agreement presupposes a distinction between two types or categories of objects: what are called *particulars* and what are called *universals*. The category of particulars includes what the nonphilosopher typically thinks of as "things"—familiar concrete objects like human beings, animals, plants, and inanimate material bodies; and the realist tells us that what is peculiar to particulars is that each occupies a single region of space at a given time. Universals, by contrast, are construed as repeatable entities. At any given time, numerically one and the same universal can be wholly and completely exhibited or, as realists typically put it, exemplified by several different spatially discontinuous particulars. Thus, different people can exemplify the same virtue at the same time; different automobiles can simultaneously exemplify the same shape; and different houses can, at a given time, exemplify literally the same color. The virtue, the shape, and the color are all universals.[5] The claim of the metaphysical realist is that familiar particulars agree in attribute in virtue of their jointly exemplifying a single universal. So there are nonrepeatable entities that stand in a special relation to repeatable entities, and this fact is what grounds attribute agreement among the familiar objects of the everyday world.

Realists typically want to claim that there is more than one kind of universal. All the cases of attribute agreement we have mentioned involve what are called *one-place* or *monadic* universals. They are universals that particulars exemplify individually or one by one; but there are also relations, universals that are exemplified by several individuals in relation to each other. Thus, *being a mile apart* is something that is exemplified by a pair of objects: one thing is a mile away from another; and it is a universal: many pairs of objects can be so related at any given time. Likewise, *being next to* is a spatial relation between objects: one object is next to another and, again, it is a universal: many pairs of objects can agree in entering into it. Both these relations are what are called *symmetrical*

relations; given any pair of objects, *a* and *b*, such that *a* bears either relation to *b*, *b*, in turn, bears that same relation to *a*. But not all relations are symmetrical. Many relations are such that pairs of objects enter into them only when taken in a certain order. Thus, *being the father of* is an *asymmetrical* relation: if one thing, *a*, is the father of another thing, *b*, then *b* is not the father of *a*. As logicians put it, it is the ordered pair, <*a*, *b*> (*a* and *b* taken in just *that* order), that exhibits the relation. The three relations we have considered are all two-place or dyadic relations; but obviously there can be three-place, four-place, and, generally, *n*-place relations.

Relations, then, are *polyadic* or *many-place* universals. But colors, virtues, and shapes are all monadic. Each is exhibited by objects taken individually. Now, many realists lump all monadic universals together under the title 'property'; but some realists (typically those influenced by the Aristotelian tradition) insist on a further distinction here. We are asked to distinguish between properties and kinds. Kinds are things like the various biological species and genera.[6] Whereas objects exemplify properties by *possessing* them, things exemplify kinds by *belonging to* them. Philosophers who draw this distinction frequently tell us that while kinds constitute the particulars that exemplify them as *what* they are, properties merely modify or characterize particulars antecedently so marked out; and they often claim that kinds are *individuative* universals. What is meant is that kinds constitute their members as individuals distinct from other individuals of the same kind as well as from individuals of other kinds. Thus, everything that belongs to the kind *human being* is marked out as a discrete individual, as one human being countably distinct and separate both from other human beings and from things of other kinds.

So attribute agreement can involve a variety of different types of universal. Several particulars can agree in belonging to a single kind; they can agree in possessing a single property; and several pairs, triples, or generally, *n*-tuples of particulars can agree in entering into a single relation. And realists want to claim that attribute agreement of any of these forms is subject to degrees. A dog and a cat agree in kind: both are mammals; but their agreement in kind is not as close as that tying two dogs. According to the realist, what gives rise to the difference in degree of agreement is the fact that the universals particulars exemplify exhibit varying degrees of generality. The more specific or determinate a shared universal, the closer is the resulting attribute agreement. Universals, then, come in hierarchies of generality. Presumably, every such hierarchy terminates in fully determinate universals, universals such that they have no less general or more determinate universals under them, and the particulars that jointly exemplify any such fully determinate universal will agree exactly in color, shape, kind, spatial relation, or whatever.

So particulars exemplify different sorts of universals of varying degrees of generality; but realists want to claim that the universals that serve to explain the attribute agreement among particulars can themselves agree in exemplifying further universals. Thus, the properties of red, yellow, and blue have various properties of tone and hue; they all belong to the kind *color*; and they enter into

relations like *being lighter than* and *being darker than*. And, of course, the universals exemplified by colors can be more or less determinate, thereby explaining why, for example, red is closer to orange than blue is.

Thus, the original insight that familiar particulars agree in attribute by virtue of jointly exemplifying a universal gives rise to a picture of considerable complexity. Particulars and *n*-tuples of particulars exemplify universals of different types: properties, kinds, and relations. Those universals, in turn, possess further properties, belong to further kinds, and enter into further relations; the same is true of these further properties, kinds, and relations; and so on, seemingly without end. And the seemingly endless series of universals that have come onto the scene enter into complicated hierarchies of generality, thereby inducing complex patterns of attribute agreement of varying degrees of generality. What began, then, as an apparently innocent extension of common sense has blossomed into a full-scale metaphysical theory, an ontology, that is a long distance from common sense.

Some might balk at the complexity of the theory, but realists want to insist that the complexity of the structure has a theoretical payoff. The structure represents a fruitful theory, one with the resources for explaining a wide range of phenomena. Although the phenomena realists claim their account explains are diverse and numerous, we will consider just two. Both bear on semantical issues, and both have played significant roles in the history of metaphysical realism. The first concerns *subject-predicate discourse*; the second bears on *abstract reference*. According to the realist, both phenomena give rise to pressing philosophical questions, and the realist insists that the theoretical machinery associated with metaphysical realism provides straightforward and satisfying answers to those questions.

Realism and Predication

The subject-predicate sentence is about as basic a form of discourse as there is. The following sentences are examples of this form of discourse:

(1) Socrates is courageous,
(2) Plato is a human being,
(3) Socrates is the teacher of Plato.

Using a sentence like one of these, we pick out or refer to a particular and go on to say something about it—to characterize or describe it in some way, to indicate what kind of thing it is, or to relate it to something else. Using (1), for example, we refer to Socrates and we say of him that he is courageous. This characterization of (1) suggests that it is only the subject term ‘Socrates’ that plays a referential role or picks out an object in (1), but metaphysical realists want to insist that such an account is incomplete. Any satisfactory analysis of (1), they claim, will show the predicate term ‘courageous’ to have referential force as well.[7]

Suppose that (1) is true. Pretty clearly, its truth depends on two things: first, what (1) says, and second, the way the world is. Both of these things are matters of structure; what (1) says is a matter of the terms that enter into its composition and the order in which they are placed. The relevant way the world is, on the other hand, is a matter of nonlinguistic structure; it is a matter of how things in a certain sector of the world are and how they are related to each other. So the truth of (1) involves a linguistic structure and a nonlinguistic structure, and the realist insists that it is because we have a correspondence between the two structures that (1) is true. It is because the linguistic structure of (1) corresponds to or mirrors the nonlinguistic structure of a certain sector of the world that (1) is true.[8] Pretty clearly, if we are to have the requisite correspondence, there must be a thing correlated with the proper name 'Socrates,' but the realist argues that (1) can be true only if 'courageous' is likewise correlated with some nonlinguistic object. As it occurs in (1), 'courageous' is not playing a purely formal role, the kind of role associated with terms (like the conjunctions 'or' and 'if' or the definite and indefinite articles) that do not enter into any relation with objects out in the world. Its role in (1) is to make contact with the world by referring to or picking out an object. So if (1) is to be true, both its subject term and its predicate term must have a referent, and the referents of these two terms must be related in a way that ensures that what (1) says is true. But, then, as it occurs in (1), 'courageous' picks out an entity such that, in virtue of being related to it, the referent of 'Socrates' is as (1) says he is—courageous.

Metaphysical realists, however, are quick to point out that 'courageous' is a general term; it is a term that can be applied to individuals other than Socrates and so can figure as predicate in true subject-predicate sentences other than (1). Suppose, for example, that not just (1), but also:

(4) Plato is courageous

is true. The argument presented for the case of (1) applies here as well. 'Courageous' is playing a referential role in (4) no less than in (1). But what is the relation between the referents of these two occurrences of 'courageous'? Pretty clearly, what we say about Plato when we predicate 'courageous' of him in (4) is precisely what we say about Socrates when we predicate 'courageous' of him in (1). And, according to the realist, that implies that whatever referential force 'courageous' has in (1) and (4), it is the same referential force in the two cases. The realist concludes that 'courageous' picks out a single entity in (1) and (4), a single entity such that in virtue of being related to it, both Socrates and Plato count as courageous.

And, of course, the same line of argument applies in the case of other true subject-predicate sentences where 'courageous' plays the predicate role. In every such sentence, 'courageous' has referential force or picks out an object; and provided the term is being used in a single sense in all these sentences, it has a single referential force in all of them. In every such sentence, it picks out or refers to a single entity, an entity such that in virtue of a relation between it and

the referent of the sentence's subject term, the sentence is true. But what metaphysical machinery is required to tell this story of the truth conditions for sentences like (1), (4), and their ilk? Realists insist that the ontological framework central to their account provides the materials for such a story. Assume that there are repeatable entities or universals and a relation of exemplification tying them to particulars, and our account of the truth conditions for sentences like (1) and (4) goes smoothly. It is because 'courageous' has as its referent a certain universal—the virtue of courage—and because each of Plato and Socrates exemplifies that universal that (1) and (4) are true.

Realists want, of course, to extend the story we have told about (1) and (4) to provide a general account of subject-predicate discourse. Predicates refer to universals, and what makes a subject-predicate sentence true is just that the referent of its subject term exemplifies the universal that is the referent of its predicate term. And the realist will typically claim that there are different kinds of universals that can be the referents of predicate terms. The predicates of subject-predicate sentences like (1), where we characterize an object or say how it is, take properties as their referents. Other subject-predicate sentences are like:

(2) Plato is a human being,

enabling us to identify what a thing is or to say what kind of thing it is. Their predicates take kinds as their referents. Finally, there are subject-predicate sentences like:

(3) Socrates is the teacher of Plato,

which enable us to say how different objects are related to each other; their predicates refer to relations.

If this analysis is to be complete, however, we need an account of the kind of referential relation that ties predicates to properties, kinds, and relations. Our paradigm of the referential relation is that between a name and its bearer, the sort of relation that ties 'Socrates' to the man Socrates; and some realists have wanted to claim that it is precisely this relation that predicates bear to universals.[9] Their typical example is a sentence like:

(5) This is red,

where we specify the color of some particular. We are told that (5) incorporates two names tied together by the copula 'is': 'this' names a certain particular, 'red' names a certain universal, and the copula expresses the relation of exemplification that ties the particular named by 'this' to the universal named by 'red.' On this account, the insight that subject-predicate truth involves a correspondence between a linguistic structure and a nonlinguistic structure gets a very strong expression; for on this view, we have a one-to-one correspondence between the linguistic expressions out of which (5) is composed and the nonlinguistic items

that are supposed to make (5) true. But while the claim that universals are named by predicates might seem attractive for a sentence like (5), when we turn to other subject-predicate sentences, we find that the analysis does not generalize very well. Consider, again,

(1) Socrates is courageous.

It is not plausible to suppose that its predicate is a name. Where a term names an entity, it can play the role of subject term in a subject-predicate sentence; and in that role, it refers to the item that it names. 'Courageous' does not, however, pass that test; it is not grammatically suited to occupy the subject position. If any term names the universal the realist wants to correlate with the predicate 'courageous,' it is the term 'courage'; and just as 'courageous' cannot play the subject role, 'courage' cannot function as a predicate. Nor is the case of 'courageous' idiosyncratic. Consider:

(6) This coin is circular,
(7) Plato is wise,

and:

(8) Alcibiades is exhausted.

In none of these cases is it plausible to claim that the predicate functions as a name of the universal it is supposed to refer to. In each case, there is another term ('circularity,' 'wisdom,' 'exhaustion') that is more plausibly construed as the name of the relevant universal. The fact that we cannot take the predicates of (1), (6), (7), or (8) to be names of universals suggests that 'red' is not playing that role in (5) either; and the fact is that it is not. 'Red,' along with other color words, is ambiguous; it can function as a noun (as in 'Red is a color'), and in that use, it is plausibly construed as a name of the relevant color; but it can also function as an adjective (as in 'red house' and 'red complexion'), and in that use, it does not name anything. In (5), the term has its adjectival use and so is no more a name there than 'courageous' is in (1).

We have been focusing on the grammatical obstacles to construing predicates as names; but those obstacles have semantical roots. A name is a singular term; it picks out its bearer and nothing else. Predicates, by contrast, are general terms and, as such, they enter into a referential relation with each of the objects of which they can be predicated. In the semanticist's jargon, they are *true of* or *satisfied by* those objects. But if their entering into that relation precludes their serving as names of universals, is there any other kind of referential relation that they might, nonetheless, bear to universals? Many realists have insisted that there is. They have claimed that in addition to being true of or satisfied by the objects of which they can be predicated, predicate terms *express* or *connote* universals.[10] Thus, 'courageous' is referentially linked to all and only courageous individuals

by the relation of satisfaction; but realists have claimed that it also expresses or connotes the universal all those individuals have in common, the virtue of courage. Likewise, 'circular' is satisfied by all and only the individuals that are circular, but realists tell us that it bears the further semantical relation of expression or connotation to the universal those individuals all share, the shape of circularity.

Toward clarifying the claim that predicates express universals, realists argue that to apply a predicate term to an object is to do more than merely identify the object as a member of a set of objects; it is to identify as well the universal in virtue of which objects belong to the set. Thus, when we say that an object is triangular, we are not merely saying that it belongs to a set of objects. We are also pointing to the property shared by all the members of the set and saying that the object in question exhibits that property. According to the realist, the fact that the use of a predicate term involves more than the mere identification of the items it is true of is shown by the fact that subject-predicate sentences like our (1)–(8) admit of paraphrases in which the reference to a universal is made explicit; (1), for example, can be paraphrased as:

(1') Socrates exemplifies courage,

and (6) can be paraphrased as:

(6') This coin exemplifies circularity.

In both cases, the original subject-predicate sentence gives way to a sentence in which there occurs a singular term that bears what appears to be the naming relation to the universal that the realist takes the predicate of the sentence to refer to or pick out. Now, realists want to claim that the possibility of such paraphrases is general, so that any subject-predicate sentence of the form '*a* is *F*' can be paraphrased by a sentence of the form '*a* exemplifies *F-ness*.' But if paraphrases of this sort are always possible, then to predicate a general term, '*F*,' of an object is just to say that the object exemplifies the universal, *F-ness*. And this implies that even if predicates do not name universals, their use in the context of a subject-predicate sentence has the force of introducing universals into discourse, of mentioning or referring to universals. There is, then, a referential relation here, one weaker or less direct than, but parasitic on the naming relation. That relation is what the realist calls expression or connotation. And the realist will, once again, typically claim that predicates can express or connote different kinds of universals. The predicate of a sentence like (1) expresses or connotes a property, and to assertively utter (1) is to say that a given object exemplifies that property by possessing or having it. The predicate of (2), by contrast, expresses a kind; and to assertively utter (2) is to say that some object exemplifies that kind by belonging to it. Finally, the predicate of (3) expresses a dyadic relation; and to use (3) to make a claim is to say that a particular pair of objects exemplify that dyadic relation by entering into it.

So predicates express or connote properties, kinds, and relations; and where we have a true subject-predicate sentence, the universal expressed by the predicate is exemplifi ed by the referent of the sentence's subject term. The realist claims that this account does what we want it to do; it explains how subject-predicate sentences can manage to correspond to the world, and it does so in a natural or intuitively satisfying way. What makes the account so natural, according to the realist, is its connections with the realist's interpretation of attribute agreement. General terms play the predicate role; and, on any theory, general terms mark cases of attribute agreement: all the items of which a given general term is true agree in attribute or are similar in some way. But the items that agree in attribute, according to the realist, all exemplify some one universal; and, on the realist's account, the general term that marks a given case of attribute agreement expresses or connotes precisely the same universal that supports or grounds that case of attribute agreement. So we have an account of predication that goes hand in hand with our account of attribute agreement, and the two accounts mesh in just the way they must if we are to provide a satisfactory account of subject-predicate truth. The universal that is the referent of a predicate term is precisely the universal that must be exemplifi ed by the referent of a subject term if that referent is to be something that instances the case of attribute agreement marked by that predicate term.

Realism and Abstract Reference

Realists want to claim that an ontology of universals provides us with the resources for explaining more than predication. They think their metaphysical theory enables us to give an intuitively satisfying account of the phenomenon of abstract reference.[11] This phenomenon makes its most obvious appearance in the use of what are called *abstract singular terms*. Examples of abstract singular terms are expressions like 'triangularity,' 'wisdom,' 'mankind,' and 'courage.' They are all singular terms: they can play the subject role; and they tend to pair off with expressions that can play the predicate role—general terms. Thus, we have 'triangularity'/'triangular,' 'wisdom'/'wise,' 'mankind'/'man,' 'courage'/'courageous,' and 'red' (in its noun use)/'red' (in its adjectival use). Now, intuitively, the terms making up each of these pairs seem to be related in a quite distinctive way: the abstract singular term appears to be a device for picking out a certain property or kind and the general term appears to be an expression true of or satisfi ed by all and only the objects that exemplify that property or kind. The realist insists that this intuitive account is correct and claims that unless we take abstract singular terms to be devices for referring to universals, we cannot provide a satisfactory account of the sentences in which they appear. The following are examples of such sentences:

(9) Courage is a moral virtue,
(10) Triangularity is a shape,
(11) Hilary prefers red to blue,

(12) Mankind is a kind,
(13) Wisdom is the goal of the philosophic life,

and so are the sentences we mentioned in our account of the referential force of predicates:

(1') Socrates exemplifies courage

and:

(6') This coin exemplifies circularity.

Realists point out that sentences like these are often true, and argue that only the metaphysical realist has the resources for explaining how they can manage to be true. The realist insists that if we are to provide an account of what these sentences say, we must hold that, as they occur in these sentences, abstract singular terms are functioning in precisely the way the intuitive account tells us they function: they are playing referential roles of the most straightforward sort; they are functioning as names of universals. But if they are playing that sort of role, the sentences in which they occur can be true only if the universals they name actually exist. So only the philosopher who endorses an ontology of universals can account for the truth of sentences in which abstract singular terms appear.

Consider (9). In (9), we pick out a certain property, the property exemplified by all and only courageous individuals, and we go on to say what kind of thing it is; we say that it is a moral virtue. So (9) is a claim about a certain property, the property the intuitive account tells us is named by the abstract singular term 'courage'; and that claim can be true only if that property exists; for surely the claim that courage is a thing of a certain kind could not be true if there were no such thing as courage. Likewise, in (10) we pick out the property exemplified by all and only triangular objects, and we say of that property that it is a shape. Thus, (10) is a claim about a certain property, the property the intuitive account tells us is the referent of the abstract singular term 'triangularity'; and the truth of (10) presupposes that the referent of that abstract term exists. It could hardly be true, after all, that triangularity belongs to a certain kind if triangularity did not exist. And analogous points could be made regarding (11)–(13), (1'), and (6'). In each case, we have an abstract singular term, and the sentence in question manages to say what it does only because the relevant abstract term is functioning in the way the intuitive account tells us it functions, only because it is playing the referential role of naming a universal. Accordingly, each of these sentences can be true only if the universal named by the constituent abstract term exists. And, of course, there are many other such sentences; and like our sample sentences, their truth presupposes the existence of the universals the intuitive account takes to be the referents of their constituent abstract singular terms. But obviously many such sentences are true, and only the metaphysical

realist, only the philosopher who holds that universals exist, can tell us how this is possible.

So the fact that sentences incorporating abstract singular terms can be true is something realists claim only they can explain. They insist, however, that what we have called abstract reference is not restricted to sentences like those we have been considering. There are sentences incorporating no abstract singular terms which, nonetheless, appear to involve a reference to things like properties, kinds, and relations.[12] The following are examples of the sorts of sentences the realist has in mind:

(14) That tomato and that fire engine have the same color,
(15) Some species are cross fertile,
(16) There are undiscovered relations tying physical particles to each other,
(17) He has the same character traits as his cousin,

and:

(18) That shape has been exemplified many times.

Although none of these sentences includes a singular term that names a universal, the realist tells us that they are all claims about universals, claims about the colors, character traits, and shapes things share, the biological kinds to which they belong, and the relations into which they enter; and insists that none of these sentences can be true unless the universals in question actually exist. Thus, (14)–(17) are straightforward assertions of the existence of universals meeting certain conditions; none of them can be true unless there exist universals meeting those conditions; and while (18) is not an explicit existence claim, its truth presupposes the existence of at least one multiply exemplifiable entity, a certain shape. So, again, we have the claim that there are sentences whose truth implies the existence of the sorts of things the realist calls universals; the realist points out that many sentences like (14)—(18) are true and concludes that only the philosopher who endorses an ontology of universals can explain this fact.

The sentences that exhibit the phenomenon of abstract reference, then, include both sentences with and sentences without abstract singular terms; but in both cases, the realist's contention is the same: that to account for their truth, we must endorse the ontology of metaphysical realism. A couple of comments about this line of argument are in order. First, it is independent of the realist's account of predication. The realist's claims about sentences like (9)–(18) presuppose no particular theory of predication. Even if we suppose that the only semantical property associated with predicates is that of being true of or satisfied by the items of which they are predicated, the fact remains that, intuitively, the use of sentences like (9)–(18) has the force of making claims about entities other than familiar concrete particulars. Indeed, it is plausible to think that this argument is actually presupposed by the realist's analysis of predication. As we have seen, when realists attempt to explicate and justify the claim that predicates take

universals as their referents, they appeal to the fact that ordinary subject-predicate sentences of the form '*a* is *F*' can be paraphrased by way of sentences of the form '*a* exemplifies *F-ness*.' But it is only because sentences of the latter form incorporate abstract singular terms, and because we take the truth of sentences incorporating such terms to commit us to the existence of universals, that we take the appeal to these paraphrases as evidence for the realist's theory of predication.

Second, the realist's claims about sentences involving abstract reference cannot be properly evaluated in isolation from alternative accounts of the role of abstract referring devices; for the warrant for those claims must be the failure of alternative analyses of sentences like (9)–(18). If a satisfactory nominalist account of the content and truth conditions of such sentences is forthcoming, then the realist's claim that the truth of these sentences commits us to an ontology of universals is gratuitous. The same is true of the earlier argument from subject-predicate truth. An adequate account of how subject-predicate sentences can correspond with nonlinguistic fact that does not construe predicates as referentially tied to universals would call into question the realist's claim that we need universals to account for subject-predicate truth. So both arguments are best understood as challenges to the nominalist to come up with systematic and intuitively attractive theories of predication and abstract reference, theories that give us an account of the metaphysical grounds of subject-predicate truth and the use of abstract referring devices without making reference to common or shared entities. As we shall see in the next chapter, nominalists have recognized the burden placed on them by the realist's argument in these two arenas and have expended considerable effort showing that such an account is possible. And given the way that the realist's account of abstract reference enters into the realist's account of predication, it is not surprising that nominalists have been most concerned to provide an account of the role of abstract singular terms. As we shall see, the realist's claim that our intuitive understanding of sentences like (9)–(18) presupposes the existence of universals is just an opening salvo. Realists realize that they must respond to alternative accounts of such sentences; but they are prepared to do so and are confident that their own analysis will be vindicated by the examination of nominalist accounts of abstract reference.

Restrictions on Realism—Exemplification

Our discussion suggests that metaphysical realists constitute a unified group defending a single doctrine, but the fact is that realists have disagreed among themselves on a range of issues. The most important bears on the generality of the theory. Our treatment of realism suggests that realists want to apply the Platonic schema across the board, so that for every case of what we would prephilosophically call agreement in attribute, the realist will posit a separate universal. Likewise, we have implied that every general term that can function predicatively in a true subject-predicate sentence expresses or connotes a distinct universal and that every semantically distinct abstract term names a unique

universal. But many realists have been unwilling to endorse such an unrestricted version of the theory. They have insisted that we place restrictions on the theory, so that universals correspond to only some of the ways things can be said to be, to only a limited pool of general terms, and to only some of the abstract terms in our language. Furthermore, the restrictions imposed on the theory have varied, so that by examining the different ways the theory has been restricted and the rationale for each restriction, we can bring to light the different forms metaphysical realism has taken.

We should begin by noting that no version of metaphysical realism can consistently endorse the completely unrestricted application of the Platonic schema or hold that every nonequivalent predicate term or every nonequivalent abstract term is associated with a separate and distinct universal. An entirely unrestricted version of the theory lands one in a notorious paradox. We can bring out the paradoxical nature of an unrestricted realism by focusing on the realist's analysis of predication. Suppose we endorse that analysis in its full generality and hold that a universal corresponds to every general term that can occupy the predicate position in a true subject-predicate sentence. Consider now the general term 'does not exemplify itself.' This term is, to be sure, syntactically complex; but we could, if we wished, introduce a single expression to replace the complex predicate, so the syntactic complexity is really an irrelevant detail. We have here a perfectly respectable general term, one true of or satisfied by all and only the things that do not exemplify themselves; and it is a general term that can function predicatively in true sentences. The expression is true, for example, of Bill Clinton, the number two, and the Taj Mahal. Since none of these things is self-exemplifying, each satisfies the predicate 'does not exemplify itself'; and the relevant subject-predicate sentences will all be true. There are, on the other hand, things, certain universals, to which the predicate does not apply. Presumably, the property of being incorporeal exemplifies itself: it has no body and so is incorporeal. Likewise, if there is such a thing as the property of being self-identical, it is identical with itself and so exemplifies itself. Accordingly, neither of these things satisfies the predicate 'does not exemplify itself.'

Now, since there are true subject-predicate sentences where this term functions predicatively, a totally unrestricted version of the realist's theory of predication will tell us that there is a property expressed or connoted by this predicate expression. For convenience, let us call it the property of being non-self-exemplifying. The assumption that there is such a property leads immediately to paradox; for the property must either exemplify itself or fail to do so. Suppose it does exemplify itself; then, since it is the property a thing exemplifies just in case it does not exemplify itself, it turns out that it does not exemplify itself. So if it does exemplify itself, it does not exemplify itself. Suppose, on the other hand, that it does not exemplify itself; then, it turns out that it does exemplify itself; for it is the property of being non-self-exemplifying. So if it does not exemplify itself, it does exemplify itself. But, then, it exemplifies itself just in case it does not, a deplorable result.[13] To avoid the paradox, we have no option but to deny that there is a universal associated with the general term 'does not

exemplify itself.' The realist's account of predication cannot hold for *all* general terms that function predicatively in true subject-predicate sentences.

It is frequently claimed that still further restrictions have to be imposed on the realist's theory. The claim is that, without additional restrictions, the realist's theory lands us in a vicious infinite regress. The contention is very old; it can be found in Plato's *Parmenides* and has been repeated again and again since the time of Plato.[14] The difficulty that is supposed to confront the realist bears on the core notion of exemplification. One way of stating the difficulty takes its origin in the realist's use of the Platonic schema for explaining attribute agreement. According to the schema, where a number of objects agree in all being *F*, their agreement is grounded in their multiple exemplification of the universal *F-ness*. The difficulty is that, for any given application of the schema, that application explains one case of attribute agreement, the original objects all being *F*, only to confront a new case, their all exemplifying *F-ness*. But, then, we have to appeal to a further universal, the exemplification of *F-ness*, and we have to say that the second case of attribute agreement holds among our original objects in virtue of their jointly exemplifying this second universal; but, then, we explain our second case of attribute agreement only to confront a third case, our original objects all agreeing in exemplifying the exemplification of *F-ness*. So we need to appeal to a third universal which will, in turn, generate still another case of attribute agreement with the resulting need for still another universal, and we are off on an endless regress through cases of attribute agreement and supporting universals. Conclusion? If we endorse the Platonic schema, the explanation that schema is supposed to provide can never be completed.

It should be obvious that the same difficulty appears to plague the realist's attempt to explain subject-predicate truth. The realist wants to claim that an arbitrary subject-predicate sentence,

(19) *a* is *F*,

is true only if the referent of '*a*' exemplifies the universal (*F-ness*) expressed by '*F*.' But, then, our original sentence, (19), is true only if a new subject-predicate sentence,

(20) *a* exemplifies *F-ness*,

is true, and it looks as though we have not completed our explanation of the truth of (19) until we have exhibited the ground of the truth of this new sentence. However, (20) incorporates a new predicate ('exemplifies *F-ness*') and it expresses a new universal (the exemplification of *F-ness*). The realist's theory tells us that (20) can be true only if the referent of '*a*' exemplifies the new universal. But that condition is satisfied only if:

(21) *a* exemplifies the exemplification of *F-ness*

is true, so it seems that our account of the truth of (19) requires an account of the truth of this third sentence. Once again, we appear to be off on an infinite regress, and once again, we have the apparent conclusion that the realist's theory cannot do what it is supposed to do.

