The Philosopher’s Toolkit: How to Be the Most Rational Person in Any Room
Course Guidebook

Professor Patrick Grim
State University of New York at Stony Brook
Professor Patrick Grim is Distinguished Teaching Professor of Philosophy at the State University of New York (SUNY) at Stony Brook, where he has taught since 1976. Having graduated with highest honors in both Anthropology and Philosophy from the University of California, Santa Cruz, Professor Grim was named a Fulbright Fellow to the University of St. Andrews, Scotland, from which he received his B.Phil. He received his Ph.D. from Boston University with a dissertation on ethical relativism and spent a year as a Mellon Faculty Fellow at Washington University. In addition to being named SUNY Distinguished Teaching Professor, he has received the President’s and Chancellor’s awards for excellence in teaching. Professor Grim was named Marshall Weinberg Distinguished Visiting Professor at the University of Michigan in 2006 and Visiting Fellow at the Center for Philosophy of Science at the University of Pittsburgh in 2007. He also has been a frequent Visiting Scholar at the Center for the Study of Complex Systems at the University of Michigan.

Professor Grim has published extensively in computational modeling and on such topics as theoretical biology, linguistics, decision theory, and artificial intelligence. His work spans ethics; philosophical logic; game theory; contemporary metaphysics; and philosophy of science, law, mind, language, and religion.

Professor Grim is the author of *The Incomplete Universe: Totality, Knowledge, and Truth* and the coauthor of *The Philosophical Computer: Exploratory Essays in Philosophical Computer Modeling* (with Gary Mar and Paul St. Denis); *Beyond Sets: A Venture in Collection-Theoretic Revisionism* (with Nicholas Rescher); and *Reflexivity: From Paradox to Consciousness* (also with Nicholas Rescher). He is the editor of *Mind and
Consciousness: 5 Questions and Philosophy of Science and the Occult and a founding coeditor of more than 30 volumes of The Philosopher’s Annual, an anthology of the best articles published in philosophy each year. He has taught two previous Great Courses: Questions of Value and Philosophy of Mind: Brains, Consciousness, and Thinking Machines.
# Table of Contents

## INTRODUCTION

Professor Biography ....................................................................................... i  
Course Scope .................................................................................................. 1

## LECTURE GUIDES

**LECTURE 1**  
How We Think and How to Think Better ...................................................... 4

**LECTURE 2**  
Cool Rationality and Hot Thought ............................................................... 13

**LECTURE 3**  
The Strategy of Visualization ...................................................................... 22

**LECTURE 4**  
Visualizing Concepts and Propositions ....................................................... 30

**LECTURE 5**  
The Power of Thought Experiments ........................................................... 40

**LECTURE 6**  
Thinking like Aristotle .................................................................................. 48

**LECTURE 7**  
Ironclad, Airtight Validity ............................................................................ 60

**LECTURE 8**  
Thinking outside the Box ............................................................................. 71

**LECTURE 9**  
The Flow of Argument ............................................................................. 78

**LECTURE 10**  
Simple Heuristics That Make Us Smart ...................................................... 87
<table>
<thead>
<tr>
<th>LECTURE</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Why We Make Misteaks</td>
<td>95</td>
</tr>
<tr>
<td>12</td>
<td>Rational Discussion in a Polarized Context</td>
<td>103</td>
</tr>
<tr>
<td>13</td>
<td>Rhetoric versus Rationality</td>
<td>111</td>
</tr>
<tr>
<td>14</td>
<td>Bogus Arguments and How to Defuse Them</td>
<td>119</td>
</tr>
<tr>
<td>15</td>
<td>The Great Debate</td>
<td>128</td>
</tr>
<tr>
<td>16</td>
<td>Outwitting the Advertiser</td>
<td>134</td>
</tr>
<tr>
<td>17</td>
<td>Putting a Spin on Statistics</td>
<td>141</td>
</tr>
<tr>
<td>18</td>
<td>Poker, Probability, and Everyday Life</td>
<td>148</td>
</tr>
<tr>
<td>19</td>
<td>Decisions, Decisions</td>
<td>156</td>
</tr>
<tr>
<td>20</td>
<td>Thinking Scientifically</td>
<td>165</td>
</tr>
<tr>
<td>21</td>
<td>Put It to the Test—Beautiful Experiments</td>
<td>173</td>
</tr>
<tr>
<td>22</td>
<td>Game Theory and Beyond</td>
<td>183</td>
</tr>
</tbody>
</table>
# Table of Contents

**LECTURE 23**  
Thinking with Models .................................................................................. 192  

**LECTURE 24**  
Lessons from the Great Thinkers .............................................................. 201  

**SUPPLEMENTAL MATERIAL**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glossary</td>
<td>209</td>
</tr>
<tr>
<td>Bibliography</td>
<td>231</td>
</tr>
<tr>
<td>Answers</td>
<td>244</td>
</tr>
</tbody>
</table>

*Answers for some sets of Questions to Consider and Exercises can be found in the “Answers” section at the end of this guidebook.*
Thinking is one of the things we do best. Wouldn’t it be great if we could do it even better? This course, *The Philosopher’s Toolkit*, gives you a set of thinking techniques designed with that goal in mind: tools for conceptual visualization, critical analysis, creative thinking, logical inference, rational decision, real-world testing, effective reasoning, and rational argument. The course uses interactive engagement to introduce a range of conceptual methods and perspectives: mind-stretching philosophical puzzles, mental exercises by example, thought experiments on which to test your powers, and deep questions to ponder.

You’ll learn “hands on” the simple heuristics that make us smart; the basic strategies of decision theory and conceptual modeling; and how to handle everyday probabilities, track a train of thought, outwit advertisers, and detect a spin in statistics. The course also emphasizes rationality in the social context: how to use the wisdom of crowds; how to analyze the flow of thought in a debate; and how to defuse fallacious reasoning, deflate rhetoric, and break through opinion polarization. You can apply your new tools of analysis in evaluating arguments on both sides in a “great debate.” Decision theory, game theory, probability, and experimental design are all introduced as part of the toolkit, but with a philosophical examination of their limitations, as well as their strengths.

In order to think better, we also have to understand why we make mistakes—the systematic ways in which our thinking goes wrong. You’ll be surprised to find how poor we are at noticing continuity changes in both movies and in everyday life; in fact, we are “change blind.” Our memories, however vivid, are created in large part by things that happened after the event: what questions were asked, how they were asked, and how the story was retold. Our estimates of risk and danger often have more to do with the vividness of a mental picture than with any clear-headed consideration of the probabilities. This course draws lessons from a wide range of literature in psychology and
cognitive science in exploring the systematic conceptual biases that mislead us all—and how we can compensate for those biases. It also focuses on questions in philosophy of mind regarding the role of emotion in thought, individual and social rationality, and the comparative powers of gut instinct versus systematic analysis. Is emotion the enemy of rationality or a force we need to better understand as a conceptual resource in its own right?

All of the techniques in the course are outlined with the history of philosophy and science as a background. The power of visualization is introduced using insights in the history of thought from Pythagoras to von Neumann. The power of thought experiments is illustrated with examples from Galileo, from Einstein, and from contemporary philosophy of mind. Important perspectives and approaches are traced to Plato, Schopenhauer, Hobbes, Pascal, and Descartes. Concepts important for analyzing empirical data are drawn from the work of philosophers Karl Popper, John Stuart Mill, and Charles Sanders Peirce but are illustrated using Newton’s experiments in a darkened room, R. A. Fisher’s “Lady Tasting Tea” experiment, and Richard Feynman’s experiment with a glass of ice water during televised hearings on the Challenger disaster. The history behind gambling and probability theory, game theory and the Cold War, and the strange tale of Aristotle’s manuscripts all form part of the story.

The Philosopher’s Toolkit is a perfect introduction to applied philosophy, designed for application in everyday life. The emphasis throughout is on practical conceptual strategies that are useful in any area of application, with interactive examples of logic in action taken from business, the media, and political debate. What are the advertising tricks that we all have to watch out for? What are the standard forms of bogus argument? How can we manage empirical data and statistical information without being misleading and without being misled? How can we have a rational discussion in a polarized environment? From these lectures, you’ll gain a wide set of useful new thinking skills, methods, and techniques but also a deeper psychological, philosophical, and historical appreciation for the thinking skills you already have.
Why do we think the way we do? How can we think better? Those are the questions at the core of this course, examined using interdisciplinary input from cognitive science, psychology, logic, and philosophy of mind. The background is the full sweep of the history of thought. Through your involvement in these lectures, you’ll have a deeper understanding of your own thinking, both good and bad. You’ll also have a “philosopher’s toolkit” of ways to make your thinking more effective, more flexible, more logical, more creative, and both more realistic and more rational.
Thinking is one of the things we do best, but wouldn’t it be great if we could do it even better? *The Philosopher’s Toolkit* is a set of tools designed with that goal in mind: tools for creative conceptualization, critical analysis, logical inference, rational decision making, effective reasoning, and rational argument. With those tools, we can be better thinkers—more creative, logical, inventive, and rational. This course is about thinking in two different but complementary ways. It’s about how we think (a descriptive task) and how we can think better (a normative task). Our aim in the course will be to develop a set of conceptual skills that is useful in all kinds of thinking.

**The Descriptive and the Normative**

- Let’s start with a simple example of the **descriptive** and **normative** tasks in thinking: Imagine we have a bat and a ball. The bat costs $1.00 more than the ball. Together, they cost $1.10. How much does the ball cost? Many people give 10 cents as the answer. That’s the descriptive side of the story.

- But if the ball costs 10 cents, and the bat costs $1.00 more than the ball, then the bat will cost $1.10, and the bat and the ball together will cost $1.20, not $1.10. Even if you saw that immediately, it’s likely that the first figure that popped into your head was 10 cents.

- There are two normative lessons to draw from this example. The first is that the intuitive and immediate answers we jump to are often wrong. The second is that it’s easy to do better. In this case, to do better, we just need to check the first answer. If the total is $1.10, then the ball must cost 5 cents.
Thinking in Patterns

- Among the topics that will come up repeatedly in this course are patterns. It is in our nature to think in patterns. That’s how we make sense of things—we look for the pattern that underlies them.

- A passage about kites illustrates this point. Without any context, the passage seems nonsensical, but once you know the subject of the text, everything falls into place. What you needed was a picture.

- Our aptitude for patterns is remarkable, but that propensity to look for patterns also has a downside. It’s possible to draw connections that aren’t really present.

- You may have experienced a situation similar to this: You take out the trash and suddenly can’t find your car keys. You assume that you must have thrown your car keys away.
  - You have just jumped to a causal connection between two events because they happened close together. The Latin name for this kind of fallacious thinking is post hoc ergo propter hoc (“after it, therefore because of it”).
  - There are many biases built into our thinking. We think in pictures and patterns, and we are extremely liable to be influenced by context.

- Here’s another question to consider: According to the Old Testament, how many animals of each kind did Moses load into the ark? The answer is not two but zero, because it was Noah, not Moses, who loaded animals into the ark. Loading animals into the ark brings up a vivid picture, and we tend to focus on the picture, neglecting the fact that something’s a bit off in the description.

- Much of the effort in advertising and marketing is designed to exploit our conceptual biases. Researchers have found, for example, that a pricing strategy of “four for $2.00,” rather than “50 cents each” leads people to buy more than one item.
• Context can bias your thinking even when you’re not aware of it, perhaps especially when you’re not aware of it.
  ○ In experiments that involved “priming” undergraduates with words associated with the elderly, researchers found that mental context results in physical differences in the undergraduates’ behavior—they walked slightly slower after hearing the words *Florida, gray, bingo*, and so on.
  ○ The reverse was also found to be true: physical context made a difference in mental behavior.

**Working toward Social Rationality**

• Much of this course is about individual **rationality**: developing conceptual tools and recognizing and avoiding conceptual traps. But much of it is also about social rationality. In argument and debate, rationality occurs in a social setting. In public thinking, there are different traps but also different possibilities.

• Argument and debate bring **rhetoric** along with them. The territory of this course includes examples of both good and bad rhetoric—debating techniques and how to detect them, rhetorical fallacies and how to defuse them. It also includes standard tricks used in advertising and how to recognize “spin” in statistics.

• *The Philosopher’s Toolkit* includes normative tips and techniques for working toward enhanced social rationality. Sometimes, we have to dig out of polarization in order to give important topics the open and rational consideration they deserve.
  ○ Negotiations of a test-ban treaty between the United States and the Soviet Union during the Cold War serve as an example. In these negotiations, onsite inspections were a major verification issue, with each side focused on the number of inspections to be allowed each year.
  ○ Both sides went into the negotiations with a number in mind, but the negotiations broke down because neither side would compromise.
○ With the benefit of hindsight, this looks like a case of crippling polarization—of social irrationality. The negotiations would likely have gone better if the conversation could have been shifted toward the real concerns: protecting sovereignty and national security on both sides while minimizing an overkill arms race.

Acquiring the Techniques of Philosophy

• The best philosophy borrows from and builds on diverse areas of thought, including psychology, statistics, mathematics, economics, the history of science, and even physics. The techniques of psychology, in particular, are wonderful for the descriptive task, providing insights into how we think. The question of how we should think typically demands a philosophical approach.

• Philosophy has been concerned with normative evaluation of thinking from its very beginning. In Plato’s dialogues, Socrates encounters various people in the Athenian marketplace and attempts to find out what they think—about justice, or knowledge, or piety—and why.
  ○ Socrates then challenges that reasoning, probing it, testing it, attempting to show inconsistencies and irrationalities. The dialogues are exercises in the normative evaluation of conventional reasoning.
  ○ Interestingly, the context is always social. For both Socrates and Plato, rationality is something we work toward together.

• With Plato’s student Aristotle, logic begins. In Aristotle, it is clear that the philosopher’s task is not merely to figure out how people think but how to think better—more systematically, more validly, more logically. The enterprise of logic is a major thread in the history of philosophy that runs from the pre-Socratics down to the present day.

• In the beginning, with such thinkers as Pythagoras, Euclid, Ptolemy, and others, all intellectual disciplines counted as
philosophy. Other disciplines later split off as well-defined topic areas and developed recognized techniques of their own.

- The remaining core is philosophy, which is, in large part, normative, concerned with how things ought to be.
  - That’s why ethics is a major subfield of philosophy, concerned not merely with how people act but how they should act.
  - That’s why logic is a major thread, concerned not merely with what we think but how we should think.
  - Whenever we want to think better, in general and across all the disciplines, it’s important to return to that normative core.

- The American philosopher Charles Sanders Peirce is thought of as the founder of philosophy of science, though he never held an academic position.
  - In 1877, Peirce wrote an article entitled “The Fixation of Belief,” evaluating different techniques of thinking side by side.
  - Only the technique of science, he claimed, is able to recognize its own fallibility and use it to its own advantage. Only science offers a way of constantly testing and criticizing current beliefs in order to work toward better ones.

- The best philosophy has always been informed by the other disciplines. Indeed, some of the best of 20th-century philosophy
has been done within other disciplines. Economists, physicists, and social scientists all have tools that should be part of the philosopher’s toolkit.

**Slow Down Your Thinking**

- Thinking is something you can only understand by doing. Ours is not just a theoretical exploration of thinking; it is also intended as a practical exploration, with an emphasis on applications. That means that this course is a joint enterprise; we will work through exercises together.

- Skill with any conceptual tool, as with any physical tool, comes with practice. You should attempt to apply the take-home lessons of each lecture in this course to the rest of your thinking outside the course. And don’t be afraid to pause the lectures; thinking often takes time.

- Let’s return to the example we started with—the bat and ball. This example comes from the work of Shane Frederick, a professor of marketing at Yale. It is used as part of a theory of thinking by Daniel Kahneman, a psychologist who received the Nobel Memorial Prize in Economic Sciences in 2002.
  - Kahneman speaks of two systems in our thinking. System 1 is fast, intuitive, good for emergencies—and often wrong. When you jump to conclusions, you’re using system 1.
  - System 2 is slower, more critical, and more systematic; here is where you check your answers.
  - It is very much in the spirit of this course that we need to use both systems in the appropriate sphere. In an emergency, a decision may have to be made fast, and our intuition may reflect a great deal of relevant past experience. But in many cases, what’s required is careful reflection instead. Much of the philosopher’s toolkit falls into that category.
• Kahneman suggests a technique for emphasizing whichever system is required: When what is called for is creative, exploratory, and intuitive thinking, make yourself smile. When what is called for is critical calculation and reflection, furrow your brow and pucker your lips in concentration. Give this approach a try and see if it works for you.

• In the next lecture, we’ll look more closely at system 1 and system 2 thinking, exploring them in terms of rationality and emotion: cool rationality and hot thought.

**Terms to Know**

**descriptive**: Used to designate a claim that merely reports a factual state of affairs rather than evaluating or recommending a course of action. Opposed to normative.

**ethics**: The field of philosophy that focuses on moral issues: ethically good actions, ethically right actions, rights, and obligations.

**logic**: The study of patterns of rational inference and valid argument.

**normative**: Used to designate a claim that is evaluative in nature or recommends a course of action, as opposed to descriptive.

**post hoc ergo propter hoc**: “After it, therefore because of it”; a fallacy based on the claim that because something followed another thing, it must have been because of that other thing. This fallacy overlooks the possibility of coincidental occurrence. Abbreviated as *post hoc*.

**rationality**: Exercising reason, that is, analytical or logical thought, as opposed to emotionality.

**rhetoric**: The skills of effective speaking and presentation of ideas; also, the techniques of persuasion, fair or foul, for either good ends or bad.
1. An important distinction in this lecture is between descriptive and normative approaches. To solidify this distinction, try these two exercises:
   (a) Give a purely descriptive account of when and where you grew up.
   (b) Give an account of when and where you grew up that includes normative elements, as well.

2. Charles Sanders Peirce prizes science because it recognizes its own fallibility. Why is it a “plus” to recognize fallibility? Does that apply to people, too?

3. This lecture included an example of change blindness. Can you offer a hypothesis for why we tend to be change blind?

Exercise

Below is an exercise in pattern recognition in literature, even for those who don’t know anything about literature. Three lines in the following poem are lifted from poet A, three from poet B, three from poet C, and three from poet D. Try to identify the lines that come from the same poet. (See “Answers” section at the end of this guidebook.)
A Grim Amalgamation
by Libby Jacobs

Away! away! for I will fly to thee,
So deep in luve am I;
For somewhere I believe I heard my mate responding to me
When the evening is spread out against the sky
And the rocks melt wi’ the sun.
There will be time, there will be time
To take into the air my quiet breath,
But not altogether still, for then she might not come
immediately to me
Before the taking of a toast and tea.
And fare thee weel, my only luve!
The love in the heart long pent, now loose, now at last
    tumultuously bursting,
Now more than ever seems it rich to die.
Robert Louis Stevenson’s *Dr. Jekyll and Mr. Hyde* illustrates a standard construction in which reason and the emotions are opposing parts of the mind. In this view, reason is at war with the emotions. The ideal of rationality, or Dr. Jekyll, is also exemplified by Mr. Spock of *Star Trek*, devoted to a life of logic and suppression of emotion. In the work of Sigmund Freud, there is also a Mr. Hyde: the id, which Freud calls the “dark, inaccessible” part of our personality. In this lecture, we will talk about what can be said for such a view and explore ways in which it turns out to be incomplete.

**Reason versus Passion**

- The ancient Greek philosophers are often cited as sources for the view of reason at war with the emotions. In *The Republic*, Plato develops a *tripartite theory of the soul*. Its three elements are *nous* (“reason”), *epithumia* (“passion”), and *thumos* (“spirit”). Plato makes it perfectly clear which element he thinks should be in control: reason.

- That classic picture of reason at war with the passions, and too often losing, echoes down the centuries. Erasmus laments that we have been given far more passion than reason; he estimates the ratio as 24 to 1. Shakespeare has Hamlet say, “Give me that man / That is not passion’s slave, and I will wear him / In my heart’s core.”

- Is emotion the enemy of rationality? To at least some extent, the answer is yes. In emotional contexts, such as conditions of stress, rationality can indeed suffer. In a series of tests conducted in the 1960s, soldiers under stress experienced a decrease in cognitive abilities by 10 percent; memory failed by almost a third.
• What holds for stress holds for strong emotion generally: It can negatively impact cognitive capacity, attention, memory—all the things required for rational decision making.

**Reason without Emotion Is Blind**

• There are certainly cases in which emotion runs amuck, swamping rationality. But it doesn’t look like we should simply try to eliminate emotion. Where would we be without it?

• The neuroscientist Antonio Damasio speaks of a patient he calls Elliot who suffered damage to a central portion of his frontal lobes, a region associated with judgment. Elliot aced all the rationality tests; his logical reasoning was intact; and his memory functioned normally, but his life had become chaos because he couldn’t order priorities.

• The diagnosis wasn’t a lack of reason; it was a lack of emotion. The impact of that lack of emotion on personal decision making and planning capacity was devastating. Because everything had the same emotional tone for him, Elliot had no reason to prioritize—no reason to do one thing rather than another.

• Reason without emotion seems to be blind, because emotion is tied to value—what psychologists call “valence.” The things we love are the things we value positively—those that have a strong positive valence. The things we hate or fear are those that have a strong negative valence. Valence is crucial for action; we need it in order to decide what to do and to follow through and take action. Without the sense of value that is embodied in emotion, we become like Elliot.

• **Heuristics** are simple rules of action, immediate and effective in ways that don’t require calculation or deliberation. It has been proposed that this is what emotions are: heuristics that give us the immediate value judgments required for action.
One of the earliest theoretical studies of the emotions was Charles Darwin’s *The Expression of Emotions in Man and Animals*. As the title makes clear, Darwin thought of emotion as something inherited from our evolutionary ancestors.

- Such a view makes sense if emotions are a fast and frugal heuristic for value in action. In dangerous situations, we may not have time to think through all options; we just need to get out of the way of danger. Fear motivates that reaction. Anger gives us a fast and frugal reaction when threatened.

- Without emotions as value-heuristics, we couldn’t act as quickly and decisively as we often need to. What we learn from Elliot is that without emotions we would lose decision making and appropriateness of action entirely.