The two regresses we have outlined might seem to have a simple moral: we must reject the metaphysical realist's account of attribute agreement and predication; and the regresses have frequently been exploited by philosophers of a nominalist bent to point up precisely this moral; but realists have often argued that the regresses have a quite different moral. They concede that the regresses must be avoided, but they think that there is an easy way to do this. We need merely to set restrictions on the use of the Platonic schema and its associated theory of predication. Confronted with the first regress, we can deny that every distinct form of attribute agreement involves a separate and distinct universal. In particular, we can deny that where the agreement consists in a number of objects exemplifying a universal, there is a further universal supporting the agreement. Likewise, in confronting the second regress, we can deny that every semantically distinct general term expresses a distinct universal. While conceding that there is a universal corresponding to the predicate of any sentence whose form is that of (19), we can deny that there are further universals corresponding to the predicates of sentences of the form of (20) or any of its successors.

So the claim is that if we restrict the applicability of the Platonic schema and the realist's theory of predication, we can avoid these regresses. One might, however, challenge the idea that any restriction is called for here. If the regresses just delineated are real, it is difficult to see why the realist should be bothered by them. Consider the contention that the use of the Platonic schema is viciously regressive. The realist claims to have a schema for providing a complete account of any given case of attribute agreement; but the alleged regress does nothing to call that claim into question. If there is, as the argument claims, an infinity of cases of attribute agreement lying behind any given case, that fact does not jeopardize the realist's use of the Platonic schema to provide a full and complete explanation of the initial case of attribute agreement. When realists tell us that our sample objects are all *F* because they all exemplify *F-ness*, they have given us a complete explanation of the original case of attribute agreement. If, as the argument claims, the explanation introduces a new case of attribute agreement, realists are free to apply the Platonic schema to the second case; but they are under no obligation to do so. In particular, the success of the original application of the schema to explain the first case of attribute agreement does not hinge on their explaining the second; and the same holds for each of the cases of attribute agreement allegedly following upon this one. So if the regress is real, it is not vicious; and, accordingly, no restriction on the use of the Platonic schema is called for.

A similar point can be made in reply to the claim that realists must set restrictions on the application of their theory of subject-predicate truth. Even if the regress allegedly requiring the relevant restriction is real, it is not vicious. If, as the argument claims, the realist explanation of the truth of (19) brings a new

true subject-predicate sentence, (20), on the scene, the realist's success in explaining the truth of (19) does not presuppose an explanation of the truth of (20). If the aim had been to eliminate or analyze away the subject-predicate form of discourse, then the emergence of (20) would be genuinely problematic. But the realist is hardly committed to supposing that it is possible to eliminate that form of discourse. Indeed, if there is a regress here, it is one that infects every attempt, realist *or* nominalist, at delineating the ontological grounds of subject-predicate truth.[15] Consider a nominalist theory of subject-predicate truth. For each subject-predicate sentence of the form '*a* is *F*,' it will identify some condition, *C*, and will tell us that the original sentence is true only if *C* is fulfilled; but then there will be a new subject-predicate sentence ('*a* is such that *C* is fulfilled'), and our original sentence can be true only if the second sentence is true. Accordingly, that theory will be every bit as regressive as the realist's. And in neither case is the alleged regress vicious. So even if there is a regress here, no restriction on the range of applicability of the realist's theory of predication is required.

But if they are not vicious, the two regresses seem to have the upshot that behind any case of attribute agreement or any true subject-predicate claim, there lies an infinite series of distinct universals. Some realists might find that fact worrisome; and in the interests of keeping the number of universals to a minimum, they might feel that the relevant restrictions need to be imposed on the realist's theories of attribute agreement and predication. But if realists are concerned about a bloated ontology, it is open to them to deny that the relevant regresses are even real. They can challenge the idea that when we say that objects agreeing in being *F* all jointly exemplify the universal, *F-ness*, we have thereby identified a second case of attribute agreement. We can say that in applying the Platonic schema to identify the ontological ground of a given case of attribute agreement, we are providing a fully articulated and metaphysically more perspicuous characterization of that case rather than introducing a new case. And in a similar vein, they can claim that the predicate of (20), 'exemplifies *F-ness*' is only syntactically or grammatically distinct from that occurring in (19), '*F*.' Semantically, they can claim, the two predicates are equivalent and so do not rest on distinct ontological foundations.

Neither of the first two attempts at arguing that the realist theory lands us in a regress that requires a restriction on that theory carries much force, then. There is, however, a third way of arguing this claim. According to most realists, this third argument poses genuine problems for their account, problems that can be resolved only by restricting the range of the theory. According to the realist, for a particular, *a*, to be *F*, it is required that both the particular, *a*, and the universal, *F-ness*, exist. But more is required; it is required, in addition, that *a exemplify F-ness*. As we have formulated the realist's theory, however, *a*'s exemplifying *F-ness* is a relational fact. It is a matter of *a* and *F-ness* entering into the relation of exemplification. But the realist insists that relations are themselves universals and that a pair of objects can bear a relation to each other only if they exemplify it by entering into it. The consequence, then, is that if we are to have the result that *a* is *F*, we need a new, higher-level form of

exemplifi cation (call it exemplification$_2$) whose function it is to ensure that *a* and *F-ness* enter into the exemplifi cation relation. Unfortunately, exemplification$_2$ is itself a further relation, so that we need a still higher-level form of exemplification (exemplification$_3$) whose role it is to ensure that *a*, *F-ness*, and exemplifi cation are related by exemplification$_2$; and obviously there will be no end to the ascending levels of exemplifi cation that are required here. So it appears, once again, that the only way we will ever secure the desired result that *a* is *F* is by denying that exemplifi cation is a notion to which the realist's theory applies.

The argument just set out is a version of a famous argument developed by F.H. Bradley.[16] Bradley's argument sought to show that there can be no such things as relations, whereas the argument we have been elaborating has the more modest aim of showing that the realist's story of what is involved in a thing's having a property, belonging to a kind, or entering into a relation cannot apply to itself. Now, some realists have held that while real, the regress just cited is not vicious.[17] They have taken the regress to be no more threatening than the first two regresses we have outlined. These realists have, however, been in the minority. Most realists have seen the regress as vicious. It is not altogether clear just why; for on the surface, the regress appears to have the same formal structure as the earlier two regresses. Of course, realists have sometimes mistakenly thought that those two regresses are problematic, so it is not surprising that they should fi nd the third regress worrisome. What is puzzling is that realists who show no concern over the original pair of regresses should be bothered by this regress. Perhaps they have felt that this regress, unlike the earlier two, makes it impossible to explain the thing we initially set out to explain—*a*' s being *F*. Perhaps they have felt that unless realists can point to some connecting mechanism whose connecting role is secured without dependence on some further, higher-level connecting mechanism, they have not succeeded in explaining why the particular, *a*, is *F*. It is not, however, obvious that this line of thinking is correct; for it is reasonable to think that once the realists have told us that *a* is *F* because *a* and *F-ness* enter into the relation of exemplifi cation, they have completed their explanation of the fact that *a* is *F*. There is, of course, something new the realist might want to go on and explain—the new fact that *a* and *F-ness* enter into the relation of exemplifi cation; however, the failure to explain this new fact would seem to do nothing to jeopardize their explanation of the original fact that *a* is *F*.

But whether we fi nd the reason compelling, the fact remains that our third regress looms large in the history of metaphysical realism. Realists have typically believed that they have no option but to stop the regress before it gets started.[18] Toward stopping the regress, they have insisted that the realist account does not apply to the notion of exemplifi cation itself. Obviously, some justification for this restriction is called for; and the justifi cation given is that exemplifi cation is not a relation. Realists claim that while relations can bind objects together only by the mediating link of exemplification, exemplification links objects into relational facts without the mediation of any further links. It is, we are told, an unmediated linker; and this fact is taken to be a primitive categorial feature of the concept of exemplifi cation. So, whereas we have so far spoken

of exemplification as a relation tying particulars to universals and universals to each other, we more accurately reflect realist thinking about the notion if we follow realists and speak of exemplification as a 'tie' or a 'nexus,' where the use of these terms has the force of bringing out the *nonrelational* nature of the linkage this notion provides.

So realists typically deny that their own account applies to the case of exemplification. Now, whether we find the restriction well motivated, we must concede that there is a bonus to this restriction; for if the realist account does not apply to the notion of exemplification, then our earlier claim that the Platonic schema cannot apply to the predicate 'does not exemplify itself' looks less like a desperate and *ad hoc* attempt at avoiding paradox. If there are reasons for supposing that the schema does not apply to the concept of exemplification, then it is only natural to suppose that it does not apply to concepts built out of that notion; and since in claiming that exemplification is not a relation, realists have some justification for denying that the schema applies to it, they would seem to have plausible grounds, independent of the threat of paradox, for excluding from the range of the schema the notion of being non-self-exemplifying.

Further Restrictions—Defined and Undefined Predicates

As I have suggested, most realists would endorse the restrictions we have so far placed on metaphysical realism; but some realists want to place further restrictions. Consider, for example, the predicate 'bachelor.' As we have formulated it, the realist's account tells us that there is a universal correlated with this predicate. Which universal is that? The property, presumably, of being a bachelor. But that universal is a property a thing has just in case it has the property of being male, the property of being a human being, and the property of being unmarried. So how many properties do we have here? We need the properties of being male, of being a human being, and of being unmarried to accommodate the predicates 'male,' 'human being,' and 'unmarried'; but do we need the further property of being a bachelor? We can give a perfectly satisfactory account of the predicate 'bachelor' by reference to the other three, apparently more basic properties, so is it not redundant to add a fourth property to our inventory? Isn't that additional property just needless clutter? But the doubt about the need to postulate an extra property for the predicate 'bachelor' can be extended quite naturally to the case of 'unmarried.' If we concede the need for a property to correspond to the predicate 'married,' do we need to posit an additional negative property in the case of 'unmarried'? Can we not say instead that 'unmarried' is true of a thing just in case it lacks the property corresponding to the predicate 'married'? Again, is it not redundant to add the negative property to our ontology? And, of course, if we concede, as it seems we must, that the predicate 'married' can be defined in terms of other more basic predicates, then the doubts we have raised about 'bachelor' and 'unmarried' can be extended even further.

These doubts have led some realists to set very severe restrictions on the analysis of predication so far delineated. They have insisted on a distinction

between what they call *undefined* and *defined* predicates.[19] The idea is that there are certain predicates that are not defined in terms of other predicates; these primitive predicates get their meaning by being directly correlated with universals. All other predicates are defined in terms of these primitive predicates. On this view, then, there is not a separate and distinct universal correlated with every semantically nonequivalent predicate; it is only in the case of the primitive or undefined predicates that this is so. The semantical properties of defined predicates can be explained by reference to the universals correlated with the primitive predicates in terms of which they are defined.

Although this way of restricting the realist analysis of predication may initially seem attractive, it has its problems. The central difficulty is that predicates do not come neatly divided into those that are basic or primitive and those that are defined. The philosopher must make the division, and it is arguable that any such division will be somewhat arbitrary. What one formalization of a language takes to be a basic or undefined predicate, another can, with equal adequacy, construe as a defined expression. This fact raises doubts about the distinctively metaphysical force of any attempt at dividing predicates into those that are primitive and those that are defined. If the distinction between undefined and defined predicates is to be a guide to what universals there are, it can hardly rest on the arbitrary decisions of a formalizer.

To avoid the charge of arbitrariness, then, the realist who finds this distinction useful will need to provide some philosophical justification for identifying certain predicates as basic. One important kind of justification that has been provided here is epistemological. Realists who have endorsed a strongly empiricist program in the theory of knowledge have insisted that the basic or primitive predicates are those that express features or characteristics that, from the empiricist's perspective, are epistemologically basic. Accordingly, it is predicates expressing colors, sounds, smells, simple shapes, and the like that are construed as primitive. Corresponding to each such nonequivalent predicate, there is said to be a distinct and separate universal; and it is claimed that all other predicates can be defined by reference to these universals.

Although the view just laid out was popular among realists in the first half of the twentieth century, it does not have many defenders nowadays. Those who endorsed the view found that a large number of predicates resist analysis in terms of merely sensory or perceptual properties. The theoretical predicates of science and moral or ethical predicates are just two cases that proved problematic for realists of the empiricist persuasion. Finding it impossible to analyze these predicates in purely perceptual terms, these realists were forced to deny that the predicates have any genuinely descriptive meaning and to endorse highly implausible accounts of their role in language. Thus, they claimed that the theoretical predicates of science are merely tools or instruments for taking us from one set of statements involving purely perceptual predicates to another such set, and that ethical predicates are nothing more than linguistic vehicles for venting our feelings or emotions about persons, their actions, and their lifestyles.

But it is not simply the empiricist themes at work in this proposal that left philosophers skeptical of the idea that a distinction between defined and undefined predicates is ontologically important. However one goes about the business of dividing predicates into primitive and defined, one is committed to the idea that every nonprimitive predicate can be defined wholly and completely by reference to the predicates taken to be primitive. But the fact is that few of the predicates of our language are like 'bachelor' in being susceptible of definition in terms of less complex predicates. Although it was invoked to make a slightly different point, Wittgenstein's famous discussion of the predicate 'game' brings out the difficulty here:

> Consider for example the proceedings that we call "games." I mean board-games, card-games, Olympic games and so on. What is common to them all? Don't say: "There *must* be something common, or they would not be called 'games'"—but *look* and *see* whether there is anything common to all.—For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat: don't think but look! Look, for example, at board-games, with their multifarious relationships. Now, pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost.—Are they all "amusing"? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball-games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck; and at the difference between skill in chess and skill at tennis. Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! and we can go through the many, many other groups of games in the same way; can see how similarities crop up and disappear.[20]

'Game' is pretty clearly not going to turn out to be a primitive predicate; but if Wittgenstein is right, the attempt to identify a set of more basic predicates whose associated properties will enable one to provide necessary and sufficient conditions for the applicability of the predicate 'game' is bound to be frustrated. 'Game' has a looser, less regimented semantical structure than a term like 'bachelor,' a structure that cannot be captured by any formal definition; and Wittgenstein wants to claim that, on this score, it is typical of most of the predicates of our language.

In the light of Wittgenstein's remarks, it is not surprising that the distinction between primitive and defined predicates does not play a major role in the work of contemporary realists. Some simply deny that the sort of restrictions those invoking the distinction meant to set on realism are appropriate.[21] They are *holists* about universals; that is, they reject any attempt at reducing one set of universals

to another. On the one hand, they are impressed by the fact that where we can provide formal definitions for predicates, any attempt at distinguishing between defined and undefined predicates is bound to be arbitrary. Accordingly, while perhaps conceding that the Platonic schema and its associated theory of predication do not apply to the notion of exemplification, they insist that the universals associated with predicates like 'bachelor' and 'unmarried' are every bit as respectable, every bit as real as those associated with predicates like 'blue' and 'red.' On the other hand, they agree with Wittgenstein that many predicates resist formal definition in terms of other, more basic predicates. However, unlike Wittgenstein, they find this fact no source of embarrassment for the realist. Thus, in response to Wittgenstein's demand to identify a universal common to all the things called games, they point to the property of being a game; and they deny that the impossibility of reducing this property to other more familiar universals is, in any way, problematic.

But other contemporary realists have insisted that even if the attempt to divide predicates into those that are primitive and those that are defined fails as an ontologically revealing way of restricting the realist's account, restrictions need to be placed on the application of the Platonic schema.[22] They agree, then, that our use of only some predicates has genuine ontological force, and they claim that it was not in the attempt to restrict the range of realism that empiricists went wrong. Where they went wrong was in their identification of the ontologically interesting predicates with merely perceptual or observational predicates, and in their claim that the relationship between ontologically revealing predicates and other predicates is one of definition or translation. These realists accuse their more holistic or antireductive colleagues of *apriorism*, the view that we can determine what universals there are by mere armchair reflection on the structure of our language. According to the holists, to determine what universals there are, we need merely look to the stock of predicates at our disposal: to every such nonequivalent predicate, there corresponds a separate and distinct universal. In opposition to this alleged apriorism, it is claimed that the question of what universals there are is an empirical question to be settled by scientific inquiry. It should come as no surprise that those metaphysical realists who rail against linguistic apriorism are typically also *scientific realists*. They hold, that is, that the empirical sciences, in particular physics, represent the criterion of what there is. Accordingly, they claim that the ontologically significant predicates are those essential to the formulation of the correct physical theory. It is, then, the predicates of physics in its finished form that have ontological force.

But if we accept this claim, what are we to make of the predicates that play no role in physical theory? For obvious reasons, the idea that there are translation rules taking us from strictly physical predicates to nonphysical predicates has not been seen as a viable option for the philosopher who seeks to couple metaphysical realism with an austere scientific realism. Instead, we find philosophers who defend the two forms of realism, presenting a number of different and competing views about the relationship between the ontologically significant framework of physical theory and the nonscientific framework of common sense.

I will mention just two. The first, less radical, view will not deny that there are universals correlated with predicates and abstract terms that are not a part of physical theory; but it gives ontological priority to the properties, kinds, and relations of physics. Those universals are construed as ontologically basic or fundamental, and other universals are taken to be dependent on them. The claim is that while the universals that do not enter into physical theory may not be reducible to or analyzable in terms of universals that do, the latter fix or determine the former. What physical kinds a thing belongs to, what physical properties it possesses, and what physical relations it enters into determines uniquely what nonphysical kinds, properties, and relations it exhibits. As it is usually put, nonphysical universals *supervene* on physical universals. On this view, once one has identified all of the physical facts (that is, all the facts recognized by the true physical theory), one has fixed all the facts, nonphysical as well as physical. So while nonphysical properties, kinds, and relations may not be analyzable in terms of the universals of physics, the latter provide the ontological foundation on which the former rest.[23]

A second, more radical account is that of the *eliminativist* who refuses to construe those predicates and abstract terms that cannot be accommodated by reference to the universals invoked in physical theory as having any ontological force.[24] As the eliminativist sees it, our ordinary nonscientific language is the expression of a theory of how the world is; and like any theory, it can be displaced by a theory that provides a more accurate representation of the structure of reality. According to the eliminativist, our best theory of the nature of the world is that delineated by mature physics. To the extent that our nonscientific account of the world is incompatible with mature physical theory, it is false. Those among its predicates and abstract terms that purport to refer to universals that cannot be incorporated in the picture of the world projected by physics are terms without a reference; the universals they purport to express or name simply do not exist. The eliminativist denies that there is anything puzzling about this. It is simply one more case where the theoretical posits of one theory are rejected in favor of those of a more adequate theory.

Are There Any Unexemplified Attributes?

While the differences we have noted have played an important role in the history of metaphysical realism, the single most important issue dividing realists bears on the idea of *unexemplified* universals. In delineating the main contours of realism, our focus has been on actual cases of attribute agreement and on the use of general terms and abstract singular terms in sentences that are actually true. One important tradition, however, would insist that this emphasis on the actual is misguided; it leads us to suppose that all universals are in fact instantiated or exemplified. Realists of this persuasion want to insist, however, that in addition to the exemplified universals, there are many properties, kinds, and relations that are not, never have been, and never will be exemplified.[25] Some of these lack instances only *contingently*; that is, they are such that they might

have been exemplifi ed, but in fact are not. Thus, doubtless there are many complex ways physical objects might have been shaped but never were; the corresponding shapes, these realists claim, are all contingently unexemplified universals. But many of these realists have gone on to claim that, in addition to universals that only contingently go unexemplifi ed, there are attributes that are *necessarily unexemplified*, attributes such that nothing could have ever exemplified them. It is, for example, impossible that anything be both round and square. That is a way nothing could be; these realists insist that there is a corresponding attribute, one that is necessarily unexemplified.

So some realists believe that there are uninstantiated properties, kinds, and relations. Since there is some evidence that Plato believed that this is so, let us call realists of this persuasion Platonists.[26] Opposed to them are realists who insist that every universal has at least one instance at some time or other. It is plausible to think that Aristotle endorsed an ontology involving only exemplified universals; for he tells us that if everything were healthy, there would be no such thing as disease, and if everything were white, the color black would not exist.[27] Let us, then, call realists who reject the Platonist's unexemplified universals Aristotelian realists.

What are the issues separating Aristotelian realists from Platonists?[28] As a start toward answering this question, let us ask why Aristotelians object to uninstantiated universals. Aristotelians typically tell us that to endorse Platonic realism is to deny that properties, kinds, and relations need to be anchored in the spatiotemporal world. As they see it, the Platonist's universals are ontological "free fl oaters" with existence conditions that are independent of the concrete world of space and time. But to adopt this conception of universals, Aristotelians insist, is to embrace a "two-worlds" ontology of the sort we fi nd in Plato himself. On this view, we have a radical bifurcation in reality, with universals and concrete particulars occupying separate and unrelated realms. Such a bifurcation, Aristotelians claim, gives rise to insoluble problems in both metaphysics and epistemology. It is diffi cult to understand how there could be any kind of connection between spatiotemporal objects and beings completely outside space and time. Nonetheless, the realist is committed to there being such connections. After all, the cornerstone of metaphysical realism, the realistic interpretation of attribute agreement, tells us that the ontological ground of spatiotemporal particulars being the way they are, being the sorts of things they are, and being related to each other in the ways they are, just is their being connected or tied to properties, kinds, and relations. Furthermore, it is highly problematic how beings like ourselves who belong fi rmly to the spatiotemporal world of concrete particulars could ever have cognitive access to the nonspatial, nontemporal beings that Platonists tell us properties, kinds, and relations are. Since it would seem that there can be no causal relations between spatiotemporal particulars like ourselves and beings outside space and time, it looks as though the only story we could tell about our knowledge of universals is one that makes that knowledge innate or *apriori*. But Aristotelians have traditionally been skeptical of the idea of innate knowledge. They want to insist that our knowledge of properties, kinds, and

relations, like all our knowledge, has an empirical origin. Indeed, Aristotelians want to deny that we can separate or cut apart our knowledge of universals from our knowledge of concrete spatiotemporal particulars. As they see it, we grasp particulars only by grasping the kinds to which they belong, the properties they exhibit, and the relations they bear to each other; and we grasp the relevant kinds, properties, and relations, in turn, only by epistemic contact with the particulars that exemplify them.

How, in turn, do Platonists defend the idea of uninstantiated or unexemplified universals? One important strategy is to argue that precisely the same sorts of semantical considerations that lead us to posit exemplified universals support the claim that there are unexemplified universals. The Platonist will argue that it is not simply the predicates of true subject-predicate sentences that take universals as their referents; the same is true of false sentences of this form. Suppose there is an object, *a*, and a person, *P*, such that *P* falsely believes that:

(19) *a* is *F*

is true. *P* might well assertively utter (19). Although what *P* asserts in uttering (19) is false, *P* has asserted something. But what? Had (19) been true, in assertively uttering (19), *P* would have asserted that the object, *a*, exemplifies the universal, *F-ness*. The Platonist will argue that what *P* asserts in uttering (19) cannot depend on whether (19) is true or false, so what *P* falsely asserts in uttering (19) has to be the same thing *P* would have asserted had (19) been true. Thus, *P* asserts, falsely it turns out, that *a* exemplifies *F-ness*. But, the Platonist will go on, *F* might have been a general term, a shape-predicate, say, true of or satisfied by no object that exists, has existed, or will exist. So the semantical considerations that lead us to suppose that there are exemplified universals support an ontology of unexemplified universals as well; and, the Platonist may go on to argue, *F* could just as well have been a predicate that is necessarily true of nothing, so that the same argument would seem to justify the belief that there are necessarily unexemplified properties, kinds, and relations.

The Platonist will typically insist that all universals, whether exemplified or not, are *necessary beings*. Unlike the *contingently* existing particulars of common sense that exist but need not, properties, kinds, and relations are such that their nonexistence is impossible. Toward showing this, the Platonist tells us that for every property, the claim that it is a property is not just true, but necessarily true. Now, the Platonist insists that just as the truth of a claim about an object presupposes the actual existence of the object, the necessary truth of a claim about this or that object presupposes the necessary existence of the object. A necessary truth, the Platonist insists, is one that could not fail to be true; and where a necessary truth is a claim about a given object, the object in question could not fail to exist. So every property is such that it could not fail to exist; every property is a necessary being; and analogous points hold with regard to kinds and relations. So the Platonist insists that we distinguish between the

existence of a property, kind, or relation and its exemplification or instantiation. Whereas the latter may be contingent, the former never is.

In criticizing Aristotelians, the Platonist will argue that by failing to draw this distinction, the Aristotelian makes the existence of a universal depend upon the existence of something to exemplify it and thereby turns things upside down. Universals were brought on the scene to explain attribute agreement among particulars, to explain why concrete particulars are the way they are. Universals, then, are supposed to be ontologically prior to the particulars that exemplify them. On the Aristotelian view, however, things turn out just the reverse. The existence of a universal turns out to depend on there being particulars that are this or that sort of things, are characterized in this or that way, or are related to each other in this or that way. Such a view undermines the core insight motivating metaphysical realism.

Finally, although some realists (including, perhaps, Plato himself) are willing to endorse a "two-worlds" ontology, many Platonists will claim that Aristotelians are just wrong to suppose that the metaphysical problems of a "two-worlds" theory have to infect an ontology of unexemplifi ed universals. They will insist that, on their view, the nexus of exemplifi cation serves to tie universals and particulars, and they will claim that although this notion is ontologically basic or primitive, it is a perfectly respectable notion, one that the Aristotelian no less than the Platonist is committed to. And they will argue that the Aristotelian's contention that the Platonist faces insoluble epistemological problems is overblown. They will insist that while some universals have no instances in the spatiotemporal world, many do; and they will claim that our knowledge of exemplifi ed universals can be captured by a thoroughgoing empiricism. As they see it, we come to have cognitive access to these universals simply by experiencing the spatiotemporal particulars that exemplify them; whatever other knowledge we have of universals is grounded in our knowledge of these exemplifi ed universals. Thus, we come to know about some unexemplifi ed universals by extrapolation from our empirically based knowledge of instantiated properties, kinds, and relations. If there are universals that have no identifi able relations to the exemplifi ed universals we meet in our day-to-day commerce with the world, then Platonists will concede that we have no knowledge of such universals; but they will deny that this is surprising. They will claim, rather, that this is just what we would have expected.

Notes

1 An exception, of course, is the conceptual schemer we discussed in the Introduction.
2 *Parmenides* 130E–131A in Hamilton and Cairns (1961).
3 For twentieth-century expressions of the view we meet in the *Parmenides*, see Russell (1912: Chaps. IX and X), Strawson (1959: Chaps. V and VI), Donagan (1963), Wolterstorff (1973), Loux (1978a), and Armstrong (1989a).
4 The terms 'realism' and 'metaphysical realism' are the standard labels for this view; but the terms are also used to refer to a view about the nature of truth, the view that

there is a mind-independent world correspondence which renders each of our beliefs determinately true or false. Used in this sense, realism stands opposed to what is called *antirealism*. The conceptual schemers we discussed in the Introduction take an antirealistic stance on the nature of truth, whereas the defenders of a traditional conception of metaphysics as the attempt to characterize the general structure of reality are, in this latter sense, realists; but philosophers who are realists about truth can be, and often are, nominalists about attribute agreement. See Chapter 7, where the contrast between realism and antirealism is explained in depth.

5 Not all philosophers think the distinction between universals is as straightforward as I suggest. See, for example, Macbride (2005).

6 Other examples of kinds include the various ontological categories; they are simply the highest or most general kinds. Obviously, the philosopher who denies that there are kinds will need to find some metaphysically neutral way of characterizing what he is doing when he does metaphysics or attempts to identify the categories of being.

7 For a very clear statement of the view that subject-predicate discourse presupposes the existence of universals, see Donagan (1963: especially pp. 126–33). Where, as here, a paper appears in Loux (1976a), page references are to that volume.

8 For a more extensive treatment of correspondence and truth, see "Nominalisms about Propositions" in Chapter 4, and Chapter 7.

9 This kind of account is defended by Gustav Bergmann. See, for example, "The Philosophy of Malebranche," in Bergmann (1959: 190–1).

10 See, for example, Wolterstorff (1973: 85); Chap. V of Strawson (1959); and Loux (1978a: 30–3).

11 For an extended treatment of abstract reference and its ontological underpinnings, see Chap. IV of Loux (1978a).

12 See, for example, Roderick Chisholm, "Properties and states of affairs intentionally considered," in Chisholm (1989: 141–2) and van Inwagen (2006).

13 This is just the property version of what is called Russell's Paradox. In its more familiar class version, the paradox has as its upshot the moral that there is not a class for every membership condition. If there were, then there would be a class whose members are all and only the classes that are not members of themselves. But if there were such a class, then either it would be a member of itself or it would not be a member of itself. In either case, we would have a contradiction.