The 18th-century philosopher David Hume offers an extreme form of the claim that emotions should rule.

- Value, Hume says, comes not from reason but from something else. It is not contrary to reason, Hume says, “to prefer the destruction of the whole world to the scratching of my finger.” That is an issue of value, not an issue of reason.

- Hume turns Plato’s claim completely on its head. It is not reason that should rule, he says, but the passions. “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.”

**Reason or Emotion? Both.**

- What’s the answer? Are emotions the enemy of reason, to be subjugated and suppressed, or should emotions be in control? This question represents a logical fallacy, called a “false dilemma.” A false dilemma relies on a bogus problem set-up with only two options, when in fact, those are not the only options. In this case, we need both reason and emotion—cool rationality and hot thought.

- For many people, at least up to a point, emotional arousal actually improves performance. This is a familiar fact among athletes. That
is what pep talks in the locker room are all about: getting players emotionally pumped and raring to win. Athletes in competition often want performance stress. Beyond a certain point, however, stress becomes overstress. With too much stress, performance declines quickly.

- In the case of both emotion and stress, that pattern of performance is known as the “inverted U.” Performance rises on one side of the upside-down U, stays fairly high for a stretch, but then plummets dramatically. A mild level of emotion and even stress seem necessary for alertness and full involvement. But in conditions of excessive emotion and stress, performance declines.

- Without stress, people tend to rely on analysis of the problem, both wide and deep. They consider many options and think each of them through. We can think of that as the purely analytical approach. Under stress, in contrast, people attend to the general outline of the problem, rather than a wide range of specifics. They focus on significantly fewer options. We can think of that as the intuitive approach.

- We can portray the intuitive approach with either a negative or a positive spin. In a negative light, there is a reduction in the careful consideration of alternative options, in the tracking of options at length, and in the use of working memory and attention to detail. On the positive side, there is an increase in the focus of attention to just the critical issues and elements, avoiding inessentials.

- Recent brain studies at the University of California seem to confirm the conclusion that we need both reason and emotion. Researchers there noted both the analytical and the intuitive in the context of gambling. Brain scans of gamblers revealed that the best players used both the amygdala (associated with emotion) and the prefrontal cortex (associated with control). Rational decision making seems to require both.
• Some decisions demand instantaneous action in an immediate context. For those, we can rely on the heuristics of gut instinct and emotional response. But some decisions—human decisions, at least—aren’t about what to do immediately. Some decisions are about what to do in the next few months, or over the next five years, or in the course of a lifetime.
  ○ For those kinds of decisions, we don’t need to rely on emotion alone; we have every reason to recruit the analytical and reflective abilities of the prefrontal cortex. In those cases, it would be wrong not to consider numerous options or to think through potential consequences in detail and depth.
  ○ But for extremely short-range decisions, it’s a good thing that we have immediate emotional responses.

How Can We Make Better Decisions?
• The fact that decision making can decay in conditions of extreme emotional stress leads to a few simple lessons we can apply in our lives.

• Avoid making decisions—particularly long-range decisions—in extreme emotional situations. Decisions about living wills and funeral arrangements, for example, should be made before the need arises so that the entire brain can be engaged.
- Write up a list of pros and cons, and then balance the pros against the cons.
  - Clearly lay out goals and alternatives, and be careful to explore all alternatives. Expectation of consequences should be based on the best evidence available.
  - Make sure value is clearly included. What is the importance of different goals? Use emotional response as an indicator, but be aware of ways in which short-term emotion might be a bad indicator for long-term goals.

- Monitor your own emotional reactions in the decision-making process. If the “rational” considerations seem to lean one way, are you disappointed? If they lean another way, does that make you anxious? Emotional response in the midst of the attempt at rational decision making is telling you something about what you value. You may discover that you value things in different ways than you thought you did.

**Deferred Gratification**

- A fascinating experiment regarding short-term emotional response and long-term control is the famous “marshmallow test” conducted at Stanford in the late 1960s and early 1970s.

- Individual nursery school children were brought into a room and seated at a table that had a marshmallow on it. The children were offered a choice: Eat the marshmallow right away, or wait for a few minutes while the researcher left the room. If they waited, they would get another marshmallow when the researcher came back and could eat them both. If they wanted to eat the first marshmallow before the researcher got back, they were told to ring a bell. The researcher would come back and they could eat the first marshmallow, but they wouldn’t get a second one.

- Some children ate the single marshmallow immediately. Some restrained themselves by deliberately looking away or by toying with the marshmallow. That kind of restraint indicates a capacity
for **deferred gratification**—for rational control of emotion in decision making.

- The children were tracked through later life, and researchers found that those who couldn’t wait for the second marshmallow seemed to have more behavioral problems later—with maintaining attention, sustaining friendships, and handling stress. In some cases, those who were able to wait scored significantly higher on SATs.

- The lesson here is: In order to increase rationality in our decision making, we should perhaps all learn to wait for that second marshmallow.

### Terms to Know

**amygdala**: A small region deep within the brain that is associated with emotion.

**deferred gratification**: The ability to restrain oneself from taking an immediate payoff in order to obtain a larger payoff later.

**epithumia**: Appetite or passion; according to Plato, the second element of the soul.

**fallacy**: A form of argument in which the premises appear to support a conclusion but, in fact, do not; the term is often used to refer to familiar types of logical mistakes that may be used to trick or mislead.

**heuristics**: Simple guides to action or rules of thumb that allow us to act or make a decision without calculation or deliberation.

**intuitive approach**: An approach to problem solving that focuses on the general outline of a problem, attending to a few critical issues and generating a quick decision after consideration of relatively few alternative solutions. Opposed to an analytic approach, which examines a wide range of specifics and generates an exhaustive list of alternative solutions.
**nous**: Reason; according to Plato, the first element of the soul and that which should rule.

**prefrontal cortex**: The anterior portion of the brain, which lies just behind the forehead, heavily involved in complex planning and decision making.

**thumos**: Spirit or fortitude shown in battle; according to Plato, the third element of the soul.

**tripartite theory of the soul**: Plato’s theory that the soul includes three parts: **nous**, **epithumia**, and **thumos**.

**valence**: A psychological term describing the positive or negative value (often emotional) that we ascribe to things.

**value**: Desirability; psychological research indicates that relative valuations tend to be inconsistent over time and sensitive to context.

---

### Suggested Reading

Damasio, *Descartes’ Error*.

Hume, “Of the Passions.”

Plato, *The Republic*.

Thagard, *Hot Thought*.

---

### Questions to Consider

1. A major theme of this lecture is that emotion is a helpful “fast and frugal heuristic” for some kinds of decision making but detrimental for others.
   (a) From your own experience, give an example in which you think an immediate emotional decision was the right kind of decision to make.
(b) From your own experience, give an example in which you think an immediate emotional decision turned out badly.

2. The Stoics thought rationality should rule and the emotions should be suppressed. Hume claimed that reason should be the slave of the passions—that the emotions should rule and reason should be suppressed. If you had to choose one of those extreme positions, which would you choose?

3. What would life be like if you had no emotion regarding anything at all? What would life be like if all your reactions were purely emotional?

Exercises

The Spanish artist Francisco de Goya (1746–1828) has a famous etching called The Sleep of Reason Brings Forth Monsters (view it at http://www.rijksmuseum.nl/images/aria/rp/z/rp-p-1921-2064.z.). What does the title tell you about Goya’s stance on reason and the emotions?

Watch one of the many replications of the Stanford marshmallow experiment available online (http://www.youtube.com/watch?v=6EjJsPylEOY or http://vimeo.com/7494173). What do you think the children in your family would do in that situation? What do you think you would have done as a child?
The purpose of this course is to give you some conceptual tools for effective problem solving, logical thinking, and rationality in general. This lecture will help you to understand the most important tool of all for logical thinking and effective problem solving: visualization. As we’ll see in this lecture, visualization is a tool you already have; our goal is to increase your respect for that tool and encourage you to put it to work.

What Is Visualization?

- Let’s start with an example for you to try in your head. Imagine you have a long rectangular board. You cut it into two lengths, one of which is three times the length of the other. You now cut the smaller of those lengths again into two pieces, one of which is twice as long as the other. The smallest piece is now 8 inches square. How long was the board you started out with?

- Next, draw the board, following the instructions above. How long was the board? Once you have a picture, the answer may seem pretty obvious. The piece you divided into three must be three times the length of the 8-inch square, so it’s 24 inches long. The original board must be four times that—96 inches, or 8 feet long.

- When you try to do a problem like this in your head, it’s complicated, but it’s not when you can see the problem—even in your mind’s eye. That’s the power of visualization.

Pattern Recognition

- People are much better than computers at pattern recognition. Think of our ability to recognize human faces. Newborns don’t have any trouble recognizing a face as a face. We don’t have any trouble recognizing the face of someone we know. We can sometimes recognize the face of a person we haven’t seen for 20
years, even with a fleeting glance. Computer scientists have been trying to replicate that ability for decades.

- Computers are faster and more accurate at all kinds of symbolic calculation, but when it comes to information processing and pattern recognition, nothing beats the human visual system.

How Philosophy Uses Visualization
- There is a long history of visualization in the philosopher’s toolkit. In *The Republic*, Plato outlines a theory of the relation between the visible and the intelligible world, between mere opinion and genuine knowledge. He maps out the theory visually, with an image known as Plato’s “divided line”—a division between the visible and the intelligible, further divided on each side between superficial understanding and a deeper grasp. One can almost see Plato scratching that line in the Athenian sand.

- That tradition of visualization continues through the history of thought. Another philosopher-mathematician, René Descartes, saw that a link between such equations as $y = 2x^2 + 9x + 3$ and the beautiful arc of a parabola. By treating $x$ and $y$ as coordinates in a plane, algebra became geometrically visual. We still call those Cartesian coordinates. And the conceptual tool used by Descartes was visualization.

The Limits of Cognitive Capacity
- Why does it help to work with pencil and paper rather than entirely in your head? One reason is cognitive capacity. There are just so many things we can hold in conscious attention at one time.

- One of the most cited papers in psychology is a classic from the 1950s by George Miller: “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.” The thesis is that we can only manage about seven pieces of information at one time, plus or minus two. That’s the limit of a phone number, without the area code.
• Not surprisingly, this thesis depends on context, on what the seven items are, and how the information is “chunked.” The truth is that there is no single magic number that measures human cognitive capacity, but Miller makes this point that our cognitive grasp is limited.

• You run up against that capacity when you try to multiply two 3-digit numbers in your head. But if you write the two numbers down and multiply them on paper, you can do such problems easily. One benefit of writing problems out—using visualization—is that it frees up cognitive load.

**Improving Cognitive Capacity**

• Sometimes visualization can practically solve a problem by itself, as we see in the word problem concerning the roads in Brown County. If you’ve been paying attention to the theme of this lecture, you worked that problem out by making a sketch. Sketches can help in two ways.
  ○ First, once you codify each piece of information from a problem in a sketch, you no longer have to try to keep it in short-term memory.
  ○ Second, if the sketch captures all the information in the problem, it practically solves the problem itself.