14 See *Parmenides* 131E–132B in Hamilton and Cairns (1961). For more recent discussions of realism and infinite regresses, see Strawson (1959: Chap. V), Donagan (1963: 135–9); Loux (1978a: 22–7), and Armstrong (1989a: 53–7).

15 This point is nicely made in Armstrong (1989a: 54–5).

16 Bradley (1930: 17–18).

17 See Wolterstorff (1973: 102).

18 See, for example, Donagan (1963: 138); Strawson, *Individuals* (1959: 169); and Bergmann's "Meaning," in Bergmann (1964: 87–8).

19 See, for example, Donagan (1963: 128–9); and Bergmann, "Two types of linguistic philosophy," in Bergmann (1954: 122).

20 Wittgenstein (1953: 66).

21 See, for example, Loux (1978a: 20–1).

22 See, for example, Armstrong (1989a: 87).

23 For a helpful discussion of supervenience, see Jaegwon Kim, "Concepts of supervenience," in Kim (1993: 53–78).

24 The issues discussed here are typically discussed in the philosophy of mind, where the status of the qualitative features of consciousness present problems for philosophers who endorse a strong version of materialism and hold that what exist are simply the objects postulated by our best physical theories. For a nice discussion of these issues and a statement of the eliminativist strategy, see Paul Churchland (1990: especially Chap. II).
25 See, for example, Donagan (1963: 131–3) and Loux (1978a: Chap. V).
26 See *Phaedo* 73A–81A and *Republic* 507B–507C in Hamilton and Cairns (1961).
27 See *Categories* 11 (14^a8–10) in McKeon (1941). A contemporary version of the Aristotelian view is defended in Armstrong (1989a: 75–82).
28 Most of the issues central to the dialectic that follows are discussed in Donagan (1963), Loux (1978a), and Armstrong (1989a); and Chisholm, "Properties and states of affairs intentionally considered," in Chisholm (1989: 141–2).

Further Reading

For the classical sources of metaphysical realism, the beginning student should read Plato's *Phaedo*, Books V–VII of the *Republic* and the opening sections of the *Parmenides*. Aristotle's discussions of Plato's views make for difficult reading, but the intrepid student is directed to *Metaphysics A*.6, *Metaphysics B*, and *Metaphysics Z*.13–16. Modern defenses of realism are often technical, but the student who reads chapters IX and X of Russell (1912), Donagan (1963), and Armstrong (1989a), Chisholm, "Properties and states of affairs intentionally considered," in Chisholm (1989) and van Inwagen (2006) should have a good foundation for reading any of the literature mentioned in the notes. The pieces by Russell and Armstrong can be found in *Metaphysics: Contemporary Readings*.

EUTHYPHRO

Characters

Socrates

Euthyphro

Scene—The Hall of the King*

2 EUTHYPHRO: What in the world are you doing here in the king's hall, Socrates? Why have you left your haunts in the Lyceum? You surely cannot have a suit before him, as I have.

SOCRATES: The Athenians, Euthyphro, call it an indictment, not a suit.

b EUTHYPHRO: What? Do you mean that someone is prosecuting you? I cannot believe that you are prosecuting anyone yourself.

SOCRATES: Certainly I am not.

EUTHYPHRO: Then is someone prosecuting you?

SOCRATES: Yes.

EUTHYPHRO: Who is he?

SOCRATES: I scarcely know him myself, Euthyphro; I think he must be some unknown young man. His name, however, is Meletus, and his district Pitthis, if you can call to mind any Meletus of that district—a hook-nosed man with lanky hair and rather a scanty beard.

EUTHYPHRO: I don't know him, Socrates. But tell me, what is he prosecuting you for?

c SOCRATES: What for? Not on trivial grounds, I think. It is no small thing for so young a man to have formed an opinion on such an important matter. For he, he says, knows how the young are corrupted, and who are their corrupters. He must be a wise d man who, observing my ignorance, is going to accuse me to the state, as his mother, of corrupting his friends. I think that he is the only one who begins at the right point in his political reforms; for his first care is to make the young men as good as possible, just as a good farmer will take care of his young plants first, and, after he has done that, of the 3 others. And so Meletus, I suppose, is first clearing us away who, as he says, corrupt the young men growing up; and then, when he has done that, of course he will turn his attention to the older men, and so become a very great public benefactor. Indeed, that is only what you would expect when he goes to work in this way.

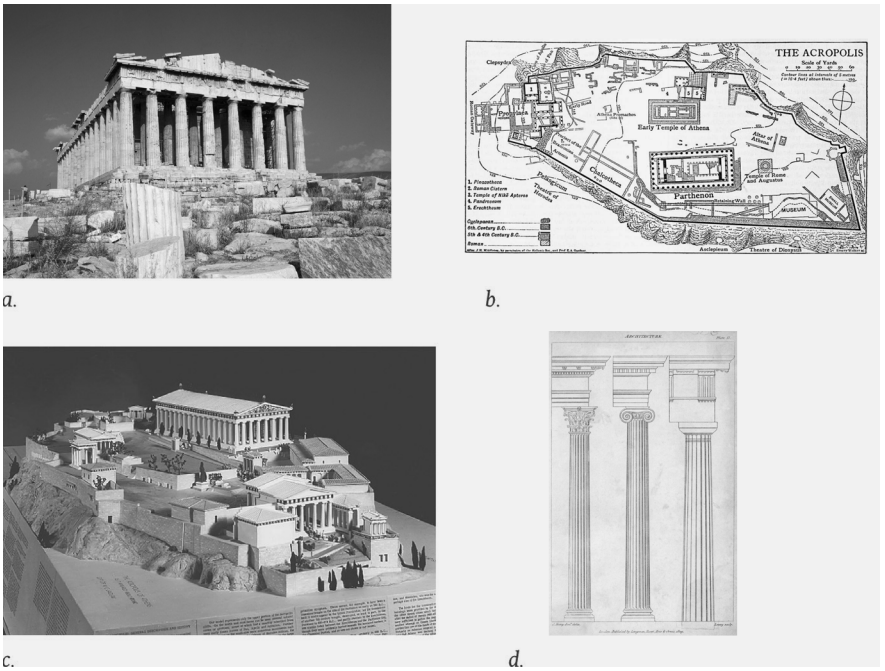
EUTHYPHRO: I hope it may be so, Socrates, but I fear the opposite. It seems to me that in trying to injure you, he is really setting to work by striking a blow at the foundation of the state. But how, tell me, does he say that you corrupt the youth?

b SOCRATES: In a way which sounds absurd at first, my friend. He says that I am a maker of gods; and so he is prosecuting me, he says, for inventing new gods and for not believing in the old ones.

EUTHYPHRO: I understand, Socrates. It is because you say that you always have a divine guide. So he is prosecuting you for introducing religious reforms; and he is going into court to arouse prejudice against you, knowing that the multitude are easily prejudiced

*The anachronistic title "king" was retained by the magistrate who had jurisdiction over crimes affecting the state religion.

Plato: *Euthyphro,* Apology, Crito, translated by F.J. Church (Pearson/Library of the Liberal Arts, 1987).



The Acropolis and the Parthenon

a. The *Parthenon*, Athens, built 477–438 B.C. The Parthenon, dedicated to Athena, patron deity of Athens, was at one period rededicated to the Christian Virgin Mary and then later became a Turkish mosque. In 1687 a gunpowder explosion created the ruin we see today. The Doric shell remains as a monument to ancient architectural engineering expertise and to a sense of classical beauty and order. (©*James Davis/Eye Ubiquitous/*Corbis)

b. Restored plan of the Acropolis, 400 B.C. The history of the Acropolis is as varied as the style and size of the temples and buildings constructed atop the ancient site. (*Public Domain*)

c. This model of the Acropolis of Athens recreates the complexity of fifth century B.C. public space, which included centers for worship, public forum, and entertainment. (With permission of the *Royal Ontario Museum* © *ROM*)

d. Doric, Ionic, and Corinthian columns with their characteristic capitals. (©*AS400 DB/Corbis*)

about such matters. Why, they laugh even at me, as if I were out of my mind, when I talk about divine things in the assembly and tell them what is going to happen; and yet I have c
never foretold anything which has not come true. But they are resentful of all people like us. We must not worry about them; we must meet them boldly.

SOCRATES: My dear Euthyphro, their ridicule is not a very serious matter. The Athenians, it seems to me, may think a man to be clever without paying him much attention, so long as they do not think that he teaches his wisdom to others. But as soon as they think that he makes other people clever, they get angry, whether it be from d
resentment, as you say, or for some other reason.

EUTHYPHRO: I am not very anxious to test their attitude toward me in this matter.

SOCRATES: No, perhaps they think that you are reserved, and that you are not anxious to teach your wisdom to others. But I fear that they may think that I am; for my love of men makes me talk to everyone whom I meet quite freely and unreservedly, and

without payment. Indeed, if I could I would gladly pay people myself to listen to me. If then, as I said just now, they were only going to laugh at me, as you say they do at you, it would not be at all an unpleasant way of spending the day—to spend it in court, joking and laughing. But if they are going to be in earnest, then only prophets like you can tell where the matter will end.

EUTHYPHRO: Well, Socrates, I dare say that nothing will come of it. Very likely you will be successful in your trial, and I think that I shall be in mine.

SOCRATES: And what is this suit of yours, Euthyphro? Are you suing, or being sued?

EUTHYPHRO: I am suing.

SOCRATES: Whom?

4 EUTHYPHRO: A man whom people think I must be mad to prosecute.

SOCRATES: What? Has he wings to fly away with?

EUTHYPHRO: He is far enough from flying; he is a very old man.

SOCRATES: Who is he?

EUTHYPHRO: He is my father.

SOCRATES: Your father, my good man?

EUTHYPHRO: He is indeed.

SOCRATES: What are you prosecuting him for? What is the accusation?

EUTHYPHRO: Murder, Socrates.

b SOCRATES: Good heavens, Euthyphro! Surely the multitude are ignorant of what is right. I take it that it is not everyone who could rightly do what you are doing; only a man who was already well advanced in wisdom.

EUTHYPHRO: That is quite true, Socrates.

SOCRATES: Was the man whom your father killed a relative of yours? But, of course, he was. You would never have prosecuted your father for the murder of a stranger?

EUTHYPHRO: You amuse me, Socrates. What difference does it make whether the murdered man were a relative or a stranger? The only question that you have to ask is, did the murderer kill justly or not? If justly, you must let him alone; if unjustly, c you must indict him for murder, even though he share your hearth and sit at your table. The pollution is the same if you associate with such a man, knowing what he has done, without purifying yourself, and him too, by bringing him to justice. In the present case the murdered man was a poor laborer of mine, who worked for us on our farm in Naxos. While drunk he got angry with one of our slaves and killed him. My father therefore bound the man hand and foot and threw him into a ditch, while he sent to Athens to ask the priest what he should do. While the messenger was gone, he entirely neglected the man, thinking that he was a murderer, and that it would be no d great matter, even if he were to die. And that was exactly what happened; hunger and cold and his bonds killed him before the messenger returned. And now my father and the rest of my family are indignant with me because I am prosecuting my father for the murder of this murderer. They assert that he did not kill the man at all; and they say that, even if he had killed him over and over again, the man himself was a mur- e derer, and that I ought not to concern myself about such a person because it is impious for a son to prosecute his father for murder. So little, Socrates, do they know the divine law of piety and impiety.

SOCRATES: And do you mean to say, Euthyphro, that you think that you understand divine things and piety and impiety so accurately that, in such a case as you have stated, you can bring your father to justice without fear that you yourself may be doing something impious?

EUTHYPHRO: If I did not understand all these matters accurately, Socrates, I should not be worth much—Euthyphro would not be any better than other men. 5

SOCRATES: Then, my dear Euthyphro, I cannot do better than become your pupil and challenge Meletus on this very point before the trial begins. I should say that I had always thought it very important to have knowledge about divine things; and that now, when he says that I offend by speaking carelessly about them, and by introducing reforms, I have become your pupil. And I should say, "Meletus, if you acknowledge b Euthyphro to be wise in these matters and to hold the correct belief, then think the same of me and do not put me on trial; but if you do not, then bring a suit, not against me, but against my master, for corrupting his elders—namely, myself whom he corrupts by his teaching, and his own father whom he corrupts by admonishing and punishing him." And if I did not succeed in persuading him to release me from the suit or to indict you in my place, then I could repeat my challenge in court.

EUTHYPHRO: Yes, by Zeus! Socrates, I think I should find out his weak points if he were to try to indict me. I should have a good deal to say about him in court long before c I spoke about myself.

SOCRATES: Yes, my dear friend, and knowing this I am anxious to become your pupil. I see that Meletus here, and others too, seem not to notice you at all, but he sees through me without difficulty and at once prosecutes me for impiety. Now, therefore, please explain to me what you were so confident just now that you knew. Tell me what d are righteousness and sacrilege with respect to murder and everything else. I suppose that piety is the same in all actions, and that impiety is always the opposite of piety, and retains its identity, and that, as impiety, it always has the same character, which will be found in whatever is impious.

EUTHYPHRO: Certainly, Socrates, I suppose so.

SOCRATES: Tell me, then, what is piety and what is impiety?

EUTHYPHRO: Well, then, I say that piety means prosecuting the unjust individual who has committed murder or sacrilege, or any other such crime, as I am doing now, whether he is your father or your mother or whoever he is; and I say that impiety means e not prosecuting him. And observe, Socrates, I will give you a clear proof, which I have already given to others, that it is so, and that doing right means not letting off unpunished the sacrilegious man, whosoever he may be. Men hold Zeus to be the best and the most just of the gods; and they admit that Zeus bound his own father, Cronos, for 6 wrongfully devouring his children; and that Cronos, in his turn, castrated his father for similar reasons. And yet these same men are incensed with me because I proceed against my father for doing wrong. So, you see, they say one thing in the case of the gods and quite another in mine.

SOCRATES: Is not that why I am being prosecuted, Euthyphro? I mean, because I find it hard to accept such stories people tell about the gods? I expect that I shall be found at fault because I doubt those stories. Now if you who understand all these matters so well agree in holding all those tales true, then I suppose that I must yield to your b authority. What could I say when I admit myself that I know nothing about them? But tell me, in the name of friendship, do you really believe that these things have actually happened?

EUTHYPHRO: Yes, and more amazing things, too, Socrates, which the multitude do not know of.

SOCRATES: Then you really believe that there is war among the gods, and bitter hatreds, and battles, such as the poets tell of, and which the great painters have depicted c in our temples, notably in the pictures which cover the robe that is carried up to the

Acropolis at the great Panathenaic festival? Are we to say that these things are true, Euthyphro?

EUTHYPHRO: Yes, Socrates, and more besides. As I was saying, I will report to you many other stories about divine matters, if you like, which I am sure will astonish you when you hear them.

SOCRATES: I dare say. You shall report them to me at your leisure another time. At present please try to give a more definite answer to the question which I asked you just
d now. What I asked you, my friend, was, What is piety? and you have not explained it to me to my satisfaction. You only tell me that what you are doing now, namely, prosecuting your father for murder, is a pious act.

EUTHYPHRO: Well, that is true, Socrates.

SOCRATES: Very likely. But many other actions are pious, are they not, Euthyphro?

EUTHYPHRO: Certainly.

SOCRATES: Remember, then, I did not ask you to tell me one or two of all the many
e pious actions that there are; I want to know what is characteristic of piety which makes all pious actions pious. You said, I think, that there is one characteristic which makes all pious actions pious, and another characteristic which makes all impious actions impious. Do you not remember?

EUTHYPHRO: I do.

SOCRATES: Well, then, explain to me what is this characteristic, that I may have it to turn to, and to use as a standard whereby to judge your actions and those of other men, and be able to say that whatever action resembles it is pious, and whatever does not, is not pious.

EUTHYPHRO: Yes, I will tell you that if you wish, Socrates.

SOCRATES: Certainly I do.

7 EUTHYPHRO: Well, then, what is pleasing to the gods is pious, and what is not pleasing to them is impious.

SOCRATES: Fine, Euthyphro. Now you have given me the answer that I wanted. Whether what you say is true, I do not know yet. But, of course, you will go on to prove that it is true.

EUTHYPHRO: Certainly.

SOCRATES: Come, then, let us examine our statement. The things and the men that are pleasing to the gods are pious, and the things and the men that are displeasing to the gods are impious. But piety and impiety are not the same; they are as opposite as possible—was not that what we said?

EUTHYPHRO: Certainly.

SOCRATES: And it seems the appropriate statement?

b EUTHYPHRO: Yes, Socrates, certainly.

SOCRATES: Have we not also said, Euthyphro, that there are quarrels and disagreements and hatreds among the gods?

EUTHYPHRO: We have.

SOCRATES: But what kind of disagreement, my friend, causes hatred and anger? Let us look at the matter thus. If you and I were to disagree as to whether one number
c were more than another, would that make us angry and enemies? Should we not settle such a dispute at once by counting?

EUTHYPHRO: Of course.

SOCRATES: And if we were to disagree as to the relative size of two things, we should measure them and put an end to the disagreement at once, should we not?

EUTHYPHRO: Yes.

SOCRATES: And should we not settle a question about the relative weight of two things by weighing them?

EUTHYPHRO: Of course.

SOCRATES: Then what is the question which would make us angry and enemies if we disagreed about it, and could not come to a settlement? Perhaps you have not an d
answer ready; but listen to mine. Is it not the question of the just and unjust, of the honorable and the dishonorable, of the good and the bad? Is it not questions about these matters which make you and me and everyone else quarrel, when we do quarrel, if we differ about them and can reach no satisfactory agreement?

EUTHYPHRO: Yes, Socrates, it is disagreements about these matters.

SOCRATES: Well, Euthyphro, the gods will quarrel over these things if they quarrel at all, will they not?

EUTHYPHRO: Necessarily.

SOCRATES: Then, my good Euthyphro, you say that some of the gods think one e
thing just, the others another; and that what some of them hold to be honorable or good, others hold to be dishonorable or evil. For there would not have been quarrels among them if they had not disagreed on these points, would there?

EUTHYPHRO: You are right.

SOCRATES: And each of them loves what he thinks honorable, and good, and just; and hates the opposite, does he not?

EUTHYPHRO: Certainly.

SOCRATES: But you say that the same action is held by some of them to be just, and by others to be unjust; and that then they dispute about it, and so quarrel and fight 8
among themselves. Is it not so?

EUTHYPHRO: Yes.

SOCRATES: Then the same thing is hated by the gods and loved by them; and the same thing will be displeasing and pleasing to them.

EUTHYPHRO: Apparently.

SOCRATES: Then, according to your account, the same thing will be pious and impious.

EUTHYPHRO: So it seems.

SOCRATES: Then, my good friend, you have not answered my question. I did not ask you to tell me what action is both pious and impious; but it seems that whatever is pleasing to the gods is also displeasing to them. And so, Euthyphro, I should not be sur- b
prised if what you are doing now in punishing your father is an action well pleasing to Zeus, but hateful to Cronos and Uranus, and acceptable to Hephaestus, but hateful to Hera; and if any of the other gods disagree about it, pleasing to some of them and displeasing to others.

EUTHYPHRO: But on this point, Socrates, I think that there is no difference of opinion among the gods: they all hold that if one man kills another unjustly, he must be punished.

SOCRATES: What, Euthyphro? Among mankind, have you never heard disputes c
whether a man ought to be punished for killing another man unjustly, or for doing some other unjust deed?

EUTHYPHRO: Indeed, they never cease from these disputes, especially in courts of justice. They do all manner of unjust things; and then there is nothing which they will not do and say to avoid punishment.

SOCRATES: Do they admit that they have done something unjust, and at the same time deny that they ought to be punished, Euthyphro?

EUTHYPHRO: No, indeed, that they do not.

SOCRATES: Then it is not the case that there is nothing which they will not do and say. I take it, they do not dare to say or argue that they must not be punished if they have
d done something unjust. What they say is that they have not done anything unjust, is it not so?

EUTHYPHRO: That is true.

SOCRATES: Then they do not disagree over the question that the unjust individual must be punished. They disagree over the question, who is unjust, and what was done and when, do they not?

EUTHYPHRO: That is true.

SOCRATES: Well, is not exactly the same thing true of the gods if they quarrel about justice and injustice, as you say they do? Do not some of them say that the others are doing something unjust, while the others deny it? No one, I suppose, my dear friend,
e whether god or man, dares to say that a person who has done something unjust must not be punished.

EUTHYPHRO: No, Socrates, that is true, by and large.

SOCRATES: I take it, Euthyphro, that the disputants, whether men or gods, if the gods do disagree, disagree over each separate act. When they quarrel about any act, some of them say that it was just, and others that it was unjust. Is it not so?

EUTHYPHRO: Yes.

9 SOCRATES: Come, then, my dear Euthyphro, please enlighten me on this point. What proof have you that all the gods think that a laborer who has been imprisoned for murder by the master of the man whom he has murdered, and who dies from his imprisonment before the master has had time to learn from the religious authorities what he should do, dies unjustly? How do you know that it is just for a son to indict his father and to prosecute him for the murder of such a man? Come, see if you can make it clear
b to me that the gods necessarily agree in thinking that this action of yours is just; and if you satisfy me, I will never cease singing your praises for wisdom.

EUTHYPHRO: I could make that clear enough to you, Socrates; but I am afraid that it would be a long business.

SOCRATES: I see you think that I am duller than the judges. To them, of course, you will make it clear that your father has committed an unjust action, and that all the gods agree in hating such actions.

EUTHYPHRO: I will indeed, Socrates, if they will only listen to me.

c SOCRATES: They will listen if they think that you are a good speaker. But while you were talking, it occurred to me to ask myself this question: suppose that Euthyphro were to prove to me as clearly as possible that all the gods think such a death unjust, how has he brought me any nearer to understanding what piety and impiety are? This particular act, perhaps, may be displeasing to the gods, but then we have just seen that piety and impiety cannot be defined in that way; for we have seen that what is displeas-
d ing to the gods is also pleasing to them. So I will let you off on this point, Euthyphro; and all the gods shall agree in thinking your father's action wrong and in hating it, if you like. But shall we correct our definition and say that whatever all the gods hate is impious, and whatever they all love is pious; while whatever some of them love, and others hate, is either both or neither? Do you wish us now to define piety and impiety in this manner?

EUTHYPHRO: Why not, Socrates?

SOCRATES: There is no reason why I should not, Euthyphro. It is for you to consider whether that definition will help you to teach me what you promised.

EUTHYPHRO: Well, I should say that piety is what all the gods love, and that impiety is what they all hate. e

SOCRATES: Are we to examine this definition, Euthyphro, and see if it is a good one? Or are we to be content to accept the bare statements of other men or of ourselves without asking any questions? Or must we examine the statements?

EUTHYPHRO: We must examine them. But for my part I think that the definition is right this time.

SOCRATES: We shall know that better in a little while, my good friend. Now consider this question. Do the gods love piety because it is pious, or is it pious because they love it? 10

EUTHYPHRO: I do not understand you, Socrates.

SOCRATES: I will try to explain myself: we speak of a thing being carried and carrying, and being led and leading, and being seen and seeing; and you understand that all such expressions mean different things, and what the difference is.

EUTHYPHRO: Yes, I think I understand.

SOCRATES: And we talk of a thing being loved, of a thing loving, and the two are different?

EUTHYPHRO: Of course.

SOCRATES: Now tell me, is a thing which is being carried in a state of being carried because it is carried, or for some other reason? b

EUTHYPHRO: No, because it is carried.

SOCRATES: And a thing is in a state of being led because it is led, and of being seen because it is seen?

EUTHYPHRO: Certainly.

SOCRATES: Then a thing is not seen because it is in a state of being seen: it is in a state of being seen because it is seen; and a thing is not led because it is in a state of being led: it is in a state of being led because it is led; and a thing is not carried because it is in a state of being carried: it is in a state of being carried because it is carried. Is my meaning clear now, Euthyphro? I mean this: if anything becomes or is affected, it does not become because it is in a state of becoming: it is in a state of becoming because it becomes; and it is not affected because it is in a state of being affected: it is in a state of being affected because it is affected. Do you not agree? c

EUTHYPHRO: I do.

SOCRATES: Is not that which is being loved in a state either of becoming or of being affected in some way by something?

EUTHYPHRO: Certainly.

SOCRATES: Then the same is true here as in the former cases. A thing is not loved by those who love it because it is in a state of being loved; it is in a state of being loved because they love it.

EUTHYPHRO: Necessarily.

SOCRATES: Well, then, Euthyphro, what do we say about piety? Is it not loved by all the gods, according to your definition? d

EUTHYPHRO: Yes.

SOCRATES: Because it is pious, or for some other reason?

EUTHYPHRO: No, because it is pious.

SOCRATES: Then it is loved by the gods because it is pious; it is not pious because it is loved by them?

EUTHYPHRO: It seems so.

SOCRATES: But, then, what is pleasing to the gods is pleasing to them, and is in a state of being loved by them, because they love it?

EUTHYPHRO: Of course.

SOCRATES: Then piety is not what is pleasing to the gods, and what is pleasing to the gods is not pious, as you say, Euthyphro. They are different things.

e EUTHYPHRO: And why, Socrates?

SOCRATES: Because we are agreed that the gods love piety because it is pious, and that it is not pious because they love it. Is not this so?

EUTHYPHRO: Yes.

SOCRATES: And that what is pleasing to the gods because they love it, is pleasing to them by reason of this same love, and that they do not love it because it is pleasing to them.

EUTHYPHRO: True.

SOCRATES: Then, my dear Euthyphro, piety and what is pleasing to the gods are
11 different things. If the gods had loved piety because it is pious, they would also have loved what is pleasing to them because it is pleasing to them; but if what is pleasing to them had been pleasing to them because they loved it, then piety, too, would have been piety because they loved it. But now you see that they are opposite things, and wholly different from each other. For the one is of a sort to be loved because it is loved, while the other is loved because it is of a sort to be loved. My question, Euthyphro, was, What is piety? But it turns out that you have not explained to me the essential character of piety; you have been content to mention an effect which belongs to it—namely, that all
b the gods love it. You have not yet told me what its essential character is. Do not, if you please, keep from me what piety is; begin again and tell me that. Never mind whether the gods love it, or whether it has other effects: we shall not differ on that point. Do your best to make clear to me what is piety and what is impiety.

EUTHYPHRO: But, Socrates, I really don't know how to explain to you what is in my mind. Whatever statement we put forward always somehow moves round in a circle, and will not stay where we put it.

SOCRATES: I think that your statements, Euthyphro, are worthy of my ancestor
c Daedalus.* If they had been mine and I had set them down, I dare say you would have made fun of me, and said that it was the consequence of my descent from Daedalus that the statements which I construct run away, as his statues used to, and will not stay where they are put. But, as it is, the statements are yours, and the joke would have no point. You yourself see that they will not stay still.

EUTHYPHRO: Nay, Socrates, I think that the joke is very much in point. It is not my
d fault that the statement moves round in a circle and will not stay still. But you are the Daedalus, I think; as far as I am concerned, my statements would have stayed put.

SOCRATES: Then, my friend, I must be a more skillful artist than Daedalus; he only used to make his own works move, while I, you see, can make other people's works move, too. And the beauty of it is that I am wise against my will. I would rather that our statements had remained firm and immovable than have all the wisdom of Daedalus and
e all the riches of Tantalus to boot. But enough of this. I will do my best to help you to explain to me what piety is, for I think that you are lazy. Don't give in yet. Tell me, do you not think that all piety must be just?

EUTHYPHRO: I do.

12 SOCRATES: Well, then, is all justice pious, too? Or, while all piety is just, is a part only of justice pious, and the rest of it something else?

EUTHYPHRO: I do not follow you, Socrates.

*Daedalus' statues were reputed to have been so lifelike that they came alive.

SOCRATES: Yet you have the advantage over me in your youth no less than your wisdom. But, as I say, the wealth of your wisdom makes you complacent. Exert yourself, my good friend: I am not asking you a difficult question. I mean the opposite of what the poet [Stasinus] said, when he wrote:

> "You shall not name Zeus the creator, who made all things: for where there is fear there also is reverence." b

Now I disagree with the poet. Shall I tell you why?

EUTHYPHRO: Yes.

SOCRATES: I do not think it true to say that where there is fear, there also is reverence. Many people who fear sickness and poverty and other such evils seem to me to have fear, but no reverence for what they fear. Do you not think so?

EUTHYPHRO: I do.

SOCRATES: But I think that where there is reverence there also is fear. Does any man feel reverence and a sense of shame about anything, without at the same time c dreading and fearing the reputation of wickedness?

EUTHYPHRO: No, certainly not.

SOCRATES: Then, though there is fear wherever there is reverence, it is not correct to say that where there is fear there also is reverence. Reverence does not always accompany fear; for fear, I take it, is wider than reverence. It is a part of fear, just as the odd is a part of number, so that where you have the odd you must also have number, though where you have number you do not necessarily have the odd. Now I think you follow me?

EUTHYPHRO: I do.

SOCRATES: Well, then, this is what I meant by the question which I asked you. Is there always piety where there is justice? Or, though there is always justice where there is piety, yet there is not always piety where there is justice, because piety is only a part d of justice? Shall we say this, or do you differ?

EUTHYPHRO: No, I agree. I think that you are right.

SOCRATES: Now observe the next point. If piety is a part of justice, we must find out, I suppose, what part of justice it is? Now, if you had asked me just now, for instance, what part of number is the odd, and what number is an odd number, I should have said that whatever number is not even is an odd number. Is it not so?

EUTHYPHRO: Yes.

SOCRATES: Then see if you can explain to me what part of justice is piety, that I e may tell Meletus that now that I have been adequately instructed by you as to what actions are righteous and pious, and what are not, he must give up prosecuting me unjustly for impiety.

EUTHYPHRO: Well, then, Socrates, I should say that righteousness and piety are that part of justice which has to do with the careful attention which ought to be paid to the gods; and that what has to do with the careful attention which ought to be paid to men is the remaining part of justice.

SOCRATES: And I think that your answer is a good one, Euthyphro. But there is one little point about which I still want to hear more. I do not yet understand what the 13 careful attention is to which you refer. I suppose you do not mean that the attention which we pay to the gods is like the attention which we pay to other things. We say, for instance, do we not, that not everyone knows how to take care of horses, but only the trainer of horses?

EUTHYPHRO: Certainly.

SOCRATES: For I suppose that the skill that is concerned with horses is the art of taking care of horses.

EUTHYPHRO: Yes.

SOCRATES: And not everyone understands the care of dogs, but only the huntsman.

EUTHYPHRO: True.

b SOCRATES: For I suppose that the huntsman's skill is the art of taking care of dogs.

EUTHYPHRO: Yes.

SOCRATES: And the herdsman's skill is the art of taking care of cattle.

EUTHYPHRO: Certainly.

SOCRATES: And you say that piety and righteousness are taking care of the gods, Euthyphro?

EUTHYPHRO: I do.

SOCRATES: Well, then, has not all care the same object? Is it not for the good and benefit of that on which it is bestowed? For instance, you see that horses are benefited and improved when they are cared for by the art which is concerned with them. Is it not so?

EUTHYPHRO: Yes, I think so.

c SOCRATES: And dogs are benefited and improved by the huntsman's art, and cattle by the herdsman's, are they not? And the same is always true. Or do you think care is ever meant to harm that which is cared for?

EUTHYPHRO: No, indeed; certainly not.

SOCRATES: But to benefit it?

EUTHYPHRO: Of course.

SOCRATES: Then is piety, which is our care for the gods, intended to benefit the gods, or to improve them? Should you allow that you make any of the gods better when you do a pious action?

EUTHYPHRO: No indeed; certainly not.

SOCRATES: No, I am quite sure that that is not your meaning, Euthyphro. It was for
d that reason that I asked you what you meant by the careful attention which ought to be paid to the gods. I thought that you did not mean that.

EUTHYPHRO: You were right, Socrates. I do not mean that.

SOCRATES: Good. Then what sort of attention to the gods will piety be?

EUTHYPHRO: The sort of attention, Socrates, slaves pay to their masters.

SOCRATES: I understand; then it is a kind of service to the gods?

EUTHYPHRO: Certainly.

SOCRATES: Can you tell me what result the art which serves a doctor serves to produce? Is it not health?

EUTHYPHRO: Yes.

e SOCRATES: And what result does the art which serves a ship-wright serve to produce?

EUTHYPHRO: A ship, of course, Socrates.

SOCRATES: The result of the art which serves a builder is a house, is it not?

EUTHYPHRO: Yes.

SOCRATES: Then tell me, my good friend: What result will the art which serves the gods serve to produce? You must know, seeing that you say that you know more about divine things than any other man.

EUTHYPHRO: Well, that is true, Socrates.

SOCRATES: Then tell me, I beg you, what is that grand result which the gods use our services to produce?

EUTHYPHRO: There are many notable results, Socrates.

SOCRATES: So are those, my friend, which a general produces. Yet it is easy to see 14
that the crowning result of them all is victory in war, is it not?

EUTHYPHRO: Of course.

SOCRATES: And, I take it, the farmer produces many notable results; yet the principal result of them all is that he makes the earth produce food.

EUTHYPHRO: Certainly.

SOCRATES: Well, then, what is the principal result of the many notable results which the gods produce?

EUTHYPHRO: I told you just now, Socrates, that accurate knowledge of all these matters is not easily obtained. However, broadly I say this: if any man knows that his b
words and actions in prayer and sacrifice are acceptable to the gods, that is what is pious; and it preserves the state, as it does private families. But the opposite of what is acceptable to the gods is sacrilegious, and this it is that undermines and destroys everything.

SOCRATES: Certainly, Euthyphro, if you had wished, you could have answered my main question in far fewer words. But you are evidently not anxious to teach me. Just c
now, when you were on the very point of telling me what I want to know, you stopped short. If you had gone on then, I should have learned from you clearly enough by this time what piety is. But now I am asking you questions, and must follow wherever you lead me; so tell me, what is it that you mean by piety and impiety? Do you not mean a science of prayer and sacrifice?

EUTHYPHRO: I do.

SOCRATES: To sacrifice is to give to the gods, and to pray is to ask of them, is it not?

EUTHYPHRO: It is, Socrates.

SOCRATES: Then you say that piety is the science of asking of the gods and giving d
to them?

EUTHYPHRO: You understand my meaning exactly, Socrates.

SOCRATES: Yes, for I am eager to share your wisdom, Euthyphro, and so I am all attention; nothing that you say will fall to the ground. But tell me, what is this service of the gods? You say it is to ask of them, and to give to them?

EUTHYPHRO: I do.

SOCRATES: Then, to ask rightly will be to ask of them what we stand in need of e
from them, will it not?

EUTHYPHRO: Naturally.

SOCRATES: And to give rightly will be to give back to them what they stand in need of from us? It would not be very skillful to make a present to a man of something that he has no need of.

EUTHYPHRO: True, Socrates.

SOCRATES: Then piety, Euthyphro, will be the art of carrying on business between gods and men?

EUTHYPHRO: Yes, if you like to call it so.

SOCRATES: But I like nothing except what is true. But tell me, how are the gods benefited by the gifts which they receive from us? What they give is plain enough. Every good thing that we have is their gift. But how are they benefited by what we give them? 15
Have we the advantage over them in these business transactions to such an extent that we receive from them all the good things we possess, and give them nothing in return?

EUTHYPHRO: But do you suppose, Socrates, that the gods are benefited by the gifts which they receive from us?

SOCRATES: But what *are* these gifts, Euthyphro, that we give the gods?

EUTHYPHRO: What do you think but honor and praise, and, as I have said, what is acceptable to them.

b SOCRATES: Then piety, Euthyphro, is acceptable to the gods, but it is not profitable to them nor loved by them?

EUTHYPHRO: I think that nothing is more loved by them.

SOCRATES: Then I see that piety means that which is loved by the gods.

EUTHYPHRO: Most certainly.

SOCRATES: After that, shall you be surprised to find that your statements move about instead of staying where you put them? Shall you accuse me of being the Daedalus that makes them move, when you yourself are far more skillful than Daedalus was, and make them go round in a circle? Do you not see that our statement has come
c round to where it was before? Surely you remember that we have already seen that piety and what is pleasing to the gods are quite different things. Do you not remember?

EUTHYPHRO: I do.

SOCRATES: And now do you not see that you say that what the gods love is pious? But does not what the gods love come to the same thing as what is pleasing to the gods?

EUTHYPHRO: Certainly.

SOCRATES: Then either our former conclusion was wrong or, if it was right, we are wrong now.

EUTHYPHRO: So it seems.

SOCRATES: Then we must begin again and inquire what piety is. I do not mean to
d give in until I have found out. Do not regard me as unworthy; give your whole mind to the question, and this time tell me the truth. For if anyone knows it, it is you; and you are a Proteus whom I must not let go until you have told me. It cannot be that you would ever have undertaken to prosecute your aged father for the murder of a laboring man unless you had known exactly what piety and impiety are. You would have feared to risk the anger of the gods, in case you should be doing wrong, and you would have been
e afraid of what men would say. But now I am sure that you think that you know exactly what is pious and what is not; so tell me, my good Euthyphro, and do not conceal from me what you think.

EUTHYPHRO: Another time, then, Socrates. I am in a hurry now, and it is time for me to be off.

SOCRATES: What are you doing, my friend! Will you go away and destroy all my hopes of learning from you what is pious and what is not, and so of escaping Meletus?
16 I meant to explain to him that now Euthyphro has made me wise about divine things, and that I no longer in my ignorance speak carelessly about them or introduce reforms. And then I was going to promise him to live a better life for the future.

1 The Why, What and How of Philosophy of Economics

Overview

The philosophy of economics is at the same time an ancient and a very recent discipline. It is ancient in that the world's greatest economists beginning with Aristotle were also or mainly philosophers, and many of their contributions should be classified as contributions to the *philosophy* of economics rather than the *science* of economics, narrowly understood. With the increasing specialization and professionalization of academic disciplines that occurred in the nineteenth century, economics was separated from philosophy and developed, especially after the Second World War, a mainstream paradigm that was hostile to philosophical reflection. At the same time, philosophers of science were mainly interested in natural science and thus tended to ignore economics and other social sciences. It is only in the last 30 or so years that we can once more experience a mutual interest and exchange, and witness the development of academic institutions that focus on the intersection of economics and philosophy. In that sense, then, the discipline is a novel one. This chapter will explore why philosophy of economics is a subject worth studying, explaining its various branches and the overall approach and narratives in evidence in this book.

Two Opposing Paradigms

When I am being introduced to someone I haven't met before and my new acquaintance asks me what I am doing, they often look at me in surprise, puzzlement or sheer disbelief when I tell them that I am a "philosopher of economics." Aren't philosophy and economics two completely different kettles of fish? Isn't economics a science that deals in facts which can be expressed in figures and equations, and isn't philosophy a discipline belonging to the humanities, more akin to the arts than the sciences, and dealing with ideas rather than data? Somewhat more provocatively, aren't economists cold-hearted proponents of free markets and individual responsibility and philosophers naive believers in idealistic principles and the human good?

No doubt there is something to these stereotypes. Observing economists and philosophers at their respective academic conferences gives some evidence beyond the platitudes that the two fields are indeed dominated by quite different cultures. But to some extent both disciplines have become more open to ideas from the outside, and values other than their own now play a role in each discipline. To give just a couple of examples of sub-disciplines of economics and philosophy where untypical attitudes have a great influence on the debate, take happiness economics and formal ethics. Happiness economics studies the causes and effects of subjective well-being. It is highly interdisciplinary and often combines economic analysis with work from other fields such as psychology, sociology and philosophy. It is different from traditional welfare economics in that it rests on a radically different concept of well-being (see Chapter 12 for details). Whether this or that concept of well-being is adequate is of course one of the major issues any philosopher working on ethical theory has to address. Formal ethics, in turn, is a branch of philosophy that addresses traditional philosophical issues using tools drawn from economics such as rational-choice theory and game theory. The values of rigor and mathematical elegance, formally more characteristic of economics than of philosophy departments, surely influence the debate in no small measure.

The separation of economics and philosophy into two disciplines has in fact occurred fairly recently and is to a large extent artificial. Indeed, many of the world's greatest economists were also, or even mainly, philosophers: Adam Smith, David Hume, John Stuart Mill, Karl Marx, William Stanley Jevons, to some extent John Maynard Keynes and more recently Amartya Sen. Aristotle is often said to be the first economist, but of course he is better known as one of the greatest philosophers of all time.

The separation into distinct disciplines has to do with the general trend towards greater specialization that all sciences have experienced and continue to experience but also with a more specific stance towards science, including social science. This stance, sometimes called "modernism" (McCloskey 1983), takes the view that (a) science investigates facts, and only facts; (b) factual knowledge is exhausted by what can be ascertained on the basis of observations and experiments.

Both clauses (a) and (b) serve to separate the science of economics from other kinds of inquiry. Science, we are told, examines facts, or what there is, and not values, or what ought to be. According to this view, the economist *qua* scientist abstains from value judgments. Objective, scientific knowledge is value-free. Value judgments are a problem for ethicists. Moreover, in order to be objective, our knowledge has to be based on observable features of the world. Science deals with verifiable states of affairs, not with speculations that go beyond what is accessible to the senses. A classic statement of this perspective stems from the greatest of the empiricists, David Hume:

> When we run over libraries, persuaded of these principles, what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: For it can contain nothing but sophistry and illusion.
>
> (Hume 1999 [1748]: 211; original emphasis)

Less radical thinkers of this orientation would perhaps stop short of Hume's call to arms but they'd nevertheless agree that there is a clear separation between science, which is based on numbers and observable facts, and metaphysics, which is based on hypothesis and speculation.

We are therefore facing a dichotomy: what is versus what ought to be; and what is ascertainable by observation versus what is speculative. According to this view, then, economists *qua* scientists stay on the safe side of these dichotomies. By contrast, philosophers *qua* ethicists deal with value judgments and what ought to be; philosophers *qua* metaphysicians deal with speculations about the ultimate constituents of reality.

Parallel developments in both economics and philosophy in the second half of the twentieth century have helped to overcome these dichotomies. On the one hand, at least some economists, while still emphasizing the distinction between so-called positive (or factual) and normative (or evaluative) economics have come to realize that they cannot shy away from value judgments altogether. Especially through the work of Amartya Sen it has transpired that an economist *qua* scientist needs to engage in ethical query (e.g., Sen 1987). On the other hand, economists have stopped insisting that economic knowledge is exhausted by what is observable. To give one example, economists now actively participate in discussions concerning the notion of cause, a concept once deemed too metaphysical for scientists (Hoover 2009). To give another, the notion of "revealed preference," once endorsed by the profession because it allows economists to avoid making assumptions concerning unobservable states of affairs, has come under severe attack (Hausman 2012).

Philosophy as a discipline has changed, too. While continuing to deal with traditional ethical questions such as those concerning the nature of the

good and principles of justice, as well as traditional metaphysical questions such as those concerning the nature of causality and laws of nature, philosophers do so in ways increasingly informed by and continuous with science. Much of recent philosophy therefore resembles science to some extent, for example in the use of empirical information, mathematical modeling and sometimes even experimental methods.

In other words, economics and philosophy have drawn closer to each other by economists having started to ask questions that were once considered philosophical in the pejorative sense of "non-scientific," and by philosophers having started to address their questions in ways that resemble science more closely than traditional philosophy. A consequence of this convergence is that a lot of work that is now being done in economics and philosophy departments, discussed at academic conferences and published in economics journals or philosophy journals resists straightforward categorization as either "straight economics" or "straight philosophy." It is work at this intersection of economics and philosophy that this book is concerned with.

The recent (2008–) financial crisis provides an independent reason why philosophy of economics is an area of research of potentially very high significance, both academic and practical. Many commentators, among them Nobel prize-winning economists, have blamed the crisis on a failure of economics as a discipline. Here are some prominent voices:

> Of all the economic bubbles that have been pricked [since 2008], few have burst more spectacularly than the reputation of economics itself.
>
> (*The Economist* 2009)

> Last year, everything came apart.
>
> Few economists saw our current crisis coming, but this predictive failure was the least of the field's problems. More important was the profession's blindness to the very possibility of catastrophic failures in a market economy.
>
> (Krugman 2009a)

> The main cause of the crisis was the behavior of the banks—largely a result of misguided incentives unrestrained by good regulation. …
>
> There is one other set of accomplices—the economists who provided the arguments that those in the financial markets found so convenient and self-serving. These economists provided models—based on unrealistic assumptions of perfect information, perfect competition, and perfect markets—in which regulation was unnecessary.
>
> (Stiglitz 2009)

Paraphrasing Krugman and Stiglitz, we might say that among the causes of the financial crisis were economic models that were idealized to such an extent and in such a way that they couldn't be used for salient purposes

such as predicting financial crises like the present one and underwriting policy interventions such as banking regulation. But are these critics justified in their allegations? Are they right in their rejection of current mainstream models and in blaming the crisis on these models? An unsympathetic observer might point out that both Krugman and Stiglitz have axes to grind. Specifically, both are supporters of alternative economic paradigms. Krugman's article pursues an unashamed Keynesian agenda; Stiglitz is well known for his advocacy of models of imperfect and incomplete information (which have, of course, different regulatory implications).

Krugman and Stiglitz thus criticize mainstream economics for using bad theories—theories that make assumptions about markets that are unrealistic, theories that do not allow for "catastrophic failure" despite its apparent empirical reality. Krugman's article points to another aspect. In his analysis of what went wrong Krugman writes, "As I see it, the economics profession went astray because economists, as a group, mistook beauty, clad in impressive-looking mathematics, for truth" (Krugman 2009a). Krugman, in other words, criticizes economists for using bad methodology: pursuing mathematical elegance rather than truth leads to models that both fail to predict significant economic events and fail to provide good policy advice.

Philosophers, or more specifically philosophers of science, think about questions concerning theory assessment and scientific methodology professionally. More generally speaking, philosophers of science are interested in metaphysical and epistemological issues raised by the sciences. Metaphysical issues are those pertaining to the most fundamental building blocks of nature and society such as objects, properties, individuals, laws of nature, social norms, causality and modality (possibility and necessity). Epistemological issues concern the ways in which scientists find out about these in experiments, measurements and observation. If philosophers of science are any good at their jobs, the tools, concepts and theories they come up with should help us judge whether Krugman's and Stiglitz's points concerning theory assessment and methodology are as compelling as they make them seem.

There is a third aspect. Some have argued, like Stiglitz, that the behavior of the banks was one of the main causes of the crisis. But unlike him, they do not see the failure in unrealistic idealizations but rather in inappropriate moral foundations of the economics future bankers are taught at business schools. An article in the British newspaper *The Guardian* argues:

> It is business schools, after all, which flooded the banking world with graduates of their prestigious MBA courses. They then helped the economy to nosedive.
>
> One US website recently dubbed business schools the "academies of the apocalypse" and named and shamed dozens of international high-flying MBAs—"toxic bankers and scammers"—from Harvard MBA graduate Henry Paulson, secretary of the treasury under President Bush, who spoke

> vehemently against government regulation of Wall Street, to deposed HBOS [Halifax Bank of Scotland] chief executive Andy Hornsby.
>
> (James 2009)

And it is not only the failure of business schools to integrate courses on corporate social responsibility and business ethics that is to blame, as the *Guardian* article continues to suggest. Rather, the economic paradigm students are taught, in business schools and universities, can be used—or abused—to justify Gordon Gekko's "greed is good" maxim. One of the first things economics students learn is that there is a mechanism called the "invisible hand" by which markets magically transform the pursuit of self-interest into social benefit. Slightly more advanced students learn that Adam Smith's invisible-hand hypothesis has been confirmed mathematically by the so-called first fundamental theorem of welfare economics.

If the financial crisis is a social bad, as most of us would agree, and if it was brought about by freely operating markets, there must be a mistake in the invisible-hand idea. Perhaps greed isn't so good after all. At any rate, there is some reason to doubt whether economics rests on a solid ethical foundation.

Once more, philosophers should be in a good position to advance debates concerning ethical foundations. Ethics is one of the major branches of philosophy, and philosophers have been debating problems of ethics since the very beginning of Western philosophy. So surely they should have come up with some concepts, tools and ideas to aid economists?

The Philosophy of Economics: Interpreting Theory, Methodology and Ethics

Philosophers of economics are philosophers whose work focuses on the theoretical, methodological and ethical foundations of economics. It is no accident therefore that I selected criticisms of economics which point to theoretical, methodological and ethical challenges the financial crisis raises. In this section I will explain in slightly more detail what these three branches of philosophy of economics comprise.

The main theoretical framework in economics is given by theories of rational choice. These come in various guises, applicable to situations of certainty, risk, uncertainty, strategic situations and group decisions. To examine the foundations of rational-choice theory means to examine the axioms and principles that underlie such theories, to assess whether they are justifiable as axioms and principles of *rational* choice and adequate as descriptive accounts of *actual* choice.

Not every economic theory is one of rational choice, however. There are also macro theories such as the Solow–Swan growth model, the quantity theory of money, the IS/LM model and numerous microeconomic laws such as Say's law, Hotelling's law, the laws of supply and demand. To examine the foundations of economic theory also means to examine how to interpret

economic models and laws and ask about the metaphysical underpinnings of economics: Are there laws of nature in economics? What role do causal relations play? Are there only individuals or also social wholes?

Part I of this book will look in detail at a selection of these issues. Chapter 3 will examine rational-choice theory under certainty and risk and Chapter 4 game theory. Choice theory under uncertainty will be discussed briefly in Chapter 14. The notions of law, causation and causal mechanism will be examined in Chapters 5 and 6. Chapter 7 studies models and idealization in economics.

The second branch of philosophy of economics is methodology, which, as the name suggests, looks at the methods economists employ in testing their theories and establishing facts, laws and causal relations. Economists usually refer to something very specific when they use the term method: "A Three-Stage Optimization Method to Construct Time Series International Input–Output Database"; "A Quadratic Approximation Method for Agency Models with Frequent Actions"; "An Instrumental Variable Estimation Method of Nonparametric Models." Philosophers talk more abstractly and are more concerned about foundational issues. For them, there are observational and experimental methods. Observational methods generate data passively. Relevant for economics are mainly the measurement of economic indicators such as GDP, unemployment and inflation and the statistical analysis of data using regression. Experimental methods give economists a more active role: they design the set-up, manipulate a variable, observe and record the result and only then analyze it statistically. Experimental economics is now a thriving field within economics, and the use of randomized experiments has become popular in development economics in recent years.

To examine the methodological foundations of economics means to learn how these methods work, under which conditions they work, and what kinds of questions they can answer. Part II of this book is devoted to these issues. Chapter 8 will do so for the measurement of economic indicators, Chapter 9 for econometric methods, Chapter 10 for economic experiments and Chapter 11 for randomized field studies.

The third branch of philosophy of economics comprises the *ethical aspects of economics*. Economics involves ethical issues to no small degree. Consider a famous passage from Milton Friedman's essay "The Methodology of Positive Economics":

> An obvious and not unimportant example is minimum-wage legislation. Underneath the welter of arguments offered for and against such legislation there is an underlying consensus on the objective of achieving a "living wage" for all, to use the ambiguous phrase so common in such discussions. The difference of opinion is largely grounded on an implicit or explicit difference in predictions about the efficacy of this particular means in furthering the agreed-on end.
>
> (M. Friedman 1953: 5)

Friedman here claims that a "living wage for all" is an agreed-upon end while the means to achieve this end are disputed among economists. I doubt that at Friedman's time there was indeed widespread consensus regarding this policy objective but let us assume he was right. Does consensus regarding a policy objective mean that to take means to achieve it is morally justified? Is the end itself justified? Is rational discussion possible or are differences in value judgments "differences about which men can ultimately only fight" (M. Friedman 1953: 5)? These are the kinds of questions a philosopher of economics asks when he is concerned with the ethical aspects of economics.

Welfare economics is the branch of economics that addresses normative questions such as these. To examine the ethical foundations of economics mainly means to examine the ethical foundations of welfare economics, and that means to examine welfare, principles of distributive justice and ethical issues the market raises. Part III of this book will look at these matters, in this order, in Chapters 12–14.

I said earlier that to examine the ethical foundations of economics *mainly* means to examine the ethical foundations of welfare economics, because in fact ethical judgments abound even in what is called "positive" economics (that is, descriptive or explanatory economics). This is because many of the methods economists employ require ethical judgments in order to function well. This is especially noteworthy in the case of consumer price inflation measurement: to measure consumer price inflation appropriately judgments about the value of changes in the quality of goods have to be made; but such judgments require a notion of consumer welfare and are therefore ethical in nature (see Chapter 8).

Chapter 15 is a concluding chapter that aims to bring together various strands of thought encountered in the book. It discusses libertarian paternalism, a recent highly acclaimed and controversial policy proposal. As we will see, to understand this proposal, knowledge about rational-choice theory, experimental methodology, theories of well-being, justice and market failure is crucial. My hope is that readers will feel more confident about assessing such a proposal after reading the book than before.

The Aims of Economics and the Problem of Purpose

Economists pursue a variety of aims using a variety of scientific tools and methods. A classical statement due to Carl Menger ascribes the tripartite goal of explanation–prediction–control to economics (Menger 1963). According to this view, economists aim to explain past events and regularities; they further aim to anticipate future events and thereby help with policy and planning; and lastly they aim to provide recipes for successful interventions in the economy.

Menger's account is a nice starting point for a discussion of the aims of economics but it is incomplete. The description of economic phenomena

using indicators and statistics is an auxiliary aim for the more ultimate goal of explanation–prediction–control but it is also a highly important aim in itself (cf. Sen 1983a). That the inflation and growth rates in a country *X* are such-and-such, and that *Y* percent of its inhabitants live in poverty is important information, quite independently of whether that information is used for further scientific or practical purposes.

Economists also contribute to a number of normative discussions. To provide adequate normative accounts of rationality, well-being and, perhaps, justice must therefore be regarded as among the aims of economics. Not every working economist actually participates in these foundational debates but the times where these debates were left exclusively to philosophers are long gone.

The reason to engage in this rudimentary discussion of the aims of economics here is that many of the discussions that follow are meaningless unless conducted in the context of a relatively well-specified scientific purpose, and the aims introduced here can provide such a purpose. For example, there is no sense in which it is advisable or not advisable *simpliciter* for economists to investigate causal mechanisms. But it may be advisable for economists to investigate mechanisms given they aim to explain economic phenomena, because according to a widely held view descriptions of mechanisms provide explanations of outcomes of interest. Similarly, it is at least ambiguous to ask whether rational-choice theory is adequate as such. But it may well be adequate as a descriptive account of how people actually make choices or as a normative account of how people ought to make choices, or both, or neither.

Scientific practices, then, should be evaluated against a purpose, preferably one that is pursued by the scientists themselves and not imposed from outside by a philosopher. If that is so, a complication arises: most scientific practices are used for a variety of purposes and their adequacy or appropriateness is relative to that purpose. Causal mechanisms, for instance, are investigated for at least three purposes. First, as already mentioned, they are used to explain economic phenomena. The second and third uses relate to causal inference. It is often difficult to ascertain whether one economic variable (say, money) causes another (say, nominal income) by means of statistical analyses alone, essentially because all statistical models are underdetermined by data (see Chapter 9). When underdetermination is a serious problem, it is sometimes recommended to investigate possible mechanisms through which *X* might cause *Y*:

> However consistent may be the relation between monetary change and economic change, and however strong the evidence for the autonomy of the monetary changes, we shall not be persuaded that the monetary changes are the source of the economic changes unless we can specify in some detail the *mechanism* that connects the one with the other.
>
> (Friedman and Schwarz 1963: 59; emphasis added)

The third use is to ascertain the generalizability of already established causal claims. What is true of the economic system of the United States in the period 1867 to 1960 might not hold true thereafter or for other countries. Knowledge of mechanisms has been recommended as useful for deciding whether causal claims can be exported from one setting to another.

A methodological recommendation to investigate the causal mechanism(s) responsible for some economic phenomenon of interest may therefore be a good recommendation in one context but a bad one in another. This, however, creates a problem: each of the three branches of philosophy of economics is in fact itself multidimensional because each theory, method and ethical principle should be evaluated relative to a purpose, and the purposes of economics are multifarious.

I have simplified my life to some extent by focusing on one salient purpose in Part I: the explanation of economic phenomena. I do not think that explanation is the only aim of economics, or that it is particularly important. But focusing on explanation provides a nice narrative and organizing principle for Part I. I therefore begin with an introductory discussion of the topic of scientific explanation in Chapter 2. Other purposes will be salient in Part II (especially description in Chapter 8 and policy in Chapter 11) and Part III (for instance, moral reflection and policy in Chapter 12).

Part III on ethics is similarly simplified. Welfare, markets and justice could be discussed from all sorts of perspectives. I have decided to take the "greed is good" maxim—or rather, its academic counterpart, namely, the invisible-hand hypothesis—as organizing principle. As I will explain in Chapter 12, markets transform the individual pursuit of self-interest into social benefit only under a number of controversial assumptions. These assumptions include assumptions about human welfare, the nature of markets and what matters to society. The ethical topics discussed in Part III are selected with a view to assessing the plausibility and moral justifiability of these assumptions behind the invisible-hand hypothesis.