• You can also use visualization for more abstract problems, such as the one concerning students and their quiz grades. This is really a problem in information management. Here again, visualization is the key, even though the tie between people and grades is more abstract than that between roads between towns.
  ○ This is a common type of problem on standardized tests. It’s known as a “matrix problem” because it is most easily solved with a particular visualization: a rectangular array called a **matrix**.
  ○ The question here is really about how two things align: students and their grades. Visualization turns that abstract alignment into a spatial one: a checkerboard matrix with the grades lined
up along the top and the student names down the side. Spatial relationships are easier to handle than abstract ones—you can see them.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jim</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Billy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tom</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lisa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Visualizing $E = mc^2$**

- Einstein was one of the great visualizers. He said he never thought in symbols or equations but in images, feelings, and even musical architectures. The images, he said, came first. The words came later.

- In thinking through **relativity theory**, Einstein envisaged lightning strikes hitting the ends of a moving train. He asked himself what the timing of the strikes would look like to someone inside the train and outside the train.

- Einstein also employed much more abstract visualizations. He talks of gravity bending light, for example, and speaks of the curvature of space time. It has become common to visualize those ideas in terms of a bowling ball on a rubber sheet.
  - If you think of space as a square grid on a rubber sheet, what happens with light around massive bodies is similar to what the grid would look like if a bowling ball was dropped in the middle.
  - The gravitational force makes straight lines curved—just as light will curve around a massive body.
That image ties in with one of the first and most dramatic tests of some of Einstein’s work. Suppose gravity does bend light; then, the observed position of some distant source of light—such as the stars—should be different when there is and isn’t a massive body, such as the sun, between us and them. In 1919, the British astronomer Arthur Eddington organized expeditions to Africa and Brazil in order to photograph the same portion of the sky before and during an eclipse. He was able to display the variance in the plates as a vindication of Einstein’s vision.

As Einstein himself emphasized: “Discovery is not a work of logical thought, even if the final product is bound in logical form.”

Visualizing the Pythagorean Theorem

- The Pythagorean theorem has been called the single most important result in all of mathematics. And yet Pythagoras’s proof was almost certainly visual, using no math at all. Pythagoras probably conceived of the theorem as follows: “For any right triangle, the square on the hypotenuse is equal to the sum of the squares on the other two sides.” For “square” here, just think of a square.

- The triangle in the middle is a right triangle—any right triangle. The large square on the upper right is the “square on the hypotenuse.” What the theorem says is simply that the area of the square on the upper right is always the same as the sum of the area of the other two squares put together.

- Pythagoras himself saw the importance of this result. It is said that he sacrificed 100 oxen in celebration.
Mental Rotation

- It’s important to remember that your thinking isn’t limited to the space between your ears. There is no reason not to think of thinking as something that can happen using your eyes and with a pencil and paper in front of you. Thinking can indeed be more powerful in that format.

- Mental visualization is very much like using a pencil and paper. There has been a lot of work in psychology done on what is called “mental rotation.” In it, you are given shape silhouettes in two-dimensional form and asked whether it’s possible to convert one into the other by rotating them. For example, which of the shapes on the right are rotations of the one on the left?

![Shapes](image)

- In the three-dimensional version, you’re given three-dimensional shapes and asked whether the shapes are the same but somehow rotated.

- Researchers have found that the time it takes you to discover whether two rotated shapes are the same is directly related to how far you’d have to turn one to get the other. Just as it is faster to rotate something 90 degrees than a full 180 degrees, people are faster at recognizing something as rotated 90 degrees than a full 180 degrees.

- Another interesting result is that some of the same portions of the brain that are used in actual rotation are used in imagined, or mental, rotation. Seeing and imagining recruit some of the same crucial brain circuits.

- The next lecture puts the tools of visualization to work in examining the fundamental elements of all reasoning: the atoms of thought.
Terms to Know

**Cartesian coordinates:** The position of a point on a plane as indicated in units of distance from two fixed perpendicular lines, the $x$ (horizontal) axis and the $y$ (vertical) axis.

**matrix:** A rectangular array; a visualization technique using a checkerboard to represent how two variables align. One set of values is represented by rows in the matrix; a second set of values is represented by columns.

**pattern recognition:** The ability to recognize a set of stimuli arranged in specific configurations or arrays, for example, to recognize faces as faces rather than patches of color or melodies as melodies rather than merely sequences of notes.

**Pythagorean theorem:** In Euclidean geometry, the relation of the three sides of a right triangle, represented by the formula $a^2 + b^2 = c^2$. The area of the square whose side is the hypotenuse of the right triangle (the side opposite the right angle) is equal to the sum of the area of the squares of the other two sides.

**relativity theory:** Proposed as an alternative to Newtonian physics, Einstein’s theory of relativity (general and special) holds that matter and energy are interchangeable, that time moves at a rate relative to one’s rate of speed, and that space itself can be curved by gravity. In large part, relativity theory is a series of deductions from the assumption that some things are not relative. For example, assuming that the speed of light is a constant, Einstein showed that observable simultaneity will be relative to the comparative motion of two observers.

**visualization:** The process of using diagrams or imagery as an aid to thought in representing a problem, calculation, or set of possibilities.

Suggested Reading

Bronowski, *The Ascent of Man*.

Levine, *Effective Problem Solving*.

Polya, *How to Solve It*. 

Lecture 3: The Strategy of Visualization
Questions to Consider

1. The central claim of this lecture is that problems are often easier if we can recruit our visual abilities.
   (a) Give an example of an everyday problem that you already tend to think about visually.

   (b) Are there things that you think about visually and other people do not?

   (c) Take an everyday conceptual problem you are dealing with now. Is there any way you might approach that problem visually?

2. Can you suggest a hypothesis to explain why we are so good at pattern recognition and not particularly good at either large-scale memory or symbolic computation?

3. Do you think it’s true that visualizing something good happening can help make it happen? Can you suggest a hypothesis to explain why that might hold?

Exercise

John has a pretty good weekly salary. In fact, if the salary of his older brother Bill were doubled, John would make only $100 less than Bill. Bill’s salary is $50 more than that of their youngest brother, Phil. John makes $900 a week. How much does Phil make?

(a) First, try to do this problem in your head without a piece of paper in front of you.

(b) Now try it with pencil and paper; draw a horizontal line to represent salaries, with dots for where people are on that line.

(See “Answers” section at the end of this guidebook.)
Two fundamental ideas are at work in this lecture. First is the idea that concepts, the atoms of thought, lie behind our words. Those concepts bring with them their own inherent structural relationships. A simple way of visualizing those relationships is through Venn diagrams: subcategories within larger categories, subsets within larger sets. An expanded visualization is in terms of concept trees: hierarchical organizations of categories and subcategories, sets and subsets. Second is the idea of propositions behind our sentences. These are the molecules of thought—the stuff of evidence and deliberation, the working material of all rational argument. Categorical propositions can also be visualized in Venn diagrams. We’ll use that visualization to illustrate the logic of Aristotle.

What Are Concepts?

- **Concepts** are the most basic elements in all thinking. Concepts are ideas, and words represent those ideas. You can think of concepts as either things in your brain—your concepts—or as something we share as part of our culture—our concepts. But many people think of them as just flotsam and jetsam, floating randomly in a mental or cultural sea.

- The fundamental fact is that concepts do not float around independently. Physical atoms are organized by physical structure—the physical structure represented in the periodic table. As the atoms of thought, concepts come with their own inherent organizational structure, too. Crucial to all effective thinking is an ability to visualize that inherent conceptual structure and put it to use.

Words Are Not Concepts

- Is there a difference between words and concepts? Absolutely. Words are different from language to language, but concepts are not. It’s *frog* in English but *Frosch* in German and *sapo* in Portuguese and *grenouille* in French. These are different words but just one
concept. The recognition that concepts are distinct from words has been part of the philosophical toolkit at least since Aristotle, who said that, although speech is not the same for everyone, “the mental affections themselves, of which these words are the signs … are the same for all mankind.”

- We use words to express concepts, but words are different from the concepts they express. Simple as it is, that point turns out to be crucial in building good reasoning, in detecting bad reasoning, and in reasoning effectively together. Concepts and words are different: a single concept may be expressed by many different words. Even more important, very different concepts may lie behind the same word.

The Extension of a Concept
- Concepts don’t just float around in a random sea; they come with their own relationships—an inherent structure between them. For example, frogs are a kind of vertebrate; but vertebrates are not a kind of frog. Those concepts have a kind of logical order, a relational structure.

- What we’re talking about, then, is really a three-part relationship: the relationship among words, concepts, and the things those concepts apply to. Words express concepts. Concepts apply to things. We might say that words apply to things, too, secondhand. Words express the concepts that apply to things. Philosophers speak of the totality of things that a concept applies to as its extension.

- We can express precisely the same thing in terms of set theory. What are all the things that the concept “frogs” applies to? The set of all frogs. Here’s where visualization comes in. Try to imagine all the frogs, past, present, and future, gathered into one corral. What that corral contains is the extension of the concept “frog”: the set of all frogs. What is the extension of the concept of “idea”? It’s the set of all ideas—past, present, and future. Can you form a real physical corral of those? No, but you can form a conceptual corral; you can gather them in thought.
Connotation and Denotation

- There is a philosophical distinction that is important at this point: a distinction between the connotation and the denotation of a concept. That distinction has its roots in the medieval Scholastics, but it appears with particular prominence in the work of John Stuart Mill.

- Denotation concerns the things that a concept applies to—its extension. That’s different from the connotation of words or concepts—their particular emotional associations.

- Here’s a concept for you: your grandmother’s cookies. The denotation of that concept is its extension—the cookies themselves, whether you liked them or not. But such phrases as “grandmother’s fresh-baked cookies” or “grandmother’s delicious fresh-baked cookies” clearly come loaded with a positive spin. That emotional take, whether positive or negative, is the connotation.

- Denotation is crucial to solid rational argument, reasoning in terms of facts about the things themselves. Connotation is the stuff of emotional persuasion, which we know can be far from rational. For that reason alone it’s important to be able to separate the two.

Concept Trees: A Visualization Tool

- So far we’ve been talking about the organization of concepts and categories two at a time. But the relationships we’re after form much larger structures, as well. These can be visualized in concept trees.

- There is an analogy between concept trees and family trees. If you drew a family tree for the descendants of Davy Crockett, you might have Davy as the trunk, his children as the first level of branches, their children as branches off those, all the way down to the twigs: the living descendants of Davy Crockett. There are hundreds of twigs.
• A concept tree looks exactly like that, except the trunk will be the most general of a set of categories. The first branch will be a first division of subcategories. The next set of branches will be subdivisions of those—all the way down to the twigs.

Propositions: The Idea behind the Words
• Words express concepts. But you can put words together to make sentences. And what sentences express are propositions.

• A proposition is a claim, a statement, or an assertion. It’s what you mean when you say something—not the sentence you utter, not the sounds you make, but the message you’re trying to get across. Propositions are the information packets that we use sentences to express.

• Remember that concepts aren’t just words—they’re something “behind” the words: what the words are used to convey. Propositions aren’t just sentences. They’re something “behind” the sentences. They’re something extra-linguistic that we use language to convey.

• Seeing through sentences to the propositions behind them is hard for some people. What we want to know is whether a sentence asserts something. Consider this sentence: “The Surgeon General says that smoking is bad for your health.” Does that sentence assert that smoking is bad for your health? No, it merely reports what the Surgeon General says.

• Here’s why this is important: In everything we do, it is what is behind the words that matters, not the words themselves. It is the ideas that matter, not the words we use for those ideas. It is the informational content of sentences that matters, not the sentences themselves. When we talk about a solid argument, or a rational hypothesis, or a valid line of thought, it is the propositions involved that are the focus of attention.
Categorical Propositions and Venn Diagrams

- Concepts go together to make propositions in multiple ways, but we can look at some simple cases, as Aristotle did when he invented logic. These simple cases are called **categorical propositions** because they combine two categories. Here are some examples:
  - No frogs are philosophers.
  - All neurologists are MDs.
  - Some laws are unjust.
  - Some inventions are not patentable.

- The first thing to notice is that all four of these propositions link just two categories. Categories can be visualized as the circles of **Venn diagrams**. That visualization technique scales up to diagrams not only of concepts but also of their combinations. It is a diagram in logical space.

- All four of these propositions can be represented with overlapping circles for the two categories involved. Draw two overlapping circles. For the first proposition, mark the left circle “frogs” and the right one “philosophers.” That will give you three internal spaces: an area of overlap, the frogs outside that overlap, and the philosophers outside that overlap. Logicians speak of “logical space.” This is a diagram in logical space.
• A philosopher frog would be in the middle overlap area, a member of both circles. But the proposition tells us that there aren’t any frog philosophers. It tells us that the central area is empty. Black out the overlap, and you have a visualization of how those two concepts go together in that proposition.

[Diagram of overlapping circles labeled 'Frogs' and 'Philosophers']

• Consider the second proposition: “All neurologists are MDs.” Here again, we start with overlapping circles, labeled “Neurologists” and “MDs.” The proposition is that all neurologists are MDs, which tells us that one of the three spaces is empty: There are no neurologists outside the MD corral.

[Diagram of overlapping circles labeled 'Neurologists' and 'MDs']

• The other two propositions give us positive information rather than negative.
With Venn diagrams, we can go from concepts to categorical propositions. Those will come in handy a few lectures down the line. Aristotle was the real inventor of logic, a logic that dominated Western philosophy for 2,000 years. It turns out that we can outline all the logic of Aristotle using the visualization you have at your fingertips.

**Thinking Better: The Pitfalls of Categorization**

- Categorization is an essential tool of thought, but—like other tools—we have to be careful how we use it.

- We group real things in our categories, but our categories are of our own making. We group things together in terms of what we take to be important similarities. But we could be dead wrong about what is really important.

- Categorization can also fail for ethical reasons. Sometimes, it’s just plain wrong to think of people in terms of the groups they belong to rather than as individuals in their own right.
categorical proposition: In Aristotelian logic, a simple proposition that combines two categories using “all are,” “none are,” “some are,” or “some are not.” Categorical propositions can be visualized using two circles in a Venn diagram.

category: Any group of related things; the grouping is based on what are perceived as important similarities between those things.

concepts: Ideas, the basic elements or “atoms” of thought, as distinct from the words that represent those ideas.

concept tree: The hierarchical structure that visualizes the relationships within a set of related concepts.

connotation: The emotional tone or “flavor” associated with the ideas or things that words label.

denotation: The things that a concept or word applies to.

extension: In philosophy, the set or totality of things that a concept applies to.

proposition: A claim, statement, or assertion; the message or meaning behind the words in a written or spoken sentence; the information a sentence expresses or conveys.

rational argument: A way of presenting and supporting claims that relies on logical transition from premises to conclusion.

Scholastics: Philosophers practicing the dominant Western Christian philosophy of the European Middle Ages.

set theory: The branch of mathematics that studies the properties of sets and their interrelations—abstract properties of collections, regardless of what they contain.
**sentence**: A sentence is a series of words, spoken or written, that expresses a claim, statement, or assertion.

**Venn diagram**: A way of visualizing the relations between concepts and their extensions through the use of overlapping circles.

**word**: A linguistic representation of an idea; as distinct from concept.

---

**Suggested Reading**

Aristotle, *The Organon, or Logical Treatises, of Aristotle; On Interpretation; Categories*.

Carroll (Bartley, ed.), *Lewis Carroll’s Symbolic Logic*.

Kelley, *The Art of Reasoning*.

---

**Questions to Consider**

1. In one of the last exercises in the lecture, you were asked to sort these words into three groups:

   breaded interior facades instinct brains fondue

   How did you sort them? Why do you think you sorted them that way?

2. Give an example of conceptual categories that you find useful in your everyday thinking. Give an example of conceptual categories that you think are socially harmful.

3. The contemporary philosopher David Lewis may have been following Plato in the *Phaedrus* in talking of concepts “cutting nature at the joints.” How can we tell whether our concepts are really cutting nature at the joints or whether they’re just perpetrating our own misconceptions of nature’s joints?
Exercise

Paradoxes are often tied to particular categories. Here is one:

Some words apply to themselves. They “are what they say.” “Short” is a short word, for example, “old” is an old word, and “English” is an English word. We’ll call all the words like that “autological.”

Some words do not apply to themselves. They “are not what they say.” “Long” is not a long word, for example, “new” is not a new word, and “German” is not a German word. We’ll call all the words like that “heterological.”

Every word must either apply to itself or not do so. Thus, it looks like all the words divide between the autological and the heterological: the words that apply to themselves and those that don’t.

The paradox arises with this question about the concepts just introduced: Is “heterological” heterological?

If it is, then it looks like “heterological” “is what it says.” But then it’s autological, not heterological.

Thus, it looks like “heterological” must not be heterological. But then it “isn’t what it says.” That’s just what “heterological” means. Thus, if “heterological” is not heterological, it is heterological.

Do you think that this paradox has a solution? If so, what? If not, why not?
We often think of the world of imagination as something different from the world of reality: an escape into a purely mental realm of our own creation. But really understanding what is happening around us—rather than just watching it—often requires thought that extends beyond the real into contemplation of what could be, or ought to be, or even what cannot be. This lecture presents some famous thought experiments as illustrations for more general strategies of effective problem solving. What these strategies have in common is the attempt to solve real problems with something distinctly unreal: the imaginative world of thought experiments.

Going to Extremes

- Here’s a problem-solving example that emphasizes an extension of the visualization strategy: “going to extremes.”

- There are two flagpoles, each 100 feet high. Tied between the tops of those poles is a 180-foot rope that dangles 10 feet off the ground at its lowest point. How far apart are the two flagpoles?

- Consider the extremes. What is the maximum distance the two flagpoles could be apart? The answer is 180 feet, because there is a 180-foot rope between them. At that extreme, the rope would be stretched taught, 100 feet from the ground at every point.

- How about the other extreme: the minimum distance apart? The two flagpoles could be right next to each other. The 180-foot rope would just dangle down, 90 feet on each side, 10 feet off the ground. We’ve found the answer by going to extremes: The two flagpoles have to be right next to each other.
**Generalizations and the Counter-Case**

- What we want in our theories—in both science and philosophy—are **universal generalizations**: natural laws that apply in all circumstances.

- If true, generalizations are tremendously powerful. Their Achilles’ heel, however, is that it takes only one counter-case to refute them. If we say that something holds true for all of a class, it will take just one exception to prove us wrong.

- In a physical experiment, Galileo dropped a single cannonball and small musket ball from the top of the Leaning Tower of Pisa. The physics of the time was Aristotle’s physics, and what Aristotle had said was that heavy objects will fall faster.

- But when Galileo dropped the cannonball and musket ball, they hit at the same time. All it took was a single experiment, one simple case, and Aristotle’s physics was soundly refuted.

- Although Galileo’s was a physical experiment, he also offered a pure thought experiment for the same conclusion without actually dropping cannonballs: If we tie an 8-pound cannonball to a 4-pound cannonball and drop them, what will happen?

- Aristotle says that the 4-pounder will travel slower. It seems that it will hold back the 8-pound cannonball. On this reasoning, the two cannonballs tied together should move slower than the 8-pound cannonball alone. But tied together, the two cannonballs make a 12-pound weight; thus, the two together should move faster than the 8-pound cannonball alone.

- Aristotle can’t have it both ways. Without dropping anything—with just a thought experiment—Galileo put all of Aristotelian physics into doubt.
\[ E = mc^2 \] Thought Experiment

- Next to philosophers, those who love thought experiments most are physicists. Maxwell’s Demon was a thought experiment that played a major role in the history of thermodynamics. Schrödinger’s Cat was a thought experiment central to the debate over quantum mechanics.

- Here’s a thought experiment that was crucial in the development of Einstein’s theory of relativity. Just as he was a master of visualization, Einstein was a master of the thought experiment.

- Suppose a train car is moving along a track, with Mike in the middle of the car. He passes Fred, who is standing on the platform. We assume that the speed of light is a constant. Bolts of lightning hit both ends of the train car. The light from the two bolts reaches Fred at the same time. He sees them hit simultaneously.

- What does Mike see, in the middle of the car? The car is moving toward one lightning bolt, running up to and meeting the light coming from it. Thus, Mike, in the middle of the car, will intersect light from that bolt first. The car is running away from the other bolt, adding distance for the light from it to travel. That light won’t reach Mike until a little later. From Mike’s perspective in the moving train, the bolt at the front of the train hits before the one at the back.

- Einstein’s theory of relativity is, in large part, a series of deductions from the assumption that some things are not relative. What Einstein’s thought experiment shows is that if the speed of light is not relative, observable simultaneity—whether two things are seen as happening at the same time—will be.

- Einstein also followed a firm methodological principle: that theories must be written in terms of what is observable. When you rewrite this thought experiment purely in terms of what is observable, you can see how time and simultaneity will be relative to the
comparative motion of two observers. That is the intuition at the core of Einstein’s theory of special relativity.

Simplify, Simplify, Simplify

- If the problem you face seems resistant to your best attempts, think of a problem that’s like it but simpler. That’s the strategy of analogy.

- Some problems are difficult simply because they have many parts. For those, a strategy of divide and conquer can prove effective. Solve the parts independently, and then put them back together again. In this way, you turn a complex solution for a complex problem into simple solutions for simple ones.

- Both strategies can be illustrated with the classic Tower of Hanoi problem.
  - Imagine that you have three posts in a row. The right and middle posts are empty. On the left post is stacked a series of six disks, ordered from largest at the bottom to smallest at the top. Your task is to end up with all six disks, in the same order as they are now, on either of the two other poles.

  - But there are restrictions. You can move only one disk at a time. You can move a disk only from one pole to another—you can’t just put it to the side, for example, or hold it in your hand. And you can never put a larger disk on top of a smaller one. How would you proceed? How many moves do you think it would take you to get all six disks from one pole to another?

  - First, try visualization; you want to be able to see the problem. Draw a sketch of the three poles and the six disks. It’s even better if you can manipulate a three-dimensional version. All you really need are six disks, each smaller than the one before it. Your three posts can be three spots on a table in front of you.
○ If a complex problem has you blocked, try an analogous problem in a simpler version. What if the problem was about just two stacked disks or four stacked disks?