Study Questions

1 Try to think of one case each for the theoretical, methodological and ethical challenge discussed in the first section of this chapter. How can philosophy of economics help to answer it?
2 If you are an economics student, think about your curriculum. Which of your courses raise philosophical issues? Do you think they will be addressed in this book?
3 The last section of this chapter, dealing with "The Aims of Economics and the Problem of Purpose," mentions description, prediction, explanation, control and normative reflection as aims of economics. Are there other aims? Is there a hierarchy among the different aims?
4 To what extent are the problems in philosophy of science similar to those in philosophy of other sciences? To what extent are they different?
5 "A good economist has to be a good philosopher." Discuss.

Suggested Readings

This is the introductory chapter to an introductory text on philosophy of economics, so further readings should be other textbooks on philosophy of economics. Alas, there are none, at least not on the field as it is understood here. The closest one can get in aims and scope is probably Dan Hausman's *The Inexact and Separate Science of Economics* (Hausman 1992a), though this book excludes the ethical aspects of economics (but see Hausman and McPherson 2006). Its appendix contains an excellent introduction to the philosophy of science, and many of its discussions of, for instance, *ceteris paribus* laws, Mill's philosophy of economics, Milton Friedman's instrumentalism, Paul Samuelson's operationalism and the preference reversal phenomenon are still among the best one can find on these topics. There are a number of fairly introductory books on the narrower topic of economic methodology, most of which are older and have a historical focus. An exception is the recent Boumans and Davis 2010; other, earlier texts include Blaug 1992 and Caldwell 1982. Hands 2001 is accessible and comprehensive. There are also a number of handbooks in philosophy of economics such as Davis and Hands 2011 or Kincaid and Ross 2009. Most articles in the latter are closer to research papers than introductions or overviews, however. Roger Backhouse's entry in the *New Palgrave* is also a useful starting point (Backhouse 2008). For an anthology with classic and contemporary readings, see Hausman 2008.

1 Introduction

1.1 Experience and Reality

Does a tree falling in the forest make a sound when no one is around to hear it? The question is familiar to every undergraduate. One natural response is that *of course* the tree makes a sound – why shouldn't it? The tree makes a sound whether anyone is on hand to hear it or not. And, in any case, even if there are no people about, there are squirrels, birds, or at the very least bugs that would hear it crashing down.

Consider a more measured response, versions of which have percolated down through successive generations of student philosophers. The tree's falling creates sound waves that radiate outwards as do ripples on the surface of a pond, but in a spherical pattern. If these sound waves are intercepted by a human ear (or maybe – although this might be slightly more controversial – the ear of some nonhuman sentient creature) they are heard as a crashing noise. If the sound waves go undetected, they eventually peter out. Whether an unobserved falling tree makes a sound, then, depends on what you *mean* by sound. If you mean 'heard noise', then (squirrels and birds aside) the tree falls silently. If, in contrast, you mean something like 'distinctive spherical pattern of impact waves in the air', then, yes, the tree's falling does make a sound.

Most people who answer the question this way consider the issue settled. The puzzle is solved simply by getting clear on what you *mean* when you talk about sounds. Indeed, you could appreciate the original question as posing a puzzle only if you were already prepared to distinguish two senses of 'sound'. But what precisely are these two senses? On the one hand, there is the *physical* sound, a spherical pattern of impact waves open to public inspection and measurement – at any rate, open to public inspection given the right instruments. On the other hand, there is the *experienced* sound. The experienced sound depends on the presence of an *observer*. It is not, or not obviously, a public occurrence: although a sound can be experienced by many people, each observer's experience is 'private'. You can observe and measure agents' responses to experienced sounds, but you cannot measure the experiences themselves. This way of thinking about sounds applies quite generally. It

applies, for instance, to the looks of objects, to their tastes, their smells, and to ways they feel to the touch. Physicist Erwin Schrödinger (1887–1961) puts it this way in discussing sensations of color:

> The sensation of colour cannot be accounted for by the physicist's objective picture of light-waves. Could the physiologist account for it, if he had fuller knowledge than he has of the processes in the retina and the nervous processes set up by them in the optical nerve bundles and in the brain? I do not think so.
>
> (Schrödinger 1958, 90)

The picture of the universe and our place in it that lies behind such reflections has the effect of bifurcating reality. You have, on the one hand, the 'outer' material world, the world of trees, forests, sound waves, and light radiation. On the other hand, you have the 'inner' mental world, the mind and its contents. The mental world includes conscious experiences: the looks of seen objects, ways objects feel, heard sounds, tasted tastes, smelled smells. The 'external' material world comprises the objects themselves, and their properties. These properties include such things as objects' masses and spatial characteristics (their shapes, sizes, surface textures, and, if you consider objects over time, motions and changes in their spatial characteristics).

Following a long tradition, you might call those observed qualities properly belonging to material objects 'primary qualities'. The rest, the 'secondary qualities', are characteristics of objects (presumably nothing more than arrangements of objects' primary qualities) that elicit certain familiar kinds of experience in conscious observers. Experience reliably mirrors the primary qualities of objects. Secondary qualities, in contrast, call for a distinction between the way objects are *experienced*, and the way they *are*. This distinction shows itself in student reflections on trees falling in deserted forests. More fundamentally, the distinction encourages us to view conscious experiences as occurring 'outside' the material universe.

You might doubt this, confident that conscious experiences occur in brains, and regarding brains as respectable material objects. But now apply the distinction between primary and secondary qualities to brains. Brains – yours included – have assorted primary qualities. Your brain has a definite size, shape, mass, and spatial location; it is made up of particles, each with a definite size, shape, mass, and spatial location, and each of which contributes in a small way to the brain's overall material character. In virtue of this overall character, your brain would look (and presumably sound, smell, feel, and taste!) a particular way. This is just to say that your brain could be variously experienced. The qualities of these experiences, although undoubtedly related in some systematic way to the material reality that elicits them, differ from qualities possessed by any material object, including your brain. But if that is so, where do we situate the qualities of experience?

Your first instinct was to locate them in the brain. But inspection of brains reveals only familiar material qualities. An examination of a brain – even with the kinds of sophisticated instrumentation found in the laboratory of the neurophysiologist and the neural anatomist – reveals no looks, feels, heard sounds. Imagine that you are attending a performance of *Die Walküre* at Bayreuth. Your senses are assaulted by sounds, colors, smells, even tastes. A neuroscientist observing your brain while all this is occurring would observe a panoply of neurological activities. But you can rest assured that the neuroscientist will not observe anything resembling the qualities of your conscious experience.

The idea that these qualities reside in your brain, so natural at first, appears, on further reflection, unpromising. But now, if qualities of your experiences are not found in your brain, *where are they*? The traditional answer, and the answer that we seem driven to accept, is that they are located in your *mind*. And this implies, quite straightforwardly, that your mind is somehow distinct from your brain. Indeed, it implies that the mind is not a material object at all, not an entity on all fours with tables, trees, stones – and brains! Minds appear to be *non*material entities: entities with properties not possessed by brains, or perhaps by *any* material object. Minds bear intimate relations to material objects, perhaps, and especially intimate relations to brains. Your conscious experiences of ordinary material objects (including your own body) appear to reach you 'through' your brain; and the effects of your conscious deliberations have on the universe (as when you decide to turn a page in this book and subsequently turn the page) require the brain as an intermediary. Nevertheless, the conclusion seems inescapable: the mind could not itself be a material object.

1.2 The Unavoidability of the Philosophy of Mind

You might find this conclusion unacceptable. If you do, I invite you to go back over the reasoning that led up to it and figure out where that reasoning went off the rails. In so doing you would be engaging in philosophical reflection on the mind: *philosophy of mind*. Your attention would be turned, not to the latest results in psychology or neuroscience, but to commonsense assumptions with which this chapter began and to a very natural line of argument leading from these assumptions to a particular conclusion. As you begin your reflections, you might suspect a trick. If you are right, your excursion into philosophy of mind will be brief. You need only locate the point at which the trick occurs.

I think it unlikely that you will discover any such trick. Instead you will be forced to do what philosophers since at least the time of Descartes (1596–1650) have been obliged to do. You will be forced to choose from among a variety of possibilities, each with its own distinctive advantages and liabilities. You might, for instance, simply accept the conclusion as Descartes did: minds and material objects are distinct kinds of entity, distinct 'substances'.

You might instead challenge one or more of the assumptions that led to that conclusion. If you elect this course, however, you should be aware that giving up or modifying an assumption can have unexpected and possibly unwelcome repercussions elsewhere. In any case, you will have your work cut out for you. The best minds in philosophy – and many of the best outside philosophy as well – have turned their attention to these issues, and there remains a notable lack of anything resembling a definitive, uncontested view of the mind's place in the universe.

Do not conclude from this that it would be a waste of time for you to delve into the philosophy of mind. On the contrary, you can enjoy the advantage of hindsight. You can learn from the successes and failures of others. Even if you cannot resolve every puzzle, you might at least come to learn something important about your picture of the universe and your place in it. If you are honest, you will be obliged to admit that this picture is gappy and unsatisfying in many respects. This, I submit, represents an important stage for each of us in coming to terms with ourselves and our standing in the order of things.

1.3 Science and Metaphysics

Some readers will be impatient with all this. Everyone knows that philosophers only *pose* problems and never solve them. Solutions to the important puzzles reside with the sciences. So it is to science that we should turn if we are ever to understand the mind and its place in a universe of quarks, leptons, and fields. Residual problems, problems not susceptible to scientific resolution, are at bottom phony *pseudo*-problems. Answers you give to them make no difference; any 'solution' you care to offer is as good as any other.

Although understandable, this kind of reaction is ill-considered. The success of science has depended on the enforcement of well-defined divisions of labor, coupled with a strategy of divide and conquer. Consider: there is no such thing as science; there are only individual sciences – physics, chemistry, meteorology, geology, biology, psychology, sociology. Each of these sciences (and of course there are others) carves off a strictly circumscribed domain. Staking out a domain requires delimiting permissible questions. No science sets out to answer *every* question, not even every 'empirical' question. In this way, every science passes the buck. The practice of buck-passing is benign because, in most cases, the buck is passed eventually to a science where it stops. Sometimes, however, the buck is passed out of the sciences altogether. Indeed, this is inevitable. The sciences do not speak with a single voice. Even if every science were fully successful within its domain of application, we should still be left with the question of how these domains are related, how pronouncements of the several sciences are to be calibrated against one another. And this question is, quite clearly, not a question answerable from within any particular science.

Enter metaphysics. One traditional function of metaphysics – or, more particularly, that branch of metaphysics called *ontology* – is to provide a completely general, overall conception of how things are. This includes not the pursuit of particular scientific ends, but an accommodation of the pronouncements of the several sciences. It includes, as well, an attempt to reconcile the sciences with ordinary experience. In one respect, every science takes ordinary experience for granted. A science is 'empirical' insofar as it appeals to observation in confirming experimental outcomes. But the intrinsic character of observation itself (and, by extension, the character of observers) is apparently left untouched by the sciences. The nature of observation – outwardly directed conscious experience – stands at the limits of science. It is just at this point that the puzzle with which this chapter began rears its head.

Scientific practice presupposes observers and observations. In the end, however, the sciences are silent about the intrinsic nature of both. The buck is passed. Our best hope for a unified picture – a picture that includes the universe as described by the sciences and includes, as well, observers and their observations – lies in pursuing serious ontology. The buck stops here. You can, of course, turn your back on the metaphysical issues. This, however, is easier said than done. Many of those who proclaim their independence from philosophical influences, in fact, embrace unacknowledged metaphysical assumptions. In considering the nature of the mind, the question is not whether you are going to engage in metaphysical thinking, but whether you are going to do so self-consciously.

1.4 Metaphysics and Cognitive Science

This book concerns the metaphysics – the ontology – of mind. It revolves around reflections on questions about mind that fall partly or wholly outside the purview of the sciences. I should warn you that this is not a particularly fashionable endeavor. Many philosophers regard metaphysics as sterile and dated. Many more have arrived at the belief that our best bet for understanding the mind and its place in the universe is to turn our backs on philosophy altogether. These philosophers promote the idea that the philosophy of mind is, or ought to be, one component of what has come to be called *cognitive science*. Cognitive science includes elements of psychology, neuroscience, computer science, linguistics, and anthropology. What has a philosopher to offer the scientists who work in these areas? That is a good question.

Perhaps philosophers can provide some kind of unifying influence, a general picture that accommodates finer-grained assessments issuing from the scientific contributors to cognitive science. This, it would seem, is simply to engage in a kind of attenuated metaphysics. The metaphysics is attenuated to the extent that it excludes traditional ontological concerns, and excludes as well consideration of the relation sciences such as physics or chemistry bear on our uncovering the nature of the mind.

Introduction

If I sound skeptical about attempts to assimilate the philosophy of mind to cognitive science, I am. This book is premised on the conviction that the philosophy of mind is continuous with metaphysics as traditionally conceived. The difficult questions that arise in the philosophy of mind – and some would say the difficult questions *tout court* – are at bottom metaphysical questions. Such questions are, to all appearances, both legitimate and unavoidable. More to the point, philosophers can make (and in fact *have* made) progress in addressing them. This does not mean that philosophers have at hand a catalogue of fully satisfactory answers that could be succinctly reviewed and assessed in an introduction to the philosophy of mind. It does mean that you can reasonably hope to find, in subsequent chapters, some help in sorting through and eliminating options.

Am I just conceding the point: philosophers agree only on questions, not on answers? Not at all. Progress in philosophy, like progress in any domain, can be measured in two ways. You can focus on some definite goal, and ask yourself whether you are approaching that goal. But you can also ask yourself how far you have come from your starting point. And, on this count, philosophy can be said to move forward. In any case, we have little choice. Philosophical questions about the mind will not go away. They occur, even in laboratory contexts, to working scientists. And as recent widely publicized controversies over the nature of consciousness attest, ignoring such questions is not an option.

A final word about the relation philosophy of mind as I have characterized it bears to scientific endeavors. Philosophy of mind, I contend, is applied metaphysics, but metaphysics, like philosophy generally, is itself continuous with science. In engaging in metaphysics, you do not compete with, but complement, the sciences. You could think of metaphysics as concerned with the fundamental categories of being. Sorting out these categories is not a matter of engaging in empirical research, but the categories themselves are shaped in part by such research, and the nature of entities falling under the categories is only discoverable empirically, only in the kind of systematic intercourse with the universe characteristic of the sciences.

Suppose you are attracted, as Descartes and many other philosophers have been, to a substance–property ontology: the universe comprises objects, the substances, that possess assorted properties and stand in assorted relations to one another. Part of the appeal of such an ontology is its meshing with the picture of the universe we obtain from the sciences. What the substances and properties ultimately are is a matter to be determined by empirical investigation. Regarding philosophy of mind as applied metaphysics, then, is not to embrace the notion that philosophy of mind is a purely speculative, wholly a priori endeavor, an endeavor founded on reason alone, a matter of armchair reflection. Our understanding of the fundamental categories unfolds through everyday and scientific engagement with the universe. If science is the systematic investigation of the universe, then metaphysics is an ineliminable accompaniment of science.

1.5 A Look Ahead

The chapters to follow introduce a range of themes preeminent in the philosophy of mind. They do so in a way that presupposes no special background in the subject. The focus is on theories that have formed the basis of what might be regarded as the modern (or is it postmodern?) conception of mind. I have done my best to present each of these theories in a way that makes its attractions salient. Philosophers of mind have, in my judgment, been too quick to dismiss views they regard as quaint or outmoded. One result is that we are apt to forgo opportunities to learn from predecessors who, as it happens, had a good deal to teach. A second result of slighting unfashionable theories is that we risk repeating mistakes that we ought by now to have learned to avoid. I have tried to rectify this situation by providing sympathetic readings of positions that are sometimes caricatured, dismissed out of hand, or simply ignored. In so doing, I have put less weight on criticism of positions covered than do other authors. My job, as I see it, is to illuminate the territory. I leave it to you, the reader, to decide for yourself what to accept and what to reject.

This is not to say that I am neutral on topics discussed. Where I offer my opinion, however, I have tried to make clear that it *is* my opinion, a consideration to be weighed alongside other considerations. In a pair of concluding chapters I say what I think. There, I offer an account of minds and their place in the universe grounded in what I consider to be an independently plausible ontology. Chapter 12 is devoted to sketching that ontology; Chapter 13 spells out its implications for central issues in the philosophy of mind. The aim of these concluding chapters is less to convince you of the details of the view I prefer than to convince you of the importance of serious ontology for the philosophy of mind.

But this is to get ahead of the story. Chapters 12 and 13 follow on the heels of chapters devoted to the examination of a rich variety of conceptions of mind. Before venturing further, it might be worthwhile to provide a brief accounting of what you can expect in each of these intervening chapters.

Cartesian Dualism and Variations

Chapter 2 introduces Descartes's 'dualist' conception of mind. Descartes divides the world into mental and nonmental – immaterial and material – substances. Having done so, he is obliged to confront the notorious mind–body problem: how could mental and nonmental substances interact causally? Dissatisfaction with Descartes's apparent failure to answer this question bred amended versions of the Cartesian framework taken up in Chapter 3: parallelism, occasionalism, epiphenomenalism, idealism.

Parallelism, conceding the seeming impossibility of comprehending causal interaction between nonmaterial mental and material entities, supposes that mental and material substances do not interact, but undergo alterations in

parallel. *Occasionalists* introduce God as a connecting link between the mental and the material. God wills changes in both the material world and in minds in such a way that occurrences in each realm are aligned just as they would be were they directly causally related. *Epiphenomenalists* defend one way, material-to-mental causation. Mental occurrences are causally inert 'by-products' of material events (most likely events in the brain). Epiphenomenalism has enjoyed renewed popularity in recent years, especially among philosophers and neuroscientists who take consciousness seriously but see no prospect of 'reducing' conscious experiences to goings-on in the brain. *Idealists* reject the materialist component of the dualist picture. All that exists, they contend, are minds and their contents. Idealists do not simply deny that external, material objects exist; they contend that an external material world is literally *unthinkable*. The thesis that material objects exist outside the mind is judged, not false, but *unintelligible*.

Behaviorism

Idealists reject the materialist side of the dualist conception of mind: nonmental material substance is inconceivable. Materialists hold, in contrast, that every substance is a material substance (and nothing more). Chapter 4 focuses on one historically influential materialist response to Cartesianism, behaviorism.

Behaviorists hoped to show that the Cartesian conception of minds as distinct from bodies was based on a fundamental misunderstanding of what you are up to when you ascribe states of mind to yourself and to others. According to behaviorists, claims about your mind can be 'analyzed' into claims about what you do or are disposed to do, how you behave or are disposed to behave. To say that you are in pain – suffering a headache, for instance – is just to say (if the behaviorist is right) that you are holding your head, moaning, saying 'I have a headache', and the like, or at least that you are disposed to do these things. Your being in pain, then, is not a matter of your being a nonmaterial mind that is undergoing pain; it is simply a matter of your behaving in a characteristic way or being so disposed.

The Mind–Brain Identity Theory

Proponents of the identity theory, the topic of Chapter 5, side with behaviorists against the Cartesian notion that minds are immaterial substances, but stand with Cartesians against the behaviorists' contention that having a mind is nothing more than behaving, or being disposed to behave, in particular ways. Identity theorists argue that states of mind (such as having a headache, or thinking of Alice Springs) are genuine internal states of agents possessing them. These states, as neuroscience will someday reveal, are states of our *brains*. Mental states are *identical with* these brain states: mental states *are* states of brains. The identity theory appeals to anyone attracted to the

idea that minds are after all just brains. But, at the same time, the identity theory inherits problems associated with that doctrine, most especially the problem of finding a place of consciousness and qualities of conscious experiences in the brain.

Functionalism

Chapter 6 turns to functionalism, the historical successor to behaviorism and the identity theory, and certainly the present day's most widely accepted conception of mind. Functionalists identify states of mind, not with states of brains, but with *functional roles*. Your having a headache is for you to be in some state (doubtless a state of your brain) that exhibits input–output conditions characteristic of pain. (In this, functionalism resembles a dressed-up version of behaviorism.) Headaches are caused by blows to the head, excessive alcohol intake, lack of sleep, eyestrain, and the like, and they produce characteristic responses that include, but are not exhausted by, overt behavior of the sort focused on by behaviorists: head-holding, moaning, utterances of 'I have a headache'. In addition to behavior, a headache gives rise to other states of mind. (And here functionalists depart from the behaviorist contention that claims about states of mind are fully analyzable in terms of behavior and behavioral dispositions.) Your headache likely leads you to believe that you have a headache, for instance, to wish matters were otherwise, and to want aspirin.

Central to all forms of functionalism is the idea that states of mind are 'multiply realizable'. To be in a particular mental state is to be in a state that has a certain characteristic role. But many different kinds of material state could occupy or 'realize' the very same role. You, an octopus, and an Alpha Centaurian could all be in pain despite your very different physiologies (pretend that Alpha Centaurians have a silicon-based 'biology'). If being in pain were, as identity theorists suggest, solely being in a particular kind of neurological state, then octopodes and Alpha Centaurians, lacking physiologies comparable to ours, could not be in pain – an apparent absurdity. Functionalism affords a powerful model that allows for the 'abstraction' of states of mind from the hardware that 'realizes' them. One result is that dramatically different kinds of material system could all share a common *psychology*.

The Representational Theory of Mind

The Representational Theory of Mind, an important strain of mainstream functionalism, is the subject of Chapter 7. Proponents of the Representational Theory of Mind, regard minds as 'information-processing' devices. Information, in the form of 'mental representations' encoded in a Language of Thought, mediates incoming stimuli and behavioral outputs. On a view of this kind, minds could be thought of as 'software' running, not

on factory-assembled computing machines, but on neurological 'hardware' in brains. The appeal of such a picture is obvious: it promises to demystify minds and their operations, neatly integrating them into the material universe.

The Representational Theory of Mind inherits a difficulty widely associated with functionalist theories in general: the difficulty of accommodating qualities of conscious experiences. When you are in pain you are in a state of a kind that has various characteristic causes and effects. But what is salient to anyone undergoing a painful experience, being in this state is *painful*. Painfulness is qualitatively distinctive. Indeed you might think that what makes a state a *pain* state is its having this character. The difficulty is to see how the qualitative aspect of conscious experiences might be reconciled with the functionalist picture.

The Intentional Stance

Daniel Dennett, the hero (or villain!) of Chapter 8, focuses on the 'propositional attitudes': beliefs, desires, intentions, and the like. Dennett holds that the question whether a creature (or indeed anything at all) possesses a belief, or desire, or intention, turns solely on the utility of the practice of ascribing beliefs (or desires, or intentions) to it. We find it useful to describe cats, desktop computers, and even thermostats as believing this or that. Your cat *believes there is a mouse under the refrigerator*. Your desktop computer *believes the printer is out of paper* (and so alerts you to that fact); the thermostat *believes that the room is too cool* (and, in consequence, turns the furnace on).

To the extent that such attributions of belief *work*, cats, desktop computers and thermostats (and, of course, people and many other creatures) are 'true believers'. There is no further question of whether thermostats, for instance, *really* have beliefs, or whether it is just that we can get away with treating them *as though* they do. *All there is* to having a belief is to be so treatable.

The practice of ascribing beliefs, desires, and intentions is, according to Dennett, a matter of taking up a particular *stance*: the 'intentional stance'. In pursuing science, however, you would find surprising differences in creatures' responses to one another and to their environments. An understanding of these requires that you adopt the 'design stance'. In so doing, you discover that mechanisms responsible for behavior differ importantly across species. Actions indistinguishable from the intentional perspective look very different once you consider the 'design' of creatures performing them. Eventually, the design stance gives way to the 'physical stance'. This is the move from considering a creature's software to looking at its hardware.

Having a mind, then, is simply a matter of being describable from the intentional stance. The mystery of how minds are related to bodies vanishes, according to Dennett, once you recognize that truths expressible from within the intentional stance can be explained by reverting to the design stance. For

their part, design stance truths are grounded in facts uncovered from within the physical stance.

Reduction and Elimination

The thought that all there is to having a mind is being so describable, could easily lead to the more radical thought that minds are, at bottom, fictions. In Chapter 9 this possibility is explored in some detail. Perhaps our talk of minds and their contents and our practice of explaining behavior by reference to mental goings-on, are simply remnants of primitive animistic forms of explanation. We once explained the weather by the fickleness of the gods who controlled it. Later, we developed a science, meteorology, that enabled us to understand meteorological phenomena purely 'naturalistically' without appeal to conscious agents. Maybe explanations of intelligent behavior should likewise move beyond appeals to states of mind and mental processes.

One possibility is that talk of minds could be replaced by talk of states and processes unearthed by neuroscience. A second possibility takes seriously an important feature of the Representational Theory of Mind. Suppose the mind *were* animated by 'mental representations'. These would be sentences in a built-in, hard-wired 'Language of Thought'. But just as a computing machine cares nothing for the significance of symbols it processes, so minds – or their physical 'realizers', brains – care nothing for the meanings of symbols in the Language of Thought: mental processes are purely 'syntactic'. Representational 'content', central in traditional accounts of the mind, drops out of the picture, taking with it the familiar categories of belief, desire, and intention.

Consciousness

Consciousness is the 800-pound gorilla that inevitably asserts itself in the philosophy of mind. Psychology and neuroscience have made impressive advances in recent years. None of these advances, however, has brought us a step closer to understanding the 'mystery of consciousness'. Or so it seems.

What exactly *is* the 'mystery of consciousness'? It is not easy to say. You can get a feel for it, however, by reflecting on a vivid conscious experience, the sort of experience you might have in strolling on a tropical beach at sunset, for instance. You have visual experiences of the ocean, the sky, the setting sun, the sand; you feel a cool breeze and the warm sand under your feet; you hear the waves lapping the shoreline and the calls of birds; you smell the scent of flowers and salt air. These experiences are the result of your perceptual encounter with your surroundings. Your 'sensory surfaces' are stimulated, and signals are passed on to your brain where, it would seem, they issue in your experiences. If this sequence were blocked at any point or inhibited, your experiences would be diminished.

As a result of your experiences, you respond in various ways. You are led to entertain new thoughts and to continue strolling; you turn your head to find a bird that has produced a particularly striking call. A scientist studying all this could, at least in theory, follow the whole input–output sequence, or at any rate have a detailed picture of what is going on inside your body. The problem is that there is apparently a 'gap' between what a scientist could observe, and what your experiences are like. How are experiences and their 'Technicolor' qualities to fit into the scientific picture? This is the mystery of consciousness.

Faced with this mystery, scientists and philosophers have responded in various ways. Some have chosen simply to ignore the phenomenon, dismiss it as unfit for scientific study. Although a scientist is free to 'bracket' or ignore one topic for the sake of studying others, philosophers do not have this luxury. Philosophers are bound to attempt a unified picture of the universe, a picture that accommodates both the findings of psychology and neuroscience, *and* conscious experiences.

Another option is to accept conscious experiences as they are, but to assimilate them to functional states of agents. To be conscious is just to be in a particular sort of functional state, a state realized in your brain. The question is whether qualities of conscious experience can plausibly be dealt with in this fashion. Many have doubted it.

Other options include *epiphenomenalism* (conscious qualities are causally inert by-products of material processes in the brain), *panpsychism* (what you might regard as qualities of conscious experience are really part of the 'intrinsic nature' of matter), and *representationalism* (what you regard as qualities of experiences are in fact qualities you represent objects you experience as having; qualities of your beach experience are, at bottom, just qualities of objects that you are experiencing).

Clearly, then, if you take *qualia* (the term used by philosophers to designate qualities of conscious experiences) seriously, you will need to say something about what David Chalmers calls 'the hard problem': what relation does consciousness bear to material goings-on? You might, in the end, be driven to embrace a position that seemed antecedently unattractive, epiphenomenalism, for instance, or panpsychism. Before accepting a position concerning which you might have important reservations, however, you should be certain that you have exhausted the space of possibilities. A central goal of this book is to make you aware of the extent of that space and thereby to equip you to choose wisely.

Non-Reductive Physicalism

Cartesian dualism takes consciousness seriously, while at the same time making it clear why conscious experiences are not going to be encountered in scientific investigations of the material universe. That is the good news. The bad news is that Cartesian dualism makes the interface between minds and

bodies – mind–body interaction – wholly mysterious. How could substances sharing no attributes causally interact?

Suppose Descartes is wrong, however. Suppose mental properties and material properties *could* be possessed by material substances. Neuroscience research suggests that mental properties are 'grounded in' physical properties. You experience a pain, for instance, *because* your nervous system is in a particular state. Still, pains evidently differ qualitatively from anything in your nervous system. This suggests that, although the presence of mental properties depends on the presence of physical properties, mental properties are nevertheless distinct from physical properties; the mental is not 'reducible' to the physical. The result: substance monism combined with a dualism of properties.

This neat 'non-reductive physicalist' solution to the mind–body problem has recently come under fire. Mental and material properties might be properties of a single substance, but if mental and material properties are genuinely distinct, and if mental properties depend for their very existence on material properties, it is hard to see how mental properties could have a role in the production of bodily behavior. This is the Cartesian problem all over again.

Suppose that your forming the belief that a snake is in the path results in a particular bodily response (your altering course). Suppose that some material event in your brain 'realizes' this belief, and that this material realizer causes you to alter your course. The material realizer might 'underlie' or 'give rise to' various mental properties. Suppose that it does. Those properties need have no part in producing your subsequent behavior; however, they might be 'causally irrelevant'.

A fast-bowled red cricket ball cracks a batsman's rib. The ball is red, but its redness apparently has no role in the cracking. Many have thought that there are excellent reasons to think mental properties are like this; the properties are on the scene, perfectly genuine, but 'causally irrelevant'. In that case you would be left with a virulent new form of epiphenomenalism. Once again, you will need to sort through the options and find the one you regard as the most promising, perhaps only because it is the least objectionable.

Ontology and Mind

The book concludes with two chapters in which, as noted above, I lay out an account of the mind grounded in a particular ontology. The ontology, details of which occupy Chapter 12, regards substances as the basic entities. Substances possess properties, which I take to be *ways* substances are. You could think of a cricket ball as a substance. A cricket ball is red and spherical. The ball's redness and sphericity are ways it – that ball, and nothing else – is. Every property contributes distinctively to its possessor's qualities and causal powers or dispositions. Indeed, every property is *both* qualitative and dispositional: properties are powerful qualities. From this basis, I construct,

in Chapter 13, an account of the mind. The construction is tentative and sketchy, but the fundamental ideas will be clear. I regard it as an important feature of the conception I sketch that it accommodates the attractions of its competitors without inheriting (all) their liabilities. There is, as I hope to convince you, something right as well as something wrong in each of the diverse accounts of the mind taken up in earlier chapters.

A final comment. This book will have achieved its purpose if it convinces you that any philosophical account of the nature of the mind includes an important metaphysical component. I am less concerned with your agreeing with me on the details of this component. To my way of thinking, you will have made considerable progress if only you recognize that the study of mind requires a stiff measure of ontological seriousness.

Suggested Reading

A book like this should inspire readers to look more closely at primary sources, the work of philosophers bent on defending (or attacking) positions being discussed. To this end, anthologies in the philosophy of mind can be especially useful. Three new collections and an old standard merit special mention. O'Connor and Robb's *Philosophy of Mind: Contemporary Readings* (2003) assembles essays expressly selected to complement this volume. Rosenthal's *The Nature of Mind* (1991), and its updated replacement, Chalmers's *Philosophy of Mind: Classical and Contemporary Readings* (2002), cover much of the same territory. My own collection, *Philosophy of Mind: A Guide and Anthology* (2003b) includes, in addition to primary source readings, extensive introductory material.

Block, Flanagan, and Güzeldere's *The Nature of Consciousness: Philosophical Debates* (1997) focuses on consciousness and includes a valuable comprehensive introduction by Güven Güzeldere. William Lycan's *Mind and Cognition: An Anthology* (1999), Christensen and Turner's *Folk Psychology and the Philosophy of Mind* (1993), and Geirsson and Losonsky's *Readings in Mind and Language* (1996) contain, in addition to readings in philosophy of mind, selections on topics in cognitive science that will be of interest to readers hankering for empirical enlightenment. Beakley and Ludlow's *The Philosophy of Mind: Classical Problems, Contemporary Issues* (1992) combines selections from towering historical figures with present day sources in both philosophy and psychology, arranged by topic. Godfrey Vesey's *Body and Mind: Readings in Philosophy* (1964), Daniel Kolak's *From Plato to Wittgenstein: The Historical Foundations of Mind* (1997), Peter Morton's *Historical Introduction to the Philosophy of Mind: Readings with Commentary* (1997), and Daniel Robinson's *The Mind* (1999) all incorporate interesting and important historical selections.

Samuel Guttenplan's (1994) *Companion* and Stich and Warfield's (2003) *Guide* to the philosophy of mind are organized topically and provide in depth coverage of particular subjects. Gregory's *Companion to the Mind* (1987) has

broader ambitions, and could prove useful on topics in psychology and the neurosciences.

Volumes intended, as this one is, to introduce readers to the philosophy of mind include: Tim Crane's *Elements of Mind: An Introduction to the Philosophy of Mind* (2001), George Graham's *Philosophy of Mind: An Introduction* (1993), Dale Jacquette's *Philosophy of Mind* (1994), Jaegwon Kim's *Philosophy of Mind* (2010), E. J. Lowe's *An Introduction to the Philosophy of Mind* (2000a), and William Lyons's *Matters of the Mind* (2001). D. M. Armstrong in *The Mind–Body Problem: An Opinionated Introduction* (1999), Anthony Kenny in *The Metaphysics of Mind* (1989), Colin McGinn in *The Character of Mind* (1982), and Georges Rey in *Philosophy of Mind: A Contentiously Classical Approach* (1997) advance distinctive views of the mind in the course of introducing the subject. Being opinionated goes with being a philosopher. A clear view of the territory results not from occupying a single, neutral vantage point, but from acquiring familiarity with a variety of perspectives.

Braddon-Mitchell and Jackson's *The Philosophy of Mind and Cognition* (1996) and Paul Churchland's *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*, Revised Edition (1988) incorporate useful discussions of topics in the philosophy of mind and in cognitive science. Readers whose interests tend toward the empirical will benefit from a look at Bechtel *et al*., *A Companion to Cognitive Science* (1998); Joao Brãnquinho, *The Foundations of Cognitive Science* (2001); Cummins and Cummins, *Minds, Brains, and Computers: The Foundations of Cognitive Science: An Anthology* (2000); Jay Garfield, *Foundations of Cognitive Science: The Essential Readings* (1990); Gleitman *et al*., *An Invitation to Cognitive Science* (1995); and Michael Posner, *Foundations of Cognitive Science* (1989). (As these titles suggest, cognitive science has a certain obsession with its foundations.) These anthologies march alongside introductory texts that include Andy Clark, *Being There: Putting Brain, Body, and World Together Again* (1997) and *Mindware: An Introduction to Cognitive Science* (2001); James Fetzer, *Philosophy and Cognitive Science* (1991); Owen Flanagan, *The Science of the Mind* (1984); Robert Harnish, *Minds, Brains, Computers: An Historical Introduction to the Foundations of Cognitive Science* (2001); Rom Harré, *Cognitive Science: A Philosophical Introduction* (2002); and Paul Thagard, *Mind: Introduction to Cognitive Science* (1996). The online *MIT Encyclopedia of Cognitive Sciences* (Wilson and Keil, 1999) is a useful and reliable Internet resource.

In general, you should be skeptical of materials you turn up on the Internet. Disinformation swamps information; self-proclaimed philosophers often aren't. For this reason, entries in Wikipedia should be approached with extreme caution. In contrast, the online *Stanford Encyclopedia of Philosophy* (Zalta 2002) is a trustworthy source for topics in the philosophy of mind. Marco Nani's (2001) *Field Guide for the Philosophy of Mind* and Chris Eliasmith's (2003) *Dictionary of Philosophy of Mind* contain useful entries. David Chalmers's *Contemporary Philosophy of Mind: An Annotated*

Introduction

Bibliography (Chalmers 2001) is an excellent bibliographic resource. Web sites of authors included in the bibliography can also contain useful and reliable material. Hint: to locate an author's web page, try typing 'Author Name philosophy' (the author's name, followed by a space, then 'philosophy') into your favorite search engine.

CHAPTER

1 A Map of the Terrain of Ethics

CASE 1

The Boy Who Ate the Pickle

A 9-year-old youngster named Yusef Camp who lived in inner-city Washington ate a pickle that he had bought from a street vendor. Soon after eating it he went into convulsions and collapsed on the sidewalk. A rescue squad took him to the nearest emergency room where his stomach was pumped. Tests revealed that the pickle contained traces of marijuana and PCP. The boy suffered severe respiratory depression and was left unconscious, unable to breathe for an unknown period.

The emergency room personnel restored respiration by putting him on a ventilator, but they were unable to restore him to consciousness or get him breathing adequately on his own.

The physicians concluded that his brain function was irreversibly destroyed and that there was no possibility of recovery. They might have simply pronounced him dead and then stopped the ventilator, but the situation soon became more complicated. Two of the attending neurologists were convinced that the patient's brain was totally dead, but one believed that he had minor brain function still in place. So they were incapable of pronouncing the patient dead based on loss of brain function. Now the question became, What should they do? Their patient was still living but permanently unconscious, breathing only because he was on a ventilator.

The physicians pointed out that there was nothing more they could do except keep the ventilator running, perhaps indefinitely and maintain the boy in a persistent or permanent vegetative state. (The longest case on record of maintaining a patient in what is called a permanent vegetative state is over thirty-seven years.)

The parents were Muslims, members of the Nation of Islam, who firmly believed in the power of Allah. They believed that Allah would intervene if it was his will, and that it was the physicians' job to give Allah that opportunity. How should the physicians respond?

The physicians, the parents, and everyone else involved in this case face some difficult and controversial ethical choices. They need to determine the proper definition of death, the role of parents and other surrogates in deciding about medical care for a minor, the proper ethics of terminal care, the morality of using scarce medical resources, and the role that minority religious perspectives ought to play in modern, secular medical care. In order to sort out these disparate and complex ethical issues we need a map of the ethical terrain: an overview of the kinds of ethical issues at stake and the terminology for labeling the disputes. This chapter will provide a basic map of that terrain. Once that overview is in place, we can begin sorting out the issues facing Yusef Camp's parents and physicians.

THE LEVELS OF MORAL DISCOURSE

The Level of the Case

Often in biomedical ethic, the discussion begins with a case problem. Someone faces a concrete moral dilemma or two people disagree about what in a specific situation is the morally appropriate behavior. Some people may mistakenly think that ethical choices do not occur all that often in medicine. They think that an "ethics case" is an unusual, special event. In fact, ethical and other value choices occur constantly, but, fortunately, in almost all situations the ethically correct course is obvious. The decision can be made with little or no conscious thought. Ethical choices have still been made—even if the decision maker does not even realize it. He or she can rely on well-ingrained moral beliefs and get by quite adequately. Occasionally, however, the choice does not come as easily. As in the case of Yusef Camp, the choice requires more careful, conscious thought. The physician faced with a choice may turn to colleagues or to a hospital ethics committee for advice. A lay person may turn to friends or to a trusted religious or secular group for guidance.

One kind of advice may come in the form of mentioning other cases that seem similar, cases that have been resolved in the past. They may be in the form of a Biblical story or a legal case about which the culture has reached agreement. These agreed-upon cases are sometimes referred to as "paradigm cases." Most people can agree that, in matters of ethics, similar cases should be treated similarly. In fact, one of the identifying characteristics of an ethical judgment (as opposed to a matter of mere taste or preference) is this awareness that if the relevant features are similar, then cases should be treated alike. As long as people can agree on what should be done in the paradigm case and can agree that the new case is similar in all relevant respects, they will be able to resolve their problem. This approach relying on paradigm cases is sometimes called **casuistry**. As seen in Figure 1, this is the lowest or most specific level of what can be

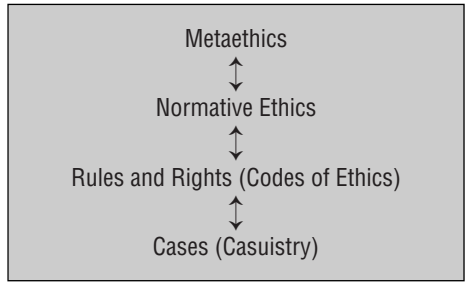


FIGURE 1 The Four Levels of Moral Discourse

considered the four major levels of moral discourse. This figure is a simplified version of the more elaborate map of the ethical terrain that appears on the front and back inside covers of this book.

Rules and Rights (Codes of Ethics)

But what if the basic ethics we learned as children does not settle the problem? What if we cannot agree on a paradigm case or cannot agree that our present problem is like the paradigm case in all relevant respects? We may, at that point, move to a second level of moral discourse, the level of **moral rules** and **rights**. Sometimes rules and rights tell us what is legal, but they may also describe what is ethical. Since not everything that is legal is also ethical (and not everything that is illegal is necessarily unethical), it will be important to note the difference. If a rule or a right is considered *ethical*, it will be seen as grounded in a moral system, an ultimate system of beliefs and norms about the rightness or wrongness of human conduct and character. Groups of rules or rights claims are sometimes called *codes of ethics*.

Yusef Camp's physicians may consult the Code of Ethics of the American Medical Association to see whether that group considers it ethical to stop treatment in such cases. His parents might consult an Islamic code. Some of the parties in the dispute may bring out the Hippocratic Oath or a "patients' bill of rights."

Sometimes the parties to an ethical dispute may cite a rule-like maxim. "Always get consent before surgery" or "a patient's medical information must be kept confidential" are examples of such maxims. These rule-like statements are usually quite specific. A large number of them would be needed to cover all medical ethical situations. If there is agreement on the rule that applies, then the case problem might be resolved at this second level.

Sometimes these maxims are stated not as rules but as *rights* claims. The statement, "a patient has a right to consent before surgery" would be an example. So would the statement "a patient has a right to have his or her medical information kept confidential." Rules are expressed from the perspective of the one who has a duty to act; rights claims from the vantage point of the one acted upon. Often rules and rights express the same moral duty from two different perspectives. "Always get consent before surgery" expresses from the health provider's point of view the same moral notion that is expressed from the patient's vantage point as "a patient has a right to consent before surgery." They are then said to be "reciprocal." If one person has a duty

to act in a certain way toward another, that other person usually can be said to have a right to be acted upon in that way.

Medical professional, religious, cultural, and political organizations sometimes gather together collections of rules or rights claims. When they do, they "codify" them or produce a *code of ethics*. They can also take the form of *oaths* as in the Hippocratic Oath or *directives* as in the "Ethical and Religious Directives for Catholic Health Facilities." When the statements are made up of rights claims, they are often called *bills of rights* as in the American Hospital Association's "Patient's Bill of Rights" or *declarations* as in the new "Universal Declaration on Bioethics and Human Rights." Chapter 2 looks at various oaths, codes, and declarations, and sees what their implications are for cases like Yusef Camp's. We will discover how controversial these codifications are. Proponents of such codes not only have to determine what rules and rights are appropriate, but also which humans (and non-humans) have the moral standing to have claims based on these rules and rights. Chapter 3 takes up this question of who has this moral standing. Here we address the question of whether Yusef Camp has the moral standing of a living human being or is already dead—according to a brain-oriented definition of death. We will also see the implications for the moral status of fetuses and non-human animals. We will at this point also confront the new controversy over the use of stem cells.

These rules and rights claims may provide enough moral guidance that the problem being disputed can be resolved. They rest, however, on the authority of the groups creating the codes (or on the inherent wisdom of the maxims themselves).

One of the controversies in ethics is how seriously these rules and rights must be taken. At one extreme, an ethical theory could include the view that there are no exceptions to the rules or rights. This view, which almost no one actually holds, is sometimes called **legalism**. At the other extreme, someone might hold that every case is so unique that no rules or rights can ever be relevant in deciding what one ought to do in a specific situation. This view, which is as implausible as legalism, is called **antinomianism**. Two intermediate positions are more plausible. **Situationalism** holds that moral rules are merely "guidelines" or "rules of thumb" that must be evaluated in each situation. The **rules of practice** view holds that rules specify practices that are morally obligatory. In this view the rules are stringently binding on conduct. Exceptions are made only in very extraordinary circumstances—much less easily than in the situationalist position. The continuum is represented in Figure 2 and in the more complete map of the ethical terrain inside the front and back covers.

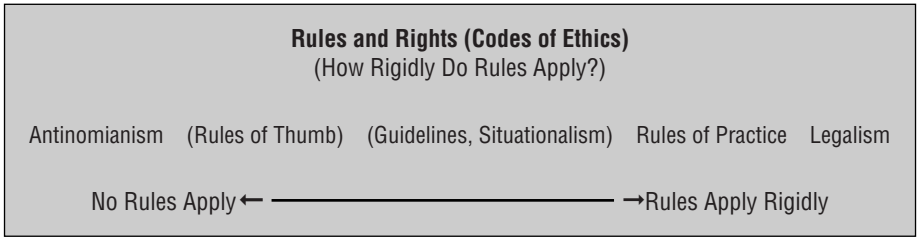


FIGURE 2 Rules and Rights

Normative Ethics

People in an ethics dispute may not be able to determine which rule or rights claim applies or how it should be applied. If the citing of various rules or rights claims cannot resolve the matter at controversy, a more complete ethical analysis may be called for. The parties may have to move to a third level of moral discourse, what can be called the level of **normative ethics**. It is at this level that the broad, basic norms of behavior and character are discussed. It is in these basic norms that rules and rights claims will be derived and defended. It is also at this level that the norms of good moral character are articulated. The key feature of these norms is that they are general: They apply to a wide range of conduct and character. If "always get consent before surgery" is a moral rule, it might be associated with some broader ethical norm, such as respect for autonomy. Since these norms (like autonomy) are very broad, only a few norms will be expected or needed in a "normative ethical theory."

ACTION THEORY As illustrated in Figure 3, normative ethics involves at least three kinds of questions. An ethical theory at the normative level, therefore, must address three separate issues. Much of recent biomedical ethics has dealt with the principles of morally right action. The focus is on the action itself; not on the character or motives of the actor. The central normative ethical question has been "what principles make actions morally right?" The answer involves some list of **moral principles** such as beneficence, nonmaleficence, respect for autonomy, or justice. These are proposals for characteristics of actions that make them morally right. Someone might claim, for example, that doing good (beneficence) or respecting autonomy will tend to make an action (or perhaps a set of actions) morally right. The principles of right action were almost the entire focus of bioethics in the 1970s and '80s and remain a dominant part of the discussion. They will be considered in more detail in Chapter 4, when we take up the principles that concentrate on producing the best possible consequences, and in Chapters 5–8, when we consider some additional principles that do not deal with maximizing good outcomes. The figures in these chapters (and the inside covers of this book) expand the map of the terrain of ethics by providing charts of possible consequence-maximizing principles and of ones that attempt to identify certain moral duties that are independent of producing good consequences.

If a bioethic includes more than one ethical principle, the **action theory** portion of normative ethics will have to address the question of how to resolve the conflicts that arise among them. There are several different possibilities for resolving these conflicts. They will be explored in Chapter 9.

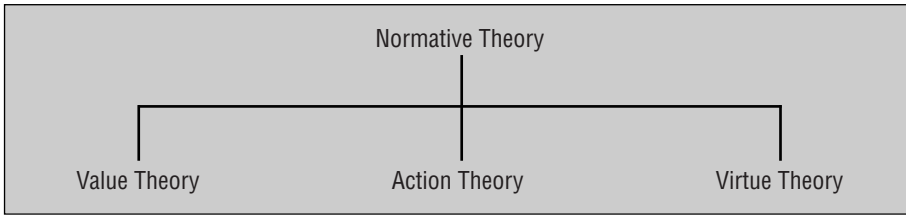


FIGURE 3 Three Questions of Normative Ethics

VALUE THEORY Since beneficence (or producing good consequences) is one possible principle of right action and nonmaleficence (or avoiding producing bad consequences) is another, a second question that has to be addressed in a full normative theory is, "What kinds of consequences are good or valuable?" This branch of normative theory is therefore called **value theory**. Some of the questions of value theory are taken up in Chapter 4, where the principles of beneficence and nonmaleficence are discussed. The map of the options for this part of normative theory is expanded in that chapter in Figure 9. Just as there are disputes about what the proper list of principles is, so there are disputes about what kinds of things are valuable. Some kinds of things, like money, seem to be valuable, but only instrumentally—because it will buy something intrinsically valuable. The real question here is, What kinds of things are intrinsically valuable? Among the standard answers are happiness, beauty, knowledge, and—importantly for biomedical ethics—health. Some people also consider morally good character to be among those things that are intrinsically valuable.

VIRTUE THEORY That brings us to the third question of normative ethics: "What kinds of character traits are morally praiseworthy?" A morally praiseworthy character trait—such as compassion or benevolence or faithfulness—is called a *virtue* and, hence, this part of normative ethics is referred to as **virtue theory**. For a fuller list of the virtues and a discussion of their role in bioethics, see Figure 1 in Chapter 11 and the discussion of virtue theory in that chapter.

The virtues refer not to the character of actions, but to the character of the people who engage in the actions. *Benevolence* and *beneficence* should be contrasted. *Benevolence* is a virtue, the virtue of willing to do good. *Beneficence* is a principle of actions, the principle of actually acting in such a way that good consequences result. One can of course will the good (show the virtue of benevolence) but end up not doing the good (being beneficent). One can also be malevolent, but nevertheless beneficent. (This person would not be of good will, but would nevertheless act in such a way that good results are produced, perhaps because the malevolent one has calculated that it is in his or her self-interest to produce the good consequences.)

This means that normative ethics involves questions of ethical principle (action theory), intrinsic goods (value theory), and good character (virtue theory). Depending on the question asked and the situation, one may be more interested in one of these questions than another. In the 1970s and '80s, for example, most biomedical ethics concentrated on the principles of right action. Theorists of the time wanted to get straight on whether an action by a physician was morally right if it was designed to produce good consequences, but simultaneously violated respect for autonomy or involved telling a lie. The bioethicists of that time did not really care very much about the character of the physician; the issue was what made his or her external conduct morally right, not whether the physician had a virtuous disposition. Ethicists who attacked the mainstream medical paternalism of the day in the name of the principle of autonomy were concerned that the benevolently paternalistic physician was *acting* immorally by violating the principle of autonomy even if his heart was in the right place. Only in the late 1980s did biomedical ethics return to the more traditional interest in the virtuous character of the health provider. Since then, there has been more of a balance between ethics concerned about actions and ethics concerned about the

character of the actors. Since action theory and virtue theory ask different questions, it is normally a mistake to think of the two as being in conflict. They are simply different aspects of the general considerations of morality.

Metaethics

Sometimes if people can get clear on which principles or virtues or intrinsic goods are at stake, they can then resolve lower-level moral disputes. They might agree on the principle of autonomy (or beneficence) being dominant, for example, and then be able to settle disputes about which moral rules or rights are legitimate. In the more interesting and complicated cases, however, the disagreement may remain intractable. The parties to a dispute may not be able to determine which principles should prevail. One person, for example, might give priority to the principle of beneficence while another might believe that autonomy should take precedence (even if respecting autonomy will lead to less good consequences, that is, being less beneficent). Or they may not agree on whether right action or virtuous character is more important. When disputes of this sort linger, the discourse must move to a fourth and final level, the level of **metaethics**.

Metaethics deals with the most basic questions of ethics: the meaning and justification of ethical terms, how people know which principles or virtues are the correct ones, and the ultimate grounding of ethics. Here we are no longer interested in the substantive questions of which actions are morally right or which traits of character are morally praiseworthy. Rather we are dealing with even more basic issues of where to look to get answers to these questions and how we can know when we have the right answer.

Religious ethics has, by now, fairly standard answers to these metaethical questions. To the religious person, claiming an action is right *means* it would be approved by the deity or is in accord with laws created by the deity. For them, to say that a character trait is virtuous is to say that it would be approved morally by God. Religious people also have well-worked-out notions of how humans can know something is ethical: by revelation and reason, by reading the scriptures, and by religious authorities such as the pope, church councils, Islamic *fatwas*, or Talmudic law. More mystical religious people may rely more on direct spiritual revelation.

Secular people are not satisfied with these positions, but have analogous answers of their own. The grounding of ethics may be in natural law or in some **contract** (actual or hypothetical) among people. Traditional secular ethics have shared with monotheistic religions the notion that ethics is **universal**; that is, for a specific moral case at a specific time and place, all people *ought* to reach the same ethical judgment about whether the behavior involved is morally right or wrong. Of course, universalists recognize that not all people actually will agree on such judgments, but they believe that there is some universal standard (such as the divine will or reason or natural law) against which people's judgments can be tested. If two people disagree, say, about whether a particular abortion in a particular set of circumstances is immoral, then at least one of them must be mistaken.

Other secular theories share with polytheistic religion the notion that there is more than one standard of reference for moral matters. These metaethical positions are called **relativist** because they hold that moral judgments are relative to the multiple standards or authorities that exist. For example, for believers in polytheistic religion,

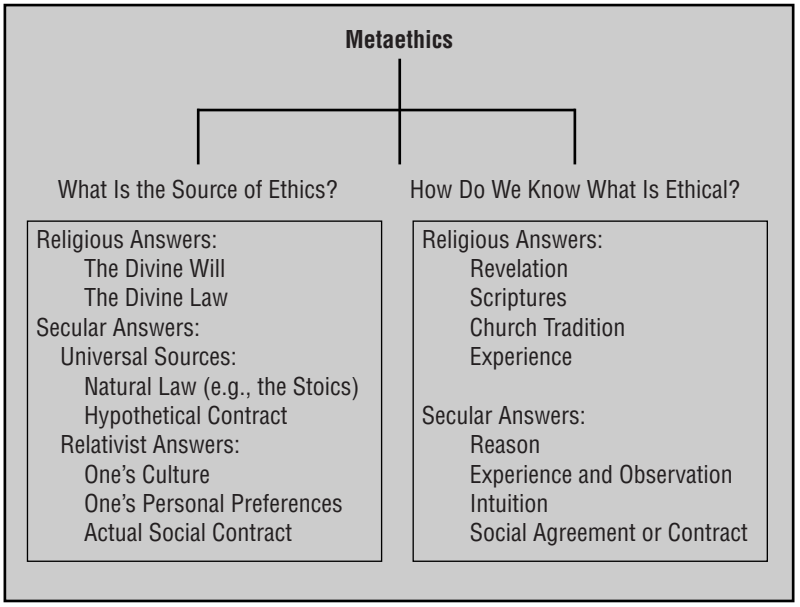


FIGURE 4 Metaethics: The Meaning and Justification of Ethical Judgments

different cultures may have different deities. One culture's god might approve of a merciful killing of a suffering patient while another culture's, while considering exactly the same case, might disapprove. Likewise, a secular ethic might be relativist if it holds that the ultimate standard of reference for moral judgments was the norms agreed upon in a particular culture. These alternative answers to the question of the source of ethical judgments are summarized in Figure 4 and in the chart on the inside covers of the book.

Metaethics also deals with a related question of how we can know the content of these moral norms. While for religious ethics the divine will or divine law is known through revelation or reason, scripture, or church tradition, in secular ethical systems it is known through reason or through empirical experience. The German philosopher Immanuel Kant based knowledge of ethics in reason; the British empiricist David Hume in the experience of sympathy. These religious and secular answers to the question of how we can know what is ethical are also included in Figure 4 and in the inside covers.

These metaethical questions take us well beyond what bioethics normally addresses. Fortunately, many have found that even if there is serious disagreement at this most abstract level, those who cannot agree on matters of religion and secular philosophy can nevertheless reach some converging consensus at the lower three levels of moral discourse. They can agree on normative ethical matters of **moral principles**, **virtue theory**, and intrinsic goods (**value theory**). They can agree on many **moral rules** and **rights**. Therefore, they can sometimes agree on what is morally right in a particular case, even if they have no agreement in metaethical matters. This is sometimes referred to as a *common morality*, an agreement on many, indeed most, ethical matters across cultures, religions, politics, and time periods.

A FULL THEORY OF BIOETHICS

Even though we need not spend much time in bioethics going all the way up the "ladder" of the levels of moral discourse, a full theory would need to climb all the way to this top level. In fact, some people would claim that traveling up this ladder from the case through rules and rights to normative theory and finally to matters of metaethics is traveling the wrong way. They hold that, in matters of ethics, one must start at the top and work one's way down. One would then first get clear on the meaning and justification of ethical claims—on metaethics—and then identify principles of right action, traits of good character, and intrinsic goods at the normative level, which would, in turn, lead to identifying lists of rules and rights, which would finally tell us how to act and what character traits one should have in particular cases. They claim one should reason from top to bottom rather than from case to the more abstract levels.

While the theorists defending the top-down approach fought bitterly with the bottom-up clinicians for the last decades of the twentieth century, there is now something of a rapprochement. More and more there is agreement that what is critical is that, for a full and consistent approach to bioethics, eventually all four of these levels must be brought into "equilibrium." It seems less and less important where one starts. If one begins with a case intuition and discovers that that intuition cannot be brought in line with firmly held beliefs about moral rules and principles, then something must give. Either one adjusts the case intuition or, if the case judgment is firm and unrelenting, then maybe the commitments at the higher levels will have to be adjusted. One will move up and down the ladder of the levels of moral discourse. Hence, in Figure 1, the arrows moving from one level of discourse to another are shown pointing in both directions. If one wants a full and consistent set of positions in bioethics, eventually a stable equilibrium is needed. The result is now often called a *reflective equilibrium*. Chapter 9 illustrates how questions at all four levels of moral discourse can be brought together to develop a set of judgments that rests in an equilibrium. In that chapter we examine the current controversies over genetic engineering and new birth technologies: in vitro fertilization, surrogate motherhood, cloning, and stem cell technologies. In these debates we are witnessing the tensions during the process of the emergence of a stable or equilibrium state in the moral debate.