○ Do you see the pattern? In your solution, the problem with two disks took 3 moves. The three-disk problem took that many moves, plus moving a bigger disk, plus that many moves again: 7 moves. The problem with four disks was really the three-disk solution, plus moving the largest, plus the three-disk solution: 15 moves.

○ Scale that up and you can see the pattern for six disks. The minimum number of steps for five disks is the number of steps to move four, plus a move for the fifth one, plus the number of steps to move the four on top of it again: 31. Six disks will be that number of moves, plus one, plus the number of moves again: 63.

A Thought Experiment in the Ethical Realm

- Not all thought experiments are out to prove something or to find the right answer. Sometimes, thought experiments are explorations in a very different sense. Here is a thought experiment in the ethical realm: the Runaway Trolley.

- A trolley is running out of control down a track. The bad news is that it is hurtling toward five people tied to the track by a mad philosopher. The good news is that you are standing next to a rail switch. All you have to do is pull the lever and the trolley will go down a second track instead. The bad news is that
the mad philosopher has also tied one person to the track before it gets to the divider.

- What do you do? Do you do nothing, allowing the five to die? Or do you pull the switch to save them, thereby sacrificing the one?

- This thought experiment was originally presented by the philosopher Philippa Foot. It was designed merely to point up differences in ethical theory. It turns out that it may also reveal differences in the way different people approach ethical problems. In ethics, utilitarians hold that the right thing to do is that which results in “the greatest good for the greatest number.” They would clearly pull the switch, sacrificing one in order to save five. What would you do?

Terms to Know

**quantum mechanics**: Developed early in the 20th century, quantum mechanics is a sophisticated theory of physics at the subatomic scale; a mathematically elegant, empirically established but philosophically puzzling theory of the very small.

**thermodynamics**: A branch of physics that is concerned with the relationships of heat, pressure, and energy.

**universal generalization**: A statement about “all” of a particular class: “All X’s are Y’s.”

Suggested Reading

Baggini, *The Pig That Wants to Be Eaten*.

Einstein (Harris, trans.), *The World As I See It*.

Sorensen, *Thought Experiments*. 
Questions to Consider

1. What were your answers in the Runaway Trolley case?
   (a) The trolley is rushing toward five victims on one track. By pulling the switch, you can divert it toward one victim on another track. Would you do so? Do you think you should?

   (b) The trolley is rushing toward five victims on one track. By pulling the switch, you can divert it toward one victim on the other track, but that victim is your daughter. Would you pull the switch? Do you think you should?

   (c) The trolley is rushing toward five victims on the track. You can stop it, but only by pushing the fat man next to you onto the tracks. Would you? Should you?

2. Frank Jackson’s black-and-white Mary thought experiment goes like this:
   (a) Mary lives in a black-and-white world, but through her black-and-white reading has learned all that science has to teach regarding color vision.

   (b) One day, she walks into the world of color and suddenly learns something new: what red looks like.

   Jackson concludes that there are facts of phenomenal consciousness—such as what red looks like—that are beyond the reach of science.

   What do you think of Jackson’s thought experiment? Does it prove his conclusion?

3. How can we explain the power of thought experiments? How can it be that thinking about something wildly hypothetical tells us something about the reality around us?
Here is a famous thought experiment regarding perception. It was sent in a letter from William Molyneux to John Locke on July 7, 1688, and is considered in Locke’s *An Essay Concerning Human Understanding*. The question is made more poignant by the fact that Molyneux’s wife had gone blind during the first year of their marriage:

Suppose a person born blind learns to feel the differences between shapes, such as spheres and cubes. If his sight were suddenly restored, would he be able to tell without touching them which were the spheres and which were the cubes?

What do you think?
In this lecture and the next, we will concentrate on the logic of Aristotle. The first attempts to analyze thinking systematically started with Aristotle, around 350 B.C.E. Aristotle’s logic was the gold standard through the Middle Ages and the Renaissance—and continued until about 150 years ago. What we’ve emphasized so far are techniques for logical thinking, with visualization as an important component. Aristotle himself lays out his logic using a visualization—one that takes up concepts and propositions. What we’re working toward is an understanding of the flow of rational argument. Aristotle is a major step in that direction.

A Brief History of Greek Philosophy

- The genealogy of Greek philosophy runs from Socrates, to Plato, to Aristotle. That transition of great minds is also a transition from oral to written culture.

- Socrates wrote nothing down. Indeed, he insisted that writing things down negatively affects our capacity for memory and even for wisdom. Plato came next, transmitting the Socratic tradition to us by making Socrates a character in his dialogues.

- Aristotle studied with Plato for 20 years. He wrote on physics, astronomy, biology, and the soul and is credited with the invention of logic. At the core of his logic—indeed of all logic—is this breathtaking idea: Maybe we can systematize thought.

Aristotle’s Goal: To Systematize Thought

- Aristotle’s theory was that if we could systematize thought, we could perhaps make it less difficult, more accurate, more effective, and faster. Aristotle’s work remained the core of logic all the way through the Middle Ages and the Renaissance. At the beginning of the 19th century, Immanuel Kant said that logic (meaning
Aristotelian logic) was a perfected science, in which no more progress could be made.

- Like all such predictions, Kant’s turned out to be wrong. By the end of the 19th century, we get the beginnings of what we consider contemporary logic, and the 20th century was one of the great centuries in the history of logic.

- The computer age is grounded in logic. The idea of formalizing, systematizing, even mechanizing thought stretches as one unbroken chain from Aristotle to the logic that drives your laptop.

**The Magic of Aristotle’s Visualization**

- A few lectures ago we talked about concepts or categories—the atoms of thought. We used Venn diagrams to visualize how two categories link together into what are called categorical propositions.

- Logic is about structure. Aristotle’s first insight was that the structure of these categorical propositions isn’t tied to their content. Thus, we can forget the Venn diagrams and represent the structures themselves using just an S for the subject term of the proposition—any subject term—and P for the predicate. With that abstraction, we are looking at four basic types of categorical propositions:

No S are P:
All S are P:

Some S are P:

Some S are not P:

- That visualization leads to a second one—Aristotle’s **square of opposition**. At the core of Aristotle’s logic is a picture of the logical relations between those kinds of propositions.
The four types of categorical propositions are on specific corners of the square. Across the top of the square are **universal propositions**: those that are all or nothing. Across the bottom are **particular propositions**: the ones that are about “some.” Running down the left side are the positive propositions; running down the right side are the negative propositions.

But the following is the real magic of Aristotle’s visualization.

First, look at the propositions on the diagonals. Those are **contradictories**. They can’t both be true, and they can’t both be
false. No matter what S and P are, the propositions on the diagonals will contradict each other.

- Now look at the relationship between the top and bottom on the left and the top and bottom on the right. In each case, the top proposition implies the one below. That means that if the top one is true, the bottom one must be true, too.

- Now think of it from the bottom to the top on each side. And think in terms of false rather than true. If the bottom one is false on either side, the top must be false, too.

- Here’s another way to express that relation. True at the top on either side gives us true at the bottom. That’s the direction of implication. False at the bottom on either side gives us false at the top. You can think of that as counterindication.

- Two other aspects of the square are a bit more obscure. They’re called contraries and subcontraries.
  - Consider the propositions across the top. Those can’t both be true: For example, it can’t be true that both all tickets are winning tickets and no tickets are winning tickets. But they could both be false—if some are winning tickets and some aren’t. The propositions across the top are called contraries: both can’t be true, but both could be false.
The propositions across the bottom have exactly the opposite relationship. Both could be true, but they can’t both be false. Those are called subcontraries.

Reasoning Better Through Aristotle’s Logic

- How does all this help in everyday reasoning? If you visualize propositions as located on the square of opposition, you will instantly know the logical relations between them.
  - Consider these two propositions: All bivalves are pelecypods. Some bivalves are not pelecypods.
  - Now, drop them onto the square. They’re diagonals. Therefore, they are contradictories that can’t both be true and can’t both be false.

- We can imagine Aristotle drawing his vision with a stick in the sand of Athens. He was probably amazed at the result, convinced that he was looking at the logical structure of the universe. Laid out in the square of opposition are all the relations between our four types of categorical propositions: universal and particular propositions, positives and negatives, contradictions and implications, and contraries and subcontraries.
Transforming Propositions

- To sum up, what the square of opposition gives us is a visualization for all of these: contradictions, contraries, implications, and counterindications. Those are the logical relations fundamental for rationality, in all the ways that rationality is important to us.

- The visualizations we’ve developed also allow us to understand the logic of certain ways of transforming propositions. We’ll be talking about transforming a proposition into its converse, or its contrapositive. For certain types of propositions, on specific corners of the square, these transforms give us immediate inferences: If the original sentence is true, so is its transform. But it will be important to remember what positions on the square these work for.
  - Starting with the converse, that transform holds for propositions in the upper right and lower left. If one of those is true, it will still be true when we switch the subject and predicate. If it’s true, it will still be true. If it’s false, it will still be false. If we go back to our Venn diagrams, we can see exactly why the converse works for the upper right and lower left.

  - Both diagrams are symmetrical. One says there is something in the overlap area and puts an $x$ there. One says there is nothing in the overlap area and blanks it out. But the overlap is the same whether we think of it as the overlap between $S$ and $P$ or the overlap between $P$ and $S$. It doesn’t matter which one we make the subject term and which we make the predicate term. What the proposition says is the same in each case. They’re equivalent.
• The other forms of proposition—the upper left and lower right—are not symmetrical. The converse is not guaranteed to hold for those. But there is a transform that does hold for those: the contrapositive.

• In order to understand the contrapositive, we have to introduce one more concept: the complement. Let’s take the category “senators.” The complement is the category of everyone who isn’t a senator—that is, non-senators.

• We take the two categories in a proposition and use their complements in order to arrive at the contrapositive. While leaving everything else the same, we switch subject and predicate, but this time, we also replace each by its complement.

• If we’re dealing with the upper-left or lower-right corners of the square, that again gives us an equivalent proposition, an immediate inference.
  ○ All senators are congressmen.

  ○ Switch subject and predicate and replace them with their complements: All non-congressmen are non-senators.

• The inference is immediate and guaranteed. But try this on your friends and you’ll see how difficult it is for people. Next time a friend says something like “All the widgets are defective,” say, “Do you mean that everything that isn’t defective isn’t a widget?” The logical inference is immediate. But you’ll be amazed how long it takes your friend to process it.

• Contrapositive works for the lower right, too: If some liquids are not flammable, some nonflammable things are liquids.

• In summary, then, for upper right and lower left, the converse gives us an immediate inference. We can just switch subject and predicate. Converse isn’t guaranteed for the other two.
• For upper left and lower right, the contrapositive gives us an immediate inference. We can switch subject and predicate and replace each by its complement. Contrapositive isn’t guaranteed for the other two.

• We’ll return to Aristotle in the next lecture, when we take our logic well beyond him.

**Terms to Know**

- **abstraction**: The thought process that allows us to derive general concepts, qualities, or characteristics from specific instances or examples.

- **Aristotelian logic**: Aristotle’s attempt to systematize thought by outlining a set of formal relations between concepts and propositions. These relations can be visualized by his square of opposition and his treatment of arguments as syllogisms.