In the case of Yusef Camp, the boy who ate the pickle, it appears that the physicians and the parents are not yet in such a stable state. If they were to start going up the ladder of the levels of moral discourse, they might turn to various codifications of moral rules and rights. Some of their options are presented in Chapter 2.

Key Concepts

Action Theory The branch of normative ethics pertaining to the principles of morally right behavior (as opposed to virtuous character, cf. Virtue Theory, Value Theory).

Antinomianism The position that ethical action is determined independent of law or rules; cf. Situationalism, Rules of Practice, Legalism.

Casuistry The approach to ethics that addresses case problems by applying paradigm or settled cases attempting to identify morally relevant similar and dissimilar features.

Contract Theory A type of metaethics that maintains that the source of moral rightness or the way of knowing what is moral stems from actual or hypothetical social agreement.

Legalism The position that ethical action consists in strict conformity to law or rules; cf. Antinomianism, Rules of Practice, Situationalism.

Metaethics The branch of ethics having to do with the meaning and justification of ethical terms and norms; cf. Normative ethics.

Moral Principles General and abstract characteristics of morally right action. The main elements of part of normative ethics called action theory; cf. Action Theory, Moral Rules.

Moral Rules Concrete statements specifying patterns of morally right conduct, sometimes believed to be derived from more abstract moral principles or, alternatively, created as summaries of patterns of individual case judgments.

Normative Ethics The branch of ethics having to do with standards of right or wrong; cf. Metaethics.

Relativism In metaethics, the position that there are multiple sources or groundings of moral judgments such as the approval of various cultures to which any correct moral judgment must conform; cf. Universalism, Situationalism.

Rights Justified moral or legal claims to entitlements or liberties often seen as taking precedence over ("trumping") considerations of consequences. Rights normally stand in a reciprocal relation with moral or legal rules; that is, if someone has a rights claim against some other party, that other party is duty-bound by a rule requiring that the right be respected.

Rules of Practice The position that rules govern practices such that actions are normally judged by rules; cf. Antinomianism, Situationalism, Legalism.

Situationalism The position that ethical action must be judged in each situation guided by, but not directly determined by, rules; cf. Antinomianism, Rules of Practice, Legalism.

Universalism The position in metaethics that there is a single source or grounding of moral judgments such as the divine will or reason to which any correct moral judgment must conform; cf. Relativism.

Value Theory The portion of normative ethics having to do with rational conceptions of the desirable. Value theory addresses the question of which outcomes are considered good consequences of actions.

Virtue Theory The portion of normative ethics having to do with virtues, that is, persistent dispositions or traits of good character in persons.

Bibliography

WORKS ON BASIC ETHICS

Beauchamp, Tom L. *Philosophical Ethics: An Introduction to Moral Philosophy*, 3rd ed. New York: McGraw-Hill, 2000.

Copp, David. *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press, 2007.

Frankena, William. *Ethics*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Gert, Bernard. *Morality: Its Nature and Justification.* New York: Oxford University Press, 2005.
Rachels, James, and Stuart Rachels. *The Elements of Moral Philosophy*. New York: McGraw-Hill, 2006.
Shafer-Landau, Russ, ed. *Ethical Theory: An Anthology*. Malden, MA: Blackwell, 2007.
Thiroux, Jacques P., and Keith W. Krasemann. *Ethics: Theory and Practice*, 10th ed. Upper Saddle River, NJ: Prentice Hall, 2009.

WORKS ON BIOMEDICAL ETHICS

Baker, Robert B., and Laurence B. McCullough. *The Cambridge World History of Medical Ethics*. Cambridge, England: Cambridge University Press, 2009.
Beauchamp, Tom L., and James F. Childress. *Principles of Biomedical Ethics*, 6th ed. New York: Oxford University Press, 2009.
Beauchamp, Tom L., LeRoy Walters, Jeffrey P. Kahn, and Anna C. Mastroianni. *Contemporary Issues in Bioethics*, 7th ed. Belmont, CA: Thompson Wadsworth, 2008.
Garrett, Thomas M., Harold W. Baillie, and Rosellen M. Garrett. *Health Care Ethics: Principles and Problems*, 5th ed. Upper Saddle River, NJ: Prentice-Hall, Inc., 2009.
Glannon, Walter. *Biomedical Ethics*. New York: Oxford University Press, 2005.
Gert, Bernard, Charles M. Culver, and K. Danner Clouser. *Bioethics: A Systematic Approach*. New York: Oxford University Press, 2006.
Jonsen, Albert R. *The Birth of Bioethics*. New York: Oxford University Press, 1998.
Park, Jennifer A., and Victoria S. Wike. *Bioethics in a Changing World*. Upper Saddle River, NJ: Pearson Education, Inc., 2010.
Pence, Gregory E. *Classic Cases in Medical Ethics: Accounts of the Cases That Have Shaped and Define Medical Ethics.* New York: McGraw-Hill Higher Education, 2008.
Pence, Gregory E. *The Elements of Bioethics*. Boston, MA: McGraw-Hill, 2007.
Steinbock, Bonnie, John D. Arras, and Alex John London, eds. *Ethical Issues in Modern Medicine: Contemporary Readings in Bioethics*, 7th ed. New York: McGraw-Hill Higher Education, 2009.
Veatch, Robert M., Amy M. Haddad, and Dan C. English. *Case Studies in Biomedical Ethics*. New York: Oxford University Press, 2010.
Veatch, Robert M. *A Theory of Medical Ethics*. New York: Basic Books, 1981.
Veatch, Robert M., ed. *Medical Ethics*, 2nd ed. Boston, MA: Jones and Bartlett, 1997.

CHAPTER 1

Validity and Why It Matters

The main topic of this book is the proper evaluation of evidence. This means the proper evaluation of arguments.

An ***argument***, again, is a set of sentences (or, as we'll sometimes call them: claims, statements, or propositions) consisting of one or more *premises* and a *conclusion*. The ***premises*** are statements that are offered as *evidence* for the conclusion, and the ***conclusion*** is the statement whose truth the argument is intended to establish. Logicians typically distinguish between deductive arguments and inductive arguments. Roughly speaking, an argument is ***deductive*** if the truth of premises would guarantee the truth of the conclusion; an argument is ***inductive*** if the truth of premises would render the truth of the conclusion probable, without guaranteeing it. Some inductive arguments are very powerful, and the probability they confer is extremely high. There's nothing *wrong* with an inductive argument just because it doesn't absolutely guarantee its conclusion. Nevertheless, inductive arguments are messier and more complicated than deductive arguments. Thus, in this chapter and Chapters 2 and 3, we will focus on the stronger and simpler kind of argument, the **deductive** argument. Simple doesn't mean easy, and the next two chapters will be a bit abstract, but please bear with us. The skills and concepts mastered here will be important for nearly all other reasoning.

Our provisional understanding of deduction is rough in two ways. First, we'll want to say quite a lot more about what's meant by "guarantee." Second, if we were to define deductive and inductive arguments as those that guarantee or make probable their conclusions, it would follow that there couldn't be *bad* arguments of either type, arguments that abjectly fail to provide the kind of support they're intended to. Consequently, we'll officially define deduction and induction in terms of the *aims* of the argument, that is, in terms of the *intentions* of the person offering the argument. Thus, we will define a ***deductive argument*** as one that *aims* at *validity*, i.e., one that purports to be valid. "Validity," of course, is a technical term that replaces the more intuitive but less precise "guarantee." Just what it means is the topic of this chapter. (We'll say more about aims and intentions a bit in this chapter, but more so in Chapter 3.)

1. DISTINGUISHING THE GOOD FROM THE BAD

The goal here is to distinguish good arguments, ones whose premises provide a genuine reason to believe the conclusion, from bad ones. The good news: You already know a lot about how to do this. From an early age, we reliably use this ability on a daily basis. So,

the task of this book is not to introduce some alien, intellectual discipline, but to develop and refine a skill you already possess. To see that we have this skill, take the following pair of examples:

(P1) All members of species X have lungs.
(P2) y is a member of species X.
(C) Therefore, y has lungs.

(P1) All members of species X have lungs.
(P2) y has lungs.
(C) Therefore, y is a member of species X.

The first is a good argument and the second is a bad one, and we confidently make that judgment. So, in some sense, we already know the difference.

The question now is: can we say what the difference is? What is it about the good argument that makes it good and the bad one that makes it bad? What is the *relevant contrast* between them, the difference that makes a difference?

What counts as a successful answer here? First, we want to know what makes arguments good *in general*, not just the first argument in particular. It is relevant and true to say, "the first argument is good, because "y has lungs" follows from y being a type X and all Xs having lungs," but that answer is too specific. It does not tell us how to evaluate arguments about economics or physics or the likelihood of rain. There's another problem with that answer. To say that the conclusion *follows from* the premises is correct, but unhelpful. If we can't say what that means in simpler terms, saying "it follows" is no more illuminating than saying the argument is good. We haven't *explained* what it is for the argument to be good. The same goes for saying that the conclusion *is a consequence of* the premises, or that the premises *imply* the conclusion, or, if you've already been exposed to some logic, that the argument is *valid*. All true, but they won't explain the idea to someone who genuinely lacks the ability to discriminate the good from the bad, or help us better understand the nature of good argumentation so we can improve our own ability.

We can sneak up on the problem by focusing on the bad argument. It has a hole in it: it could be that all Xs have lungs, but there are other species that also have lungs, and so, y could be one of those. If Xs are dogs, and cats also have lungs, then maybe y is a cat. So, the conclusion would be false.

As it happens, lots of species have lungs. But even if there weren't any other species with lungs, the premises leave open the possibility that such species exist and that is enough for the argument to have a hole in it. And this hole is what makes it a bad argument. The first argument is good because it has no hole; it's airtight: if all type Xs have lungs, and y is an X, the conclusion that y has lungs is inescapable.

There's something to this, but unfortunately, talk of holes is just a metaphor here, and so, it's too wooly to provide precise guidance. There's not *literally* a hole in the argument, as when we say there's a hole in the wall or in my sweater. For someone who doesn't already have the skill of evaluating arguments, telling them to look for holes is vague, hand-waving advice.

But it does capture something important. So, we need to figure out the precise idea to which the metaphor points. Here's a way of putting it: what is special about the first

argument is that the truth of the premises would *absolutely guarantee* the truth of the conclusion; if the premises were true, the conclusion would have to be true. Or, to put it most precisely: it is *impossible* for both the premises to be *true* and the conclusion to be *false* together. This statement is non-metaphorical, and it explains the goodness of the argument in simple terms that do not presuppose specialized logical knowledge: *impossible*, *true*, and *false*. We call arguments like this "valid."

An argument is ***valid*** if and only if it is impossible for the premises to be true and the conclusion to be false together.

The other ones, the ones that lack this special property, we call "***invalid***."

This definition fits with our two examples. What makes the first argument good is that it is absolutely impossible for it to be false that y has lungs, given that it is true that y is an X and all Xs have lungs. What makes the second example bad is that it clearly is possible for all Xs to have lungs, and for y to have lungs, and yet for y not to be an X (i.e., for it to be false that y is an X). And if we fail to recognize the disconnect between the second argument's premises and its conclusion, we are clearly allowing ourselves to be misled, to be persuaded by premises that just don't provide a good reason to believe the conclusion. On the other hand, if we allow ourselves to believe the conclusion of the first argument, given its premises, we make no such error.

More generally, in life we typically want to believe truths *and only* truths. To have any success at that goal, we need to have some kind of policy for deciding what to believe. Here's one policy: every time you are confronted with a proposition, flip a coin. If the coin comes up heads, undertake to believe the proposition; if it comes up tails, don't. This is an obviously bad policy. If you followed it, any truths you came to believe would be a matter of sheer luck, and if you acted on the beliefs you acquired, you probably wouldn't do very well. "Eating the rat poison will be a nutritious and delicious experience": Let's flip a coin.

We need a policy that tracks the truth: picks out truths and avoids falsehoods. Picking out valid arguments and rejecting invalid ones is part of such a policy, a crucial component of it. However, just paying attention to validity is not enough. Validity on its own provides *no reason* to believe the conclusion is true. And this is made explicit in our definition: all it says is that a valid argument can't have *true premises* and a false conclusion. It guarantees *conditional support* between the premises and the conclusion: *If* the premises are true, *then* the conclusion must be too. *If not*, all bets are off. Valid arguments with false conclusions are not hard to find. For instance:

(P1) All human beings have tentacles.
(P2) All creatures with tentacles live in the sea.
(C) So, all human beings live in the sea.

It's valid, but it provides no reason to believe the conclusion. Why? Because one of the premises is obviously false, and valid reasoning from a false premise provides no reason whatsoever to believe that we have a true conclusion.

Another way of putting this is to say that valid arguments are ***truth-preserving***: all true premises guarantee a true conclusion: Truth in; truth out. Falsehood in; who knows? (Unless you're lucky, a false conclusion.) Certainly, the argument gives you no reason to believe it true. So, our policy for truth-tracking should be this: believe only the conclusions

of arguments that are valid *and* that have all true premises. These arguments are important enough that we need a name for them. We'll say an argument is *sound* if and only if it is valid and has all true premises. A sound argument must have a true conclusion: *Truth-preservation + all true premises* guarantees a true conclusion.

A word of caution: these two features of an argument (i) its validity, and (ii) the actual truth values of its premises, *have nothing to do with each other*. It is worth emphasizing this point, as people often mistakenly think that the actual *truth values* of the premises and conclusion—whether the premises and conclusion happen to be true or false—can tell us whether the argument is valid or not. But this is not so. For example:

(P1) Beethoven's music is excellent.
(P2) If someone's music is still well-known centuries after their death, their music must be excellent.
(C) So, Beethoven's music is still well-known centuries after his death.

It's perfectly possible that someone might have written excellent music and also that music only survives the test to time if it is truly excellent, and yet some great composer could be unlucky enough for their work to be lost or destroyed before achieving any popularity, and so never be well-known. That's not how it was for Beethoven, but it could have happened. So, the argument is invalid, and yet the premises and conclusion are all plausibly true. Validity and true premises guarantee a true conclusion, but it doesn't work the other way around: true premises and a true conclusion guarantee nothing about the quality of reasoning.

Just to hammer home the point, let's return to our first pair of arguments:

(P1) All members of species X have lungs.
(P2) y is a member of species X.
(C) Therefore, y has lungs.

(P1) All members of species X have lungs.
(P2) y has lungs.
(C) Therefore, y is a member of species X.

You confidently judged the first valid and the second invalid. But notice, there's no way for you to even assign truth values to the premises and conclusions, *because we never even said what X and y are*. If X is dogs and y is Lassie then the sentences in both arguments are all true. If X is monarch butterflies and y is Charlie the tuna, then they're all false. None of this changes the fact that both arguments of the first kind are valid and that both arguments of the second kind are invalid. The actual truth values are *irrelevant* to assessing validity.

To summarize: how well we are reasoning from a set of assumptions does not depend on whether or not they happen to be true. When you assess validity, you should *completely ignore* whether the premises and conclusion actually happen to be true or false. Consider only the information specified in the premises and in the conclusion and determine whether there is any way at all that things could be as stated in the premises and not as stated in the conclusion. If so, it's invalid; otherwise, it is not.

Exercises 1.1

A. Evaluate whether the following are valid or invalid:

1. Americans landed on the moon in 1969. No Russians landed on the moon before 1969. So, Americans were the first to land on the moon.
2. It is always cloudy when it rains. It is cloudy now. So, it is raining now.
3. Americans first landed on the moon in 1972. No one landed on the moon before them. So, Americans were the first to land on the moon.
4. Mary is Pat's sister. So, Pat is Mary's brother.
5. All fish live in the sea. All things that have scales live in the sea. Therefore, all fish have scales.
6. No carnivore is an herbivore. John is a carnivore. Therefore, John is not an herbivore.
7. John is a friend of Brian. Brian is a friend of Jim. So, John is a friend of Jim.
8. Mary is Pat's sister. So, Pat is Mary's sister.
9. Mary is Pat's sister. So, Pat is Mary's sibling.
10. Obama is not the current president. If George W. Bush was the last president, then Obama is the current president. So, George W. Bush was not the last president.
11. If Nadal wins in straight sets in the final, he will win the tournament. Nadal will not win in straight sets in the final. So, Nadal will not win the tournament.
12. Dogs are bigger than cats. Therefore, cats are smaller than dogs.

B. Evaluate each of the following arguments for validity *and* soundness:

1. The Eiffel tower is in Paris. Paris is in France. So, the Eiffel tower is in France.
2. The Eiffel tower is in Berlin. Berlin is in France. So, the Eiffel tower is in France.
3. The Eiffel tower is in France. Paris is in France. So, the Eiffel tower is in Paris.
4. The Eiffel tower is in Berlin. Berlin is in Germany. So, the Eiffel tower is in Germany.
5. All birds have wings. All things that can fly have wings. So, all birds can fly.
6. Some dogs are pets. Some pets have four legs and a tail. So, dogs have four legs and a tail.
7. All snakes are poisonous. Pythons are snakes. So, pythons are poisonous.
8. Copper is a metal. All electrical conductors are metals. So, copper is an electrical conductor.
9. All metals are electrical conductors. Copper is a metal. So, copper is an electrical conductor.
10. India is the most populous country in Asia. China is the most populous country in the world. The world includes Asia. So, China is not in Asia.
11. Either Bernie or Hillary is going to get the nomination. Bernie can't beat Hillary. Therefore, Hillary is going to get the nomination.
12. It costs about $.50/mile to drive the average car, figuring in gas, maintenance, depreciation, and the like. Therefore, if you drive the average car and work 5 days a week, 50 weeks a year, moving 10 miles closer to work will save you about $2500/year, keeping everything else the same.

13 Obama says he's in favor of gun safety legislation. Anyone who says that means to take our guns away. So, Obama wants to take our guns away.

14 If the mind is entirely physical, it ought to be possible to create artificial minds in computers. It is possible to create artificial minds in computers. Therefore, the mind is entirely physical.

C. More on validity and soundness.

1 Consider the following argument:

> If John has pancreatic cancer, then he will be dead within 12 months.
> John has pancreatic cancer.
> So, John will be dead within 12 months.

(a) Is it valid or invalid?

(b) Suppose a new cure has been found for pancreatic cancer that is 100% effective and available to all patients. Given that information, what should we plausibly say about the above argument? Is it valid or invalid? Is it sound or unsound? Explain your answer.

(c) *Don't* suppose that a cure has been found. Nevertheless, suppose John survives for more than 12 months, i.e., it turns out that the conclusion is false. What should we now say about the argument? Is it valid or invalid? Is it sound or unsound? Explain your answer.

2 Which of the following are possible? If impossible explain why. If a genuine possibility, provide an argument as an illustrative example.

1 An argument that is sound and invalid.
2 An argument that is valid and has a false conclusion.
3 An invalid argument with a true conclusion.
4 An argument that is unsound and valid.
5 An invalid argument with true premises and a true conclusion
6 An argument that is sound and has a false conclusion.
7 An argument with a true conclusion that is unsound.
8 An argument with true premises and a true conclusion that is unsound.
9 A valid argument with a true conclusion and at least one false premise.
10 An argument that has false premises and a true conclusion that is invalid.

2. VALIDITY AND IMPOSSIBILITY

An argument is valid if, and only if, it is *impossible* for its premises to be true and its conclusion to be false. We favored this definition, because it explains a sophisticated idea, validity, in simpler terms: *impossible*, *true*, *false*. However, the first of these terms is not so simple, and we need to talk about exactly what it means.

Here's the concern. In assessing the current state of the U.S. military, a general might say, "*It is not possible* for the U.S. to successfully fight two full-scale wars at once." However, if asked whether with additional investment the U.S. could fight two full-scale wars, the very same general might say, "Sure, *it's possible* for the U.S. to successfully fight two

full-scale wars at once." Superficially, it might look like she's contradicting herself—she's asserted the very same thing to be both possible and impossible—but there's no contradiction here. In the first case, the general is taking one set of background information—the current resources of the U.S. army—as a given, and that puts certain restrictions on what is possible. In the second case, the general is not taking that as background, but considering what is possible if those resources were expanded through investment, and that gives a different specification of what is possible. If we spell out what is meant by the two sentences, they come out as "*given the U.S.'s current military capability*, this is not possible" and "*given more investment in the military*, this is possible." The italicized bits are not explicitly stated by the general but are intended to be understood by whomever she is talking to, presumably on the basis of whatever has already been said in the conversation. So, when we use the word "possible" we almost invariably mean: possible *given the background information assumed in this context*. Different background assumptions; different meaning. The word "possible" is ambiguous: in different contexts, *it means different things*.

Some notions of possibility are sufficiently important to have their own names. It is ***physically impossible*** to travel faster than the speed of light. When we say this we are claiming, rightly or wrongly, that going faster than light is incompatible with the laws of physics. It is ***psychologically impossible*** for mice to do algebra. When we say this, we are not claiming that mice have merely lacked the incentive or education to solve simultaneous equations: we're saying that such an ability is simply incompatible with the mental equipment possessed by mice. It is ***physiologically impossible*** for humans to breathe underwater (without special equipment), meaning that breathing underwater is incompatible with the capabilities of the human body. When we use these names for particular types of possibility, we are signaling which background information should be assumed in understanding our use of "impossible": physically impossible, one should assume the (known) laws of physics; physiologically impossible, assume the capabilities of the human body, and so on. Maybe the most common of these specialized notions of possibility is epistemic possibility (recall that epistemology is the study of knowledge and rational belief). To say that something is ***epistemically possible*** (for me) is to say that it's compatible with everything I know. Since you and I know different things, what's epistemically possible for me need not be epistemically possible for you, and vice versa.

Which notion of possibility is the right one for the definition of validity? Referring back to the start of the chapter, our first example is a valid argument and the second is invalid, and that's just a fact. We feel no temptation to say that the second is invalid in one context and valid in another. So, it's not going to be some squishy, context-dependent notion of possibility, where whether or not an argument is valid depends upon whatever background information *happens to be* assumed in a given context. Instead, it will be a special notion of possibility, one that assumes *no background information whatsoever*. Let's call it "***logical possibility***."

Validity is the gold standard of reasoning. When we say an argument is valid, we are saying that you can rely upon that line of reasoning *in any context whatsoever*. We are saying that it is impossible for the premises to be true and the conclusion to be false, *without assuming any unstated background information*. So, logical impossibility is not incompatibility with the facts about physics or psychology or any other body of background information. What logical impossibility boils down to is literal inconceivability. It means that we

literally *cannot make sense of* such a possibility. It may be physically impossible for things to go faster than the speed of light, but it's not logically impossible; if the laws of physics were different, say they were the ones we thought were true before relativity theory, then things could go faster than light. It is indeed physiologically impossible for humans to breathe underwater, but it's entirely conceivable that humans could have gills or some other organs that would allow us to do so. So, it's logically possible. When you want to assess what is logically possible, bracket all your background knowledge about how things actually are and consider whether you can coherently conceive of the possibility in question.

Here's a helpful, intuitive way of thinking about it. To say that something is logically possible is to say that an omnipotent being could make it true. This includes all kinds of absurd situations: it's logically possible for me to walk on water, for my car to sprout wings and fly, for cats to talk, and so on. But it doesn't include *everything* as possible. Even an omnipotent god couldn't make a four-sided triangle, a chair that is partly green but also completely red, a man that is taller than himself, etc. There's no contradiction in the idea of an omnipotent god that can't make me taller than myself. There is just *no sense to be made* of a man who is taller than himself. So, there just is no such possibility that our hypothetically all-powerful being is incapable of making happen. So, there's no conflict with omnipotence.

Returning to our original example,

(P1) All members of species X have lungs.
(P2) y is a member of species X.
(C) Therefore, y has lungs.

What this argument has going for it, which you instinctively realized made it a good argument, is that we can make no sense of the premises both being true and the conclusion being false. *Without assuming any unstated background information whatsoever*, it is impossible for the premises to be true and the conclusion to be false. It's *logically* impossible, and that's what we mean by validity.

By adopting a very inclusive conception of possibility, we end up with a very exclusive conception of impossibility. This ensures that validity (because it's defined in terms of this conception of impossibility) embodies the highest level of praise possible; *there simply couldn't be a kind of conditional support better than validity*. That is why we say validity is the gold standard of reasoning. Nevertheless, you might think it ridiculous to demand such a high standard, to define validity using a notion of possibility on which we allow that pigs could fly and cats could talk. The following arguments are *invalid* precisely because such phenomena are logically possible:

(P1) Slypork is a pig.
(C) So, Slypork can't fly.

(P1) Jeoffrey is a cat.
(C) So, Jeoffrey can't talk.

Yes, both lines of reasoning seem very sensible, *given our common-sense background knowledge about pigs and cats*: if you tell me only that Slypork is a pig, I will readily accept on

that basis that he can't fly. But that's beside the point here. We've already explained one reason why: a valid argument is one whose reasoning *can be relied upon in any situation whatsoever*, because it does not depend upon any background assumptions. These arguments simply don't have that general reliability. Why it matters, why we are insisting on validity—at least for now—is something we'll explain further in section 1.5. For the moment, let's just accept the gold standard.

To summarize the last two sections: assessing validity is an exercise in abstraction in two different ways. First, because validity is a matter of *conditional* support, we need to forget about whether the premises or conclusion are actually true. Second, because validity is a matter of logical possibility, we need to forget a host of background information about the world: what the laws of physics and psychology are, whether humans can breathe underwater, etc. We just focus on what the premises and conclusion say, and determine whether it is in any way conceivable for the premises to be true and the conclusion to be false together.

Now that you understand the technical concept of validity, there's just one more small point to understanding our definition of a ***deductive*** argument as one that aims at validity, i.e., that purports to be valid. Obviously, *arguments* don't have goals or aims, but the people who offer them do. We'll see in more detail in Chapter 3 that understanding arguments requires understanding the intentions of the authors of those arguments. For now, it's enough to note that we'll count an argument as deductive if the author *meant* for it to be valid. Thus, there will be *failed* deductive arguments: arguments that aimed at validity but aren't valid.

Box 1.1 Possibility and Necessity

Possibility and necessity are interdefined. Something is possible if and only if it's not necessarily false. Something is necessary if and only if it's not possibly false. Something is ***contingent*** if and only if it's neither necessarily true nor necessarily false.

Box 1.2 Some Important Notions of Possibility

***Logically possible*:** conceivable without contradiction
***Physically possible*:** compatible with the laws of physics
***Epistemically possible* (for S):** compatible with everything S knows

Exercises 1.2

A. For each of the following, specify whether the kind of possibility/impossibility at issue is logical or not. If not, say what we might call it, using our examples as guides (so, although we didn't mention "biological possibility," surely there is such a thing as compatibility

with the laws of biology, and so on). If it doesn't easily admit of a name, indicate what the relevant background assumptions are.

1. It's impossible for a fire to burn without oxygen.
2. Dogs can't face their "palms" toward each other.
3. A triangle must have three sides.
4. You can't go to jail for cheating on your spouse.
5. Sound must travel slower than light.
6. If Margaret and Joe are both here, then Joe must be here.
7. You can't survive a zombie bite without becoming a zombie yourself.

B. For the following pairs of sentences, determine whether it is logically possible that both be true at the same time. If so, offer a scenario that explains how. Note, this will sometimes require outlandish, though not self-contradictory, suppositions.

1. It's raining. The streets are dry.
2. I dropped this rock. It never hit the ground or any other surface.
3. Jessica died this morning. She was a half-hour late to work this afternoon.
4. Women are smarter than men. Men score higher on IQ tests than women.
5. Everyone loves Taylor Swift and her music. Taylor Swift is terribly unpopular.
6. Sam Shepard is the greatest English-speaking playwright of all time. Shakespeare existed.

C. The following pairs couldn't possibly both be true at the same time. In each case, how would you explain this fact to someone who didn't yet see why?

1. X is a triangle. X has four sides.
2. Sandy and Jules came to the party. Sandy didn't come to the party.
3. If there's an open flame, there's oxygen. There's an open flame but no oxygen.
4. I am taller than Lewis. I *am* Lewis.
5. Rene thinks he might be dreaming. Rene doesn't exist.
6. Things are going to get better, or they're going to get worse. Things are going to stay exactly the same.
7. I'm in Paris, and Paris is in France. I'm not in France.
8. Nobody's ever run a four-minute mile. Jolene ran a mile in 3:54.