- **complement**: In logic, given any category, the complement comprises all those things that do not fall in that category. For example, “senators” and “non-senators” are complements.

- **contradiction**: A statement that both asserts and denies some proposition, P, often represented in the form “P and not-P.” If either part of a contradiction is true, the other cannot be true, and thus, a contradiction P and not-P is treated as universally false.

- **contradictories**: The relationship between propositions on the diagonals of Aristotle’s square of opposition. It is a contradiction for both propositions on a diagonal to be true; if one proposition of the diagonal is true, the other must be false.

- **contrapositive**: A way of transforming categorical propositions by switching subject and predicate and replacing each with its complement. For some categorical propositions, the result is an immediate inference: the truth or falsity of the proposition is not altered. The contrapositive transformation preserves equivalence only for propositions in the upper left and lower
right on the square of opposition: the universal positive (“All S are P”) and particular negative (“Some S are not P”).

contraries: The relationship between propositions at the top left (“All S are P”) and right (“No S are P”) of Aristotle’s square of opposition. If two propositions are contraries, it is not possible for both propositions to be true, but it is possible for both propositions to be false.

converse: A way of transforming categorical propositions by switching subject and predicate. For some categorical propositions, the result is an immediate inference: the truth or falsity of the proposition is not altered. The converse preserves equivalence only for propositions in the upper right and lower left on the square of opposition: the universal negative (“No S are P”) and the particular positive (“Some S are P”).

implication: In Aristotelian logic, the relationship moving from the top to the bottom left corners or the top to the bottom right corners of the square of opposition. If the proposition on the top left corner is true, then the proposition on the bottom left corner is also true; if the proposition on the top right corner is true, then the proposition on the bottom right corner is also true. Expressed as “if all S are P, then some S are P” for the left side of the square of opposition and “if no S are P, then some S are not P” for the right side of the square of opposition.

inference: In logic, the derivation of a conclusion from information contained in the premises.

particular proposition: In logic, a proposition about “some” rather than “all”: “Some S are P” (a particular positive, e.g., some cleaning products are poisons) or “Some S are not P” (a particular negative, e.g., some cleaning products are not poisons). The particular positive occupies the lower-left corner of Aristotle’s square of opposition; the particular negative occupies the lower-right corner.

square of opposition: Aristotle’s visualization of the logical relations between categorical propositions.
**subcontraries**: The relationship between propositions at the bottom left (“Some S are P”) and right (“Some S are not P”) of Aristotle’s square of opposition. If two propositions are subcontraries, it is possible for both propositions to be true, but it is not possible for both propositions to be false.

**universal proposition**: In logic, a universal proposition refers to a claim either in the form “All S are P” (universal affirmative, e.g., all snakes are reptiles) or in the form “No S are P” (universal negative, e.g., no snakes are reptiles). The universal affirmative occupies the upper-left corner of Aristotle’s square of opposition; the universal negative occupies the upper right.

### Suggested Reading

Aristotle (Ackrill, trans.), *Categories and De Interpretatione*.

Groarke, “Aristotle’s Logic.”

### Questions to Consider

1. Are there ideas you think you could have come up with if you had been present at the right time and place? Are there ideas you think wouldn’t have thought of, even if you had been present at the right time and place?

2. Why do you think it is that Aristotle’s “exoteric” works—popular pieces for a wide popular audience—have disappeared completely, while his “esoteric” works—lectures for a small body of students—have survived?

3. One of the claims in this lecture is that the attempt to “systematize, formalize, and mechanize thought” is a goal that runs all the way from Aristotle’s logic to contemporary computers. If that is true, what do you think will be the next step in the trajectory?
Exercise

Below are elements of a square of opposition that are not in the Aristotelian positions:

- All philosophers are idealists.
- Some philosophers are not idealists.
- No philosophers are idealists.
- Some philosophers are idealists.

Fill in the Venn diagrams appropriately for each categorical proposition.

Draw solid lines between contradictories

Draw dashed lines between contraries

Draw dotted lines between subcontraries

(See “Answers” section at the end of this guidebook.)
In the last lecture, we were introduced to the logic of Aristotle in terms of a central visualization: the square of opposition. This lecture will cover what makes an argument valid and what defines ironclad, airtight validity. Here is how an argument is valid in that strongest sense: If the premises are true, the conclusion absolutely must be true. If the premises are true, it’s impossible for the conclusion to be false. We can’t get a much tighter connection than that.

What Is a Valid Argument?

• Propositions don’t just stand alone; they go together to form arguments. In arguments, we see the dynamics of thought. In the philosophical sense, an argument is a connected series of statements intended to establish a proposition. An argument in our sense represents the conceptual dynamic from proposition, to proposition, to conclusion. It is that movement of thought that is central to all rationality.

• But of course, it’s not just the intended transition that is important. We want to be able to evaluate arguments. When does a set of reasons give us a solid inference? When does evidence really support a hypothesis? When do premises really entail a conclusion? What guarantees that a transition of thought is a rational transition? Those are questions about validity.

• That’s what Aristotle was after. That’s what all of logic is after. The whole purpose of logic is to capture validity: the crucial category of argument evaluation. Perhaps we can systematize thought by capturing what makes it rational. Perhaps we can formalize validity.

Ironclad, Airtight Validity

• For an argument to be valid: If the premises are true, the conclusion must be true. That is important information, but it’s conditional
information. What we know is that the conclusion must be true if the premises are true. Following is an example; determine whether it is valid or invalid:

○ All purple vegetables are poisonous.

○ Carrots are a purple vegetable.

○ Conclusion: Carrots are poisonous.

- The conclusion is perfectly ridiculous, but the argument is also perfectly valid. Not all purple vegetables are poisonous. But suppose they were. Carrots aren’t a purple vegetable. But suppose they were. If those two premises were both true—if carrots were purple vegetables and all purple vegetables were poisonous—it would also have to be true that carrots were poisonous.

**Validity Depends on Structure**

- Suppose we have an argument in which all premises and the conclusion are true. Is that a valid argument? The answer is: not necessarily. It’s the connection that matters.

- We might have a valid argument in which all premises and the conclusion are true. That would be great, but validity alone doesn’t tell us that. We might have a valid argument in which all premises and the conclusion are false. The “purple carrots” argument above was one of those. We might, in fact, have a valid argument in which the premises are false but the conclusion is true. For example:

  ○ Everything poisonous is orange.

  ○ Carrots are poisonous.

  ○ Therefore, carrots are orange.

- That is a valid argument. If the premises are true, the conclusion must be true. Here, the premises are false, but they entail a conclusion that is true anyway. Sometimes, two wrongs do make a right.
There is only one pattern of truth and falsity that validity rules out. We will never find a valid argument that goes from true premises to a false conclusion. A valid argument is one in which, if the premises are true, the conclusion must be true.

Syllogisms
Aristotle worked out the structure of ironclad, airtight validity in terms of three-step arguments called syllogisms. A syllogism takes us from categories, to propositions, to steps in an argument. It uses three categories in all: A, B, and C. It links those categories in pairs to make categorical propositions: “All A’s are B’s”; “Some B’s are not C’s.” The syllogism itself is an argument in three propositional steps, such as the following:

○ All A’s are B’s.

○ Some B’s are not C’s.

○ No C’s are A’s.

Some syllogisms are valid. Some are not. The one above is not a valid syllogism.

Using Visualization to Determine Validity
Visualization is the key to sorting the valid syllogisms from the invalid ones. We can do all of Aristotle’s logic with the Venn diagrams we already have at our disposal.

All S are P:
No S are P:

Some S are P:

Some S are not P:

- Those are the propositional elements of syllogistic logic. But syllogisms involve three propositions. Each proposition of the syllogism—each step in the argument—uses just two of our three categories. In order to use Venn diagrams to track validity, we need to overlap three circles into a triangular Venn diagram.
  - Our first proposition, linking categories A and B, will look something like this:
○ Our second, linking B and C, will look like this:

○ Our third, linking A and C, will look like this:

○ All three categories, and all three propositions, come together in the syllogism. We put them together like this:
• The result is an overlap of three categories, just as the syllogism always involves three categories. The triangle is formed from three categorical propositions: pairs of linked circles.

**How to Visualize Logic: An Example**

• Following is an easy syllogism that is also a valid one:
  ○ All proteins are polypeptides.

  ○ All polypeptides function by elaborate folding.

  ○ Therefore, all proteins function by elaborate folding.

• Here, we have three categories: proteins, polypeptides, and “function by elaborate folding.” We label the three circles in the diagram accordingly:

![Venn diagram example](image)

• We now enter each of our premises in Venn diagram style. First, all proteins are polypeptides.

![Updated Venn diagram](image)
• All polypeptides function by elaborate folding.

• And the last step—the crucial step in telling whether a syllogism is valid: We want to look at our third proposition—the conclusion—and ask whether its information is already in the diagram we’ve constructed for our premises. If it is, the argument is valid. If it’s not, the argument is invalid. If the concluding information is already provided by the premises, it has to be true when they are, and it’s valid. If the premises don’t give us that information, the conclusion doesn’t follow from the premises, and it’s invalid.

• In order to test for validity, we want to know whether the concluding information is already there in the logical space of the premises. As you look at the diagram, ask yourself what would it mean to draw in the conclusion: “All proteins function by elaborate folding.” That would mean looking at just “proteins” and “elaborate folding”—those two circles on the left.

• For it to be true that “All proteins function by elaborate folding,” the “proteins” area outside of the overlap would be blacked out. You will see from the diagram that that area is already blacked out. There aren’t any proteins in our diagram that aren’t already squeezed into some part of the “function by elaborate folding” area.

• That means that the information in step 3—the information in the conclusion—was already there in the premises. That’s the sign of validity. If the premises are true, the conclusion must be true. The
information in the premises is sufficient to give us the information in the conclusion. The argument is valid.

The Syllogism: No New Information
• A valid argument in the deductive sense is one that squeezes out in the conclusion some information that you already had. Immanuel Kant speaks of an “analytic” proposition as one in which the predicate is already “contained” in the subject term. A valid deductive argument is one in which the information in the conclusion is already contained in the premises.

• It therefore makes sense that we can show an argument to be valid or invalid depending on whether the information in the conclusion is already there in the premises. But it also means that a syllogism will never give us any really new information.

The Three-Circle Technique Embodies Ironclad, Airtight Validity
• We can do all of Aristotle’s logic, all the logic of the medieval philosophers, all the logic that we’ve had for 2,000 years, with those three circles. The three-circle technique embodies the concept of ironclad, airtight validity. With that visualization, we can test validity for any three-step syllogism. However, the concept of ironclad, airtight validity sets a standard for inference in which all the information in a conclusion must already be contained, somehow, in the premises.

• For 2,000 years, ironclad, airtight validity was held up as the gold standard for all knowledge. But that meant all information had to flow from above—from something you already knew. Mathematics is about the only discipline that works that way. It’s no surprise that axiomatic geometry was the standard for all knowledge.

Getting to Reasoning That Offers New Information
• Aristotle’s legacy, therefore, had a downside, too. It was only with the rise of empirical science—in such figures as Galileo—that an alternative became clear. Perhaps an inference from observed
cases to universal laws is what was needed, even if it wouldn’t be ironclad, airtight, guaranteed in Aristotle’s sense.