D. Evaluate the validity of the following arguments:

1. A cat falls from the top of a tall building and strikes the ground at over 100 mph. Therefore, it is seriously injured or killed.
2. The Empire State Building is made entirely of soap bubbles. All soap bubbles disintegrate within 5 minutes. No new soap bubbles are made. So, there will be no Empire State Building in 5 minutes.
3. John, a human being, is entirely submerged in water for 10 hours without any kind of breathing apparatus. So, John dies.
4. All human beings live in pineapples under the sea. Everyone that lives in a pineapple under the sea is a friend of SpongeBob. So, every human being is a friend of SpongeBob.
5. John is unmarried. So, John is a bachelor.

3. MORE ON LOGICAL IMPOSSIBILITY

We have defined a valid argument as one for which it is *logically* impossible for the premises to be true and the conclusion to be false, and that means that the truth of the premises guarantee the truth of the conclusion, *without our assuming any background information whatsoever*. However, there's a way in which this requirement might seem confusing. Take the following argument:

(P1) John is a bachelor.
(C) Therefore, John is unmarried.

If anything is a valid argument, this is. However, someone might be concerned that we used background information in judging it as valid. To make that judgment, we needed to know that a bachelor is someone who is unmarried—we needed to *assume background information* about what "bachelor" means. This is true, but it's unavoidable and shouldn't be troubling: if we don't allow background information about what the words in the argument mean, *we can't even understand what the argument is saying*, let alone evaluate it. So, being a little more careful, what we mean is that when you evaluate the possibility of the premises being true and the conclusion being false, the *only* background information allowed is the information required for understanding the sentences that make up the argument. Consider the following four arguments:

(P1) The earth's orbit is a circle.
(C) So, the earth's orbit has no corners.

(P1) My father is 40 years older than I am.
(C) So, I am 40 years younger than my father.

(P1) Ice is just solid water.
(C) So, ice will melt when heated above 32° F.

(P1) The cue-ball struck the 8-ball with great force.
(C) So, the 8-ball moved.

Valid or invalid? The first two are valid because of what the terms in the arguments mean. A circle is a curve on which all points are equidistant from one point, the center, and a curve like that is smooth; it has no corners. And if you don't know that, either you don't understand what "circle" means or you don't understand what "corner" means. Given that you do understand, the conclusion is inescapable, Similarly, if my father is 40 years older than me, that just means that I am 40 years younger than him. Once you understand what is meant by "older than" and "younger than," it's inconceivable that the premise is true and that the conclusion is false.

Let's look at the second pair of arguments. We all know that water is liquid if it is above its freezing point, 32° F, and hence, given that ice is solid water, heating it will yield liquid water, i.e., the ice will melt. That's a piece of knowledge about

the world that we have had for so long that we have likely forgotten when first we learned it, but that's not the same thing as being part of the meaning of "water." We can see this pretty easily. Suppose we found that some chemical process resulted in a new kind of ice, a solid form of water with a different crystal structure to regular ice. And suppose that if you heat crystals of that ice, they don't liquefy until you heat them above 110° F. We would not describe that situation by saying that the new ice crystals aren't really water. We would say, "Here's a fascinating surprise about water: you can make water ice that doesn't melt even when you push its temperature way above 32° F." In fact, this is the possibility considered in Kurt Vonnegut's novel *Cat's Cradle*. A scientist discovers a novel crystal structure for water, called Ice-9, which is solid at high temperatures. (Spoiler alert: a crystal of Ice-9 is dropped into the oceans, solidifying all of earth's water, triggering an ecological disaster, and the end of humanity.)

The second invalid argument is very similar. We all know that if one pool ball hits another with great force in normal circumstances, it will cause it to move. But again, the given premise does not render the conclusion's falsehood impossible in the relevant sense; it's the given premise, plus that additional background information that guarantees the conclusion's truth. We can easily form new valid arguments by incorporating these background beliefs as premises, giving:

(P1) Ice is just solid water.
(P2) Solid water will always melt when heated above 32° F.
(C) So, ice will melt when heated above 32° F.

(P1) The cue-ball struck the 8-ball with great force.
(P2) Given the circumstances, if the 8-ball was struck with great force, it moved.
(C) So, the 8-ball moved.

These are now valid, but the originals were not.

Note an important difference between the last two invalid arguments and the two valid ones that preceded them: we *could* give those the same treatment. For example:

(P1) The earth's orbit is a circle.
(P2) Circles don't have corners.
(C) So, the earth's orbit has no corners.

This makes the original reasoning a bit more explicit, but note: it doesn't improve the conditional support for the conclusion. It *couldn't*, because the original argument is valid, and validity can't be improved on. If you don't know anything about the world, but I tell you that the earth's orbit is a circle, then I've already told you all you need to validly infer that the earth's orbit has no corners. I don't need to tell you that circles don't have corners, because if you didn't already know that, then *you didn't understand the rest of the argument*. You can't understand what a circle is and what a corner is unless you know that circles have no corners. Thus, P2 is dispensable, because it doesn't—couldn't!—tell you anything new.

By contrast, in the ice and 8-ball cases, the (P2)s are not logically necessary truths, and they're not things you must already know in order to understand the (P1)s. If you don't know anything about the world, but I tell you that ice is just solid water, I haven't told you everything you need to know to infer that ice will melt at 32° F. *You'd still need to be told that water melts at 32° F*. You could understand the terms "water," "ice," "melt," and "32" without knowing this.

Sometimes it's hard to know exactly where to draw the line between contingent, factual background knowledge of the sort that we're not allowed to presuppose in assessing validity, and necessary, meaning-related knowledge that just makes explicit what we already had to know in order to have the relevant concepts. For instance, what is the status of the following argument?

(P1) Moby Dick is a whale.
(C) Therefore, Moby Dick is a mammal.

Obviously, it doesn't hurt to explicitly add,

(P2) All whales are mammals.

It's true, and it makes the argument obviously valid. But was the argument valid without it? Is it part of the *meaning* of "whale" that whales are mammals? It is tempting to think so. In some sense you couldn't fully know what whales *really* are without knowing that they're mammals. On the other hand, P2 is a relatively recent discovery, before which, lots of people were capable of forming thoughts about whales. In *Moby Dick*, the narrator, Ishmael, believes that whales are fish. But surely he knows what the word "whale" means, even if he's badly mistaken about their nature. When someone shouts, "Avast! There blows a sperm whale off the starboard bow!" Ishmael understands perfectly well what's being claimed, and he forms the belief that there's a whale off the starboard bow. Someone, like Ishmael, could understand (P1) without knowing (P2). But *that means that (P2) does genuinely add some factual information to (P1)*. Thus, the argument is *invalid* without (P2) although valid with it. So, when in doubt, you should *always demand that the information be made explicit*. You might accidentally count some valid arguments as invalid, but in the end, no real harm comes from being too careful.

Box 1.3 Concepts and Definitions

To have the concept *square* or *cat*, etc. is just to have the ability to think thoughts about squares, cats, etc. Sometimes having a concept does seem to amount to knowing a definition: you have the concept *prime number* only if you know that it is something that is divisible only by itself and one. But having a concept isn't always a matter of knowing a definition: there isn't any particular thing you need to know about cats in order to think thoughts or understand sentences about cats. Knowing what they look like may be sufficient for having the concept; and knowing what they sound like may be sufficient; but neither is necessary.

In assessing arguments for validity, the background knowledge you're allowed to bring in is just the knowledge required for having the concepts involved in the premises and conclusions—that is, the background knowledge required for understanding these statements. Sometimes this may involve definitions, sometimes it won't.

Box 1.4 Definitions and Natural Language

The fact that dictionaries exist suggests that it's not hard to find definitions for all or most terms. But it's more complicated than this. Consider "bachelor." Suppose we define it as an unmarried male. That's not right. A 5-year-old boy doesn't count as a bachelor. Fair enough, how about unmarried male of marriageable age? A man who was married, but has been widowed, satisfies this definition. Is he a bachelor? Some may say "yes"; some may say "no." Most will shrug their shoulders and look puzzled. What about a monk, or the Pope, who have taken vows not to marry? It doesn't seem that there's a clear-cut definition of even a simple concept like *bachelor*, but we can all think thoughts about bachelors. Surprisingly, it seems that having a concept is not, in general, the same thing as knowing a definition.

Notwithstanding potential disagreements over obscure cases, we all know that bachelors are, necessarily, unmarried men. So, even if the concept is not specified by a definition, it remains that the argument from John being a bachelor to John being unmarried is indeed valid. It's worth repeating, however, that when in doubt, it's best to err on the side of caution: if you make explicit the fact that all bachelors are unmarried as an added premise, then the argument will be obviously valid.

Exercises 1.3

A. For each of the following arguments, say what background assumption would have to be added to render it a fully explicit, valid argument. Then say whether the argument was already valid without it.

1. It's a domestic cat. Therefore, it's someone's pet.
2. It's a cat. Therefore, it's warm-blooded.
3. You're voting for Gil Fulbright. So, you're a Republican.
4. It's January. So, we'll get some snow soon.
5. It's Tuesday. Therefore, the day after yesterday is Tuesday.
6. You have Lyme disease. You must have been bitten by a tick!
7. All electrons are negatively charged. So, all electrons repel each other.
8. She was your date for your senior prom? I guess she really hated her parents!

9 I had an appendectomy six weeks ago. Therefore, I have undergone a surgical procedure in the past year.
10 You're down to a half a gallon of gas and you have 140 miles to go. Therefore, you won't get there without stopping to refuel.

While we're on the subject of meaning and its relation to validity, it's worth making note of a couple of important topics, to which we shall return in more detail later in the book.

Logical Terms

We've been looking at arguments whose validity hinges on the meanings of terms like "bachelor" and "circle," but arguments whose validity depends on the meanings of broadly ***logical terms***, like "some," "all," "and," "not," "if. . .then," are even more important. Consider the following:

(P1) All cetaceans are heterotrophs.
(P2) Mauyuk is a cetacean.
(C) Mauyuk is a heterotroph.

Even if you don't know what "cetacean" and "heterotroph" mean, or who Mauyuk is, you know what "all" means, and that's enough to know that this argument is valid.

We can show this a little more clearly by drawing a diagram. (P1) says that the class of cetaceans is included in the class of heterotrophs, that you can't be a cetacean without being a heterotroph. We can illustrate this by placing a circle that represents cetaceans inside a circle that represents heterotrophs. That way, anything that falls in the C circle is going to fall in the *H* circle as well:

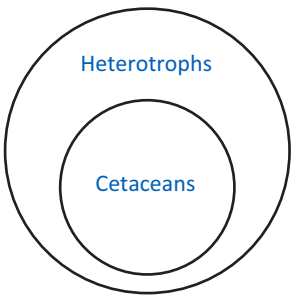


FIGURE 1.1

To represent the claim that Mauyuk is a cetacean, we use a dot or an x or something that stands for her, and we put it in the *Cetaceans* circle.

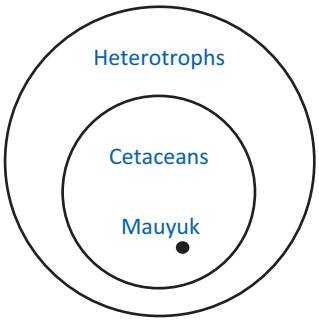


FIGURE 1.2

But now it is quite obvious that, because Mauyuk is a cetacean—she falls inside the smaller circle—she *must be* a heterotroph too—she must fall inside the larger, more inclusive circle. We can't draw a *Mauyuk* dot inside the *Cetaceans* circle without also putting it inside the *Heterotrophs* circle. This shows that from the fact that Mauyuk is a cetacean (P2) we can deduce with certainty that she is a heterotroph (C). On the other hand, if our second premise had instead claimed that Mauyuk was a heterotroph, we couldn't have validly inferred that she's a cetacean. This is because she might have—for all our premises have told us—fallen outside the *Cetaceans* circle but still inside the *Heterotrophs* circle.

As we progress, we'll see that we can learn a lot about validity by focusing on such logical terms: broad classes of arguments turn out to be valid precisely because of the patterns or forms in which such terms are used. The above case illustrates this. It just doesn't matter that the argument is about cetaceans or heterotrophs or about Mauyuk. All instances of the form

(P1) All Cs are Hs.
(P2) m is a C.
(C) m is an H.

will have to be valid. By contrast, arguments whose validity depends on specifics of the meanings of terms like "cetaceans" or "bachelor" are of little *general* interest.

Moreover, the meanings of logical terms are typically clearer and less ambiguous than the meanings associated with non-logical terms. Generally, when people say something of the form "*All* Xs are Ys" there is just one thing meant by the term "all." This also renders such terms as suitable targets in our attempt to characterize the properties of broad classes of arguments. By contrast, non-logical vocabulary is not generally so well-behaved. This brings us to the second meaning-related topic we need to briefly discuss: *equivocation*.

Equivocation

In judging the last argument valid, we reasonably assumed that the terms used in multiple locations ("cetacean," "heterotroph," "Mauyuk") meant the same thing each time they occurred. However, if a term is ambiguous, it can lead to the ***fallacy of equivocation***, where an argument uses a single word or phrase in two or more different ways, so that the

argument has the appearance of being sound, even though it isn't. Here's a toy example to illustrate the concept. Consider the following argument:

(P1) My nephew is just a kid.
(P2) Kids are baby goats.
(C) Therefore, my nephew is a goat.

Obviously, "kid" is being used differently in these two premises. The argument looks valid, but it is only valid if "kid" means the same thing in both premises. However, if we use it to mean "child," then (P1) is true but (P2) is false; if we use it to mean "baby goat," then (P2) is true but (P1) is false. So, on either reading, if the argument is valid, it's unsound; at least one of the premises is false. If, on the other hand, we use "kid" one way in (P1) and a different way in (P2), we can get two true premises, but now the argument's invalid, and again, unsound. We discuss equivocation in more detail in Chapter 10.

4. LOGIC AND THE BELIEF BIAS

We've tried to be really careful in laying out what a valid argument is and how to evaluate simple arguments. However, even if you've taken everything we've said onboard, it is still easy to be misled about an argument's validity. Consider the following argument:

(P1) Anything that has a motor needs oil.
(P2) Cars need oil.
(C) Therefore, cars have motors.

Valid or invalid? Remember that validity is about conditional support and not about the actual truth values of the premises and conclusions. Many people are inclined to think the argument is valid. Now consider another argument:

(P1) Anything that has a motor needs oil.
(P2) Opprobines need oil.
(C) Therefore, opprobines have motors.

Most people are not inclined to think that this argument is valid.[1] Notice, however, that the two arguments are perfect parallels. We're inclined to think the first argument is valid because we know that the conclusion is true. The argument, however, is invalid, as we can more easily see in the case of the second argument, where we immediately realize that things that don't have motors might also need oil for other reasons, and opprobines/cars might be among these things. This is a phenomenon known as the ***belief bias***: people tend to judge invalid arguments to be valid if the conclusion is something they already believe to be true.

Obviously this is a bad kind of mistake to make. Among other things, it will keep us stuck in a cognitive rut. If we once adopt a belief for no good reason at all, the mere fact that we now believe it will make bad reasons for that belief look like good ones.

This tendency makes it difficult for us to learn, because it makes it difficult for us to be open-minded. And of course, it makes it difficult for us to be objective; we are naturally prone to find (real) flaws in the arguments of those who disagree with us, but to overlook (real) flaws in the arguments of those who agree with us. It's hard to overstate how bad this is from the perspective of believing all and only what's true.

The culprit here, of course, is System 1: we have an automatic, intuitive sense of validity and are prone to judge accordingly.[2] But this sense of validity is highly unreliable, because it's so heavily influenced by our fallible prejudgments about the conclusions. Even if we were perfect, infallible judges about the conclusions, this would still be a bad guide to validity, since it's quite possible for an invalid argument to have true conclusions or a valid argument to have false conclusions.

Our strategy for evaluating arguments is clear: Use System 2, don't let System 1 foist beliefs on you that just seem right. System 2 will be much more reliable, at least once you've learned how to properly distinguish between valid and invalid arguments.

Exercises 1.4

A. Assess the following arguments for validity. Do so as quickly as you can, writing down whatever verdict pops into mind.

1. People who are opposed to freedom support gun control. Liberals support gun control. So, liberals are opposed to freedom.
2. People who don't care about the sick, the disadvantaged, and the elderly support cuts to welfare. Conservatives support cuts to welfare. So, conservatives don't care about the sick, the disadvantaged, and the elderly.
3. People who write graphic novels require the storytelling skills of a writer and the visual imagination of a good film maker. Film directors also need both of those skills. So, film directors write graphic novels.
4. People who totally buy into the scientific worldview are anti-religion. People who totally buy into the scientific worldview also believe in global warming. So, people who believe in global warming are anti-religion.
5. People who accept the results of well-established science accept the reality of global warming. People who accept the reality of global warming must be anti-religious. So, people who accept the results of well-established science must be anti-religious.
6. People with strong fundamentalist religious beliefs deny global warming. People who, for one reason or another, won't honestly face the evidence deny global warming. So, people with strong fundamentalist religious beliefs won't honestly face the evidence about global warming.
7. Historically, the noble potato has long been a valuable food source in poorer societies across the world. Even today, in several countries it is common for potatoes to be on the dinner table almost every night of the week. As is well-known, this is true in Ireland, but also in Poland, Peru, and several other

countries. So, we can conclude that these societies are still quite poor. They can still not be counted among the richer nations of the world.

8 Even people from the poorest of backgrounds have succeeded in life with hard work and dedication. So, anyone, even someone from a very poor background, can be successful.

B. Go back through the arguments of section A just now, and this time, take your time evaluating them, keeping in mind the fact that we're all more likely to find an argument valid if we already believe the conclusion. Was there any difference between your quick assessment and your slower, more careful assessment?

C. Which of the following arguments are valid and which are invalid?

Watch out for belief bias, i.e., uncritically accepting an argument as valid, just because you think the conclusion is true.

1 New York is bigger than Houston. Houston is bigger than San Francisco. So, New York is bigger than San Francisco.

2 Los Angeles is bigger than New York. Los Angeles is bigger than San Francisco. So, New York is bigger than San Francisco.

3 Abraham Lincoln and Bill Clinton were both U.S. presidents. Lincoln is dead, but Clinton is still alive. So, Lincoln was president before Clinton.

4 George H.W. Bush was the 41st president of the U.S. Bill Clinton defeated him in the presidential election in 1992. So, Bill Clinton was the 42nd president of the U.S.

5 The Empire State Building is in New York. You can see Brooklyn from the top of the Empire State Building. So, Brooklyn is in New York.

6 Mount Everest and K2 are both in the Himalayas. Mount Everest is taller than K2. K2 is the second tallest mountain in the Himalayas. So, Mount Everest is the tallest mountain in the Himalayas.

7 The Himalayas contain more tall mountains than any other mountain range on earth. Mount Everest is taller than any other mountain in the Himalayas. So, Mount Everest is the tallest mountain on earth.

8 Michael Phelps has won more Olympic gold medals than any other swimmer. Successful Olympic swimmers typically win more medals than other athletes. So, Phelps has won more Olympic gold medals than any other athlete.

9 Rio de Janeiro is in Brazil. Brazil is right beside Argentina. Argentina is in South America. So, Rio de Janeiro is in South America.

10 Washington D.C. is the capital of the U.S. The president's office, the Oval Office, is in the White house, which is in Washington D.C. So, the president lives in Washington D.C.

D. Go back through the arguments in 1.1 B and re-assess for validity and soundness. Note any cases where belief bias initially leads you astray in assessing for validity.

5. WHY IT MATTERS: MISSING PREMISES AND INSISTING ON VALIDITY

Validity is the highest possible standard, since it demands that it's *logically* impossible that the conclusion is false and all the premises true. You might think that the very notion of logical possibility/impossibility is rather silly and impractical, especially when we remember that absurd situations, like talking typewriters, are logically possible. Do we really need such a high standard?

First of all, not all good arguments are valid, as we'll start to see in detail in Chapter 4. But valid arguments are the cleanest, simplest kind, and it's best to start where everything is pure and simple before moving on to the messy and complicated. Also, some arguments do meet this extremely high standard. If we didn't have the concept of validity, we wouldn't be able to explain how these arguments differed from other arguments, with a weaker degree of conditional support.

Most importantly, when we insist on validity, this forces us to make explicit premises that we had left unstated. This, in turn, forces us to directly confront our unarticulated and maybe unconscious assumptions. This can have an enormous beneficial effect on our thinking.

Suppose we had a less exacting standard, one that accepted the original Ice argument as perfectly adequate. We would have a large blind spot in our understanding of the world. Someone who thinks the reasoning of that argument is adequate has just *ruled out*, without giving it a thought, the possibility that there might be different types of water ice with different melting points. If we do not require that it is *logically* impossible for the premises to be true and the conclusion to be false, we will *fail to track the truth* in an important way: we will implicitly assume we know things that we don't—in this case that *all* solid water melts at 32° F. That is why we hold arguments to the highest possible standard. If we don't, we will have intellectual blind spots.

Insisting on validity, rejecting a line of reasoning until it's clearly valid, compels us to insert the extra premise:

(P2) Solid water will always melt when heated above 32° F.

With this assumption out in the open, we no longer have that intellectual blind spot. We can now assess this premise and recognize that our evidence for it may indeed be surprisingly weak—yes, any water *I have seen* melts when heated above 32° F, but is it *obvious* that even in unusual or exotic circumstances that is *always* the case? Surely not. Answering that question demands scientific research, not just casual observation of the behavior of water under normal circumstances.

An incomplete argument is called an ***enthymeme***. Converting enthymemes into fully stated arguments is a way to drag unstated assumptions into the light. Very often, the unstated assumptions, *the missing premises*, are the weakest. In many cases, they turn out to be obviously false, or at best highly controversial and unsupported by the evidence. The belief bias means we often won't notice this unless we actively engage System 2 and self-consciously assess the stated argument's validity. So, insisting on validity, and the associated reconstruction of arguments and interrogation of the premises we uncover is

one of the most powerful tools for reasoning and arguing reliably. We'll have more to say about enthymemes and about argument reconstruction in Chapter 3, but for now, we want to note that enthymemes are very common, and that filling in the missing premises to make an argument valid is often extremely illuminating.

Through much of history, it was accepted that the earth sat entirely stationary at the center of the universe, with the sun and other objects orbiting around it. When scientists and philosophers considered the possibility that the earth orbited the sun and that it also rotated on its own axis, they were met with great skepticism. Just the rotation of the earth on its axis seemed to be ruled out by an argument from the most straightforward observational data:

(P1) If the earth rotates, we are moving at about 1,000 mph.
(P2) We don't *seem* to be moving.
(C) So, the earth isn't rotating.

This seems pretty reasonable, especially if you're living in the 1500s and have the belief bias working in favor of the argument, rather than against it. It was known that the earth's circumference was about 24,000 miles, and so, given that the earth has to rotate once every 24 hours, someone at the Equator would be moving at 1,000 mph. Even far north of that, you could work out that we would be moving at a substantial fraction of that speed. So, (P1) is true. However, the argument is not valid, and we can easily see what's missing. To be valid it needs a further premise:

(P3) If we are moving at about 1,000 miles per hour, it will seem to us that we are moving.

But why should we think *this* is true? Obviously, if I'm moving at about 1,000 miles per hour relative to the ground below me and relative to the air, then I'd feel that. But if we're all moving at the same time, would it still seem that we're moving? We can drink coffee on an airplane without it flying into our faces because, although we're moving at 500 mph, so is the coffee. Once we make (P3) explicit, we're in a position to see that it's not so obvious after all.

Most of the great innovations in science, technology, business, and a wide range of other fields resulted from someone being the first to notice that everyone else had been assuming something without ever making that assumption explicit. This is the nature of assumptions; they're things we *unreflectively* take for granted. They shape and direct our thinking without our even recognizing their influence. Consider the following, plausible but invalid, argument.

(P1) The fastest jet from Los Angeles to San Francisco takes an hour and 15 minutes.
(C) There's no way to travel from Los Angeles to San Francisco in less than an hour.

This argument presupposes that there is no mode of travel faster than a jet:

(P2) There is no faster way to travel from Los Angeles to San Francisco than by jet.

Although this is currently true, it doesn't have to stay that way. We can imagine train-sized capsules using magnetic levitation, being shot through 100-mile-long low-pressure tubes, with no friction and very little air resistance. This might get you there in half an hour. We can imagine a computerized system that makes an extremely detailed scan of your entire body in Los Angeles, that sends that scan at the speed of light to San Francisco, that disintegrates your body in Los Angeles, and that uses that scan to reconstruct you with a sophisticated 3-D printer in San Francisco. This might get you there in seconds.

The point is that we can't question our hidden assumptions until we've made them explicit. This allows us to imagine other possibilities and maybe to make them a reality. Insisting on validity forces us to explicitly articulate assumptions that we hadn't realized we'd been relying on. Sometimes these will be trivial facts, like that all whales are mammals. But sometimes we will discover that the premises we'd been relying on are substantive, questionable, and/or alterable.

All of this requires that we slow down, override our cognitive autopilot, and engage System 2. The belief bias shows us that we are *naturally strongly disposed* to accept poor arguments as decisive, simply because we already happen to believe their conclusions. Given this, if we don't deliberately verify that each argument really is valid or supply the missing premises that will make it so, we're believing what we already believed. If you don't even know what the full argument should look like, and System 1 is mindlessly giving it the thumbs up, you have little hope of spotting cases where you believe things for bad reasons. So, insisting on validity isn't just advice for the classroom or the science lab; it's practical advice for everyday life.

SUMMARY

An ***argument*** is a piece of reasoning that is intended to establish the truth of a conclusion. An argument consists of one or more premises and a conclusion. The ***premises*** are statements that are offered as ***evidence*** for the conclusion. The ***conclusion*** is the statement whose truth the argument is intended to establish.

Logicians divide arguments into two broad categories: deductive and inductive. An argument is ***deductive*** if it is intended that the truth of premises would guarantee the truth of the conclusion. An argument is ***inductive*** if it is intended that the truth of premises would render the truth of the conclusion probable or likely.

An argument is ***valid*** if, and only if, it is *(logically) impossible* for the premises to be true and the conclusion to be false together.

The terms *"possible"/"impossible"* are systematically ambiguous, i.e., they can have entirely distinct meanings in a different context. The relevant notion for defining validity is logical impossibility. Something is ***logically impossible*** if it is not conceivable without contradiction. Thus, in evaluating an argument's validity, we are not allowed to assume *background beliefs* (however commonsensical) that are not in the stated premises. Adding such beliefs to the premises of the argument means we are not assessing logical impossibility but some other notion of possibility (e.g., physical possibility, epistemic possibility). Hence, we are not assessing the stated argument's validity. The only background beliefs that we are allowed to use in assessing validity concern *the meanings of the words used to*

state the argument. This is unavoidable: if you don't know what the argument means, you can't understand it, let alone assess it.

A valid argument is one whose premises provide perfect ***conditional support*** for its conclusion: *if* all of its premises are true, then its conclusion must be true. To put it another way, a valid argument is *truth-preserving*. A valid argument with all true premises is ***sound***. (Hence), a sound argument has a true conclusion. However, if even one of the premises is false, then the fact that an argument is valid/truth-preserving provides no reason to believe its conclusion. Similarly, even if the premises are all true, if the argument is invalid, it provides no reason to believe its conclusion.

So, to judge whether an argument gives us reason to believe its conclusion, we assess its soundness, i.e., we must verify *two entirely distinct things*:

(i) The argument is valid.
(ii) The premises are all true.

It is easy to fail to carefully separate these two tasks, and hence, to incorrectly assess an argument's validity and/or soundness.

Moreover, we are all susceptible to ***belief bias***: judging an argument as valid, merely because we already believe its conclusion to be true. This should be a source of great concern, since it can give a powerful illusion that a belief is well supported by arguments when the arguments in question are defective.

An incompletely stated argument is an ***enthymeme***. Often the unstated/missing premises are the controversial ones. Making the missing premises explicit is an important tool for avoiding sloppy reasoning and for uncovering the reasons why people disagree about important matters. That is why we insist on validity in assessing deductive arguments.

The ***fallacy of equivocation*** is where a single word or phrase is ***ambiguous***: it has two or more meanings, so that an argument has the appearance of being sound, even though it isn't. There is no single reading of the ambiguous term that makes all the premises true.

NOTES

1 Markovits, Henry, and Guilaine Nantel. "The belief-bias effect in the production and evaluation of logical conclusions." *Memory & Cognition* 17.1 (1989): 11–17.

2 In the Introduction, section 2, we introduced dual systems theory, which distinguishes between System 1 (the intuitive processor) and System 2 (the conscious rule interpreter). System 1 is fast, automatic, and effortless, while System 2 is the opposite. In addition, System 1 operates unconsciously, so we can't tell which of the resulting intuitions are reliable and which are not (they'll all *seem* true). To ensure that your judgments are reliable, you'll need to use System 2.