- If we weaken the demands on validity, we get the possibility of rational information expansion for the first time—reasoning, together with observation, that offers new information.

- Ironclad won’t give us new information, and anything that gives us new information won’t be ironclad. We have bought the benefits of science by weakening our demands on validity, by giving up Aristotelian certainty. We’ll start talking about reasoning and validity in that wider sense a few lectures down the line.

- In the next lecture, however, we’ll take a break for something equally important but a little more fun: a workshop on creative thinking.

**Terms to Know**

**axiomatic**: Organized in the form of axioms and derivations from them. Euclidean geometry is an example of an axiomatic system.

**conclusion**: The endpoint of an argument; in a logical argument, the claim to which the reasoning flows is the conclusion.

**premise(s)**: The proposition(s) or claims that are given as support for a conclusion; in a rational argument, the reasoning flows from the premises to the conclusions.

**syllogism**: An argument using three categories (A, B, and C) that are linked in pairs to make categorical propositions (e.g., “All A’s are B’s,” or “Some B’s are not C’s”), which are then combined into a three-step argument. Some syllogisms are valid and some are not.

**validity**: An argument is valid if the conclusion follows from the premises, if the premises offer sufficient logical or evidential support for the conclusion.
An argument is deductively valid if it is impossible for all premises to be true and the conclusion to be false.

**Suggested Reading**

Kelley, *The Art of Reasoning*.

**Questions to Consider**

1. Aristotle sometimes seems to suggest that *all* reasoning can be analyzed in terms of syllogisms. Can you think of a counterexample?

2. Aristotle’s notion of ironclad, airtight validity is that it if the premises are true, it is absolutely *impossible* for the conclusion to be false. This lecture makes the point that that notion of validity was the gold standard for inference for 2,000 years.
   (a) In what ways do you think that was a positive force in the history of thought?

   (b) In what ways might it not have been so positive?

3. Knowing whether an argument is valid demands knowing whether the premises and conclusion are *actually* true. It demands knowing something about *possibility*: Could the premises be true and the conclusion be false? If we live in the actual world, how can we know things about mere possibility?

**Exercises**

More on validity and truth:
   (a) If we know that the premises of an argument are true and its conclusion is true, what do we know about validity?

   (b) If we know that the premises of an argument are false and the conclusion is false, what do we know about validity?
(c) If we know that the premises are false and the conclusion is true, what do we know about validity?

(d) If we know that the premises are true and the conclusion is false, what do we know about validity?

More on Aristotelian syllogisms:

Determine validity or invalidity for each of these using Venn diagrams:

1. No M are P. All Q are P. No M are Q.
2. All M are P. All Q are P. Some M are not Q.
3. All M are P. Some P are Q. Some M are Q.

(See “Answers” section at the end of this guidebook.)
Thinking outside the Box
Lecture 8

In this lecture, we’ll take a break from Aristotle—a break from visualizing argument structure—to talk about what may be the most important kind of thinking of all and, perhaps, the most fun: creative thinking…lateral thinking…thinking outside the box. Here, we’ll discuss a number of problems or puzzles and how they can be solved through creative thinking.

The Phenomenon of “Mental Set” or “Expectancy”

- A cowboy rides into town on Tuesday. He stays in town for exactly three days. On the first day, he works in the general store. On the second day, he works in the livery stable. On the third day, he hangs out in the sheriff’s office. He leaves promptly the next day and rides out of town on Tuesday.

- How can that be? The answer is that his horse is named Tuesday. He rides into town on Tuesday, stays three days, and rides out again on Tuesday.

- This story illustrates the phenomenon of mental set, or expectancy. You are led into thinking about things in a certain way. You aren’t able to change perspectives quickly enough to solve the puzzle. The whole setup is intended to get you to think in terms of days—one day in the general store, one day in the livery stable—blinding you to the fact that Tuesday could also be the name of a horse.

- Mental set or expectancy is a well-established psychological phenomena. We learn to anticipate particular patterns or continuations of patterns. We get stuck in a pattern rut.

Breaking Out of the Pattern Rut

- Creative thinking demands the ability to break out of the mental set that a problem brings with it. That mental set is the box. Creative thinking is thinking outside the box.
• Often, real-life problems, when they can be solved at all, can be solved only by looking at things in an entirely different way. The same holds for life’s real opportunities. Opportunities may be invisible unless you can look at a situation differently—sideways—until you can think outside the box. Real problems and real opportunities demand more than standardized thinking. For real life, we often need some real creativity.

• Here’s an example based on real events: A woman is driving across the desert, without a cell phone, and has a flat tire in the pouring rain. She manages to jack up the car and takes off the five nuts on the wheel. Just then, a flash flood washes the nuts away beyond recovery deep in the sand. How can she get the car back on the road?
  ○ Here’s a creative solution: She has lost the five nuts from one wheel but still has five nuts on each of the other three.
  ○ Borrow a nut from each of the others, use those three nuts on this wheel, and she can be on her way.

• In another case, some friends are swimming together in a pond. The area is residential; the pond is ringed with lawns and gardens, but it is deep and has a tangle of water plants at the bottom. One of the friends becomes entangled in the pond plants, underwater, just a foot from the surface of the pool. He is struggling desperately, but his friends realize they won’t be able to free him in time.
  ○ What else can the friends do?
  ○ The story is that they grabbed a garden hose from a yard nearby, cut a two-foot segment with a pair of garden shears, and gave him a makeshift snorkel to use until they could get help to work him free.

Cultivating Creative Thinking: Change Your Expectations
• Creative thinking can’t be taught with a template, but it can be cultivated. You can learn to broaden the horizon of possibilities you consider. You can learn to first recognize what the box is and then deliberately look for a way to think outside it.
• For example, take 12 toothpicks (or pencils) and arrange them into four squares. Here’s the challenge: By removing only two toothpicks and without moving any other toothpicks, create not four squares but only two. They must be squares, not rectangles. And there should be only two squares: no floating toothpicks that aren’t part of any square.

○ Here’s the solution: Remove two neighboring toothpicks from the original cross. If you do that, you’ll be looking at one large square with a smaller one tucked into one corner. By removing two toothpicks and moving none of the others, you end up with two squares, no leftovers.

○ Of course, those squares aren’t the same size, and they’re not two of the original four. If there was a mental set imposed by the problem, it was the expectation that the two remaining squares would be two of the four you started with. That was the mold you had to break in order to think about the problem in a different way.

• Here’s another exercise: Use a pencil to connect all nine dots below by drawing four straight lines, without lifting your pencil from the paper. Draw a line, then start the next one from where that line stopped, and so on, for a total of just four straight lines. Those four lines must go through the middle of all nine dots.

○ The key here is that your lines don’t all have to start and stop at a dot.

○ In order to solve the problem, your lines will have to extend beyond the box of nine and turn some corners outside of it.
Cultivating Creative Thinking: Break Habits of Thought

- Here’s another problem: Arrange 10 pennies in such a way so that you have five straight lines of pennies—five rows—each of which has exactly four pennies in the row. The solution is to draw a five-pointed star and place a penny at each of the five points and each of the five inside corners. When you do that, you’ll find that you have five lines, each with four pennies, for a total of 10.

- If you didn’t solve that one, what mental set got in the way? Was it that you expected the lines to be parallel? Was it that you expected them to cross only at right angles? If you analyze your own thinking on problems such as these, you’ll get a feeling for how you can step outside the box.

- What we’re doing is trying to develop an ability to think beyond the expected...to look for options that aren’t anticipated in the question or the situation. We’re trying to develop an ability to think creatively and perhaps even develop a habit of breaking habits of thought.

Cultivating Creative Thinking: Overcome Conceptual Limitations

- For the last problem, work side by side with a friend: You have 12 toothpicks. Give six to your friend. Lay three of those toothpicks down in front of you to form a triangle. Now, by adding just three more toothpicks, form four triangles, of precisely the shape and size as the one you just made.

- How do you get four triangles with just three more toothpicks? Here, creative thinking takes you literally into another dimension. On top of your original two-dimensional triangle, form a three-dimensional pyramid with your three additional toothpicks. There you have it: four triangles of precisely the shape and size as the original. If you want it to stand on its own, of course, you’ll need to stick the toothpicks together. (Try miniature marshmallows.)

- In order to get that problem, you had to move conceptually beyond two dimensions and into three. There is a wonderfully creative book on that theme, written in 1884 by Edwin Abbott, called Flatland:
A Romance of Many Dimensions. Abbott’s narrator is “A Square,” who lives in a two-dimensional plane, unable to conceive or understand what happens when things enter that two-dimensional world from a three-dimensional outside. It’s a charming story in its own right and a wonderful metaphor for the kinds of conceptual limitations we inevitably carry with us.

Thinking Better

- We’ve been talking all along about thinking outside the box. That’s a metaphor, of course, but some recent experiments indicate that it may be a particularly apt metaphor. One hundred undergraduates were given a word task designed to measure innovative thinking. Half of them had to do the task inside a five-foot-square box made of plastic pipe and cardboard. Half did the task outside and next to the box. Interestingly, those who were doing their thinking “outside the box” came up with 20 percent more creative solutions.

- The same kinds of differences showed up between students who had to walk a particular path as they worked on a problem, as opposed to those who could wander at will. Those who could “leave the beaten path” came up with 25 percent more original ideas.

- In the first lecture, we emphasized the idea that physical context can make a mental difference, and these experiments illustrate that point. Sometimes, creativity demands that you really do get away from it all.

- One final creative thinking puzzle: It is a dark and stormy night, and you’re driving in your sports car, which has only two seats. Suddenly, by the side of the road, you see three people stranded at a bus stop.
  - One is a stranger who is having a heart attack at that very moment. Another is a childhood friend who has often saved your life and has long been begging to ride in your sports car. The third person is the man or woman of your dreams—your soul mate—whom you may never see again. You have just one empty seat in your car. Who do you pick up?
The solution requires you to stop thinking that you have to pick up just one person and that only you can be behind the wheel. Ask your friend to drive the heart attack victim to the nearest hospital, then wrap your coat around the person of your dreams.

**Term to Know**

**mental set**: In perception, the background expectation that may influence what is perceived; that is, when one is expecting a normal deck of cards, mental set may lead one to ignore altered cards, such as a red ace of spades.

**Suggested Reading**


Burger, *Sphereland: A Fantasy About Curved Spaces and an Expanding Universe*.

Weston, *Creativity for Critical Thinkers*.

**Questions to Consider**

1. Below are a few more easy “lateral thinking” problems.
   
   (a) 1985 pennies are worth almost $20.00. Why is that?

   (b) Is it legal for a man in Tennessee to marry his widow’s sister?

   (c) A clerk in the butcher shop is 5 feet, 10 inches tall. What does he weigh?

   (d) Can you rearrange these letters to make one word? new door

   (e) A woman has two sons, born at the same minute of the same hour of the same day of the same year. But they are not twins. How can that be?

   (See “Answers” section at the end of this guidebook.)
2. Are there aspects of creativity that you think can be taught? Are there aspects you think cannot be taught?

**Exercise**

Write down 25 things you can do with a bucket of water. Time yourself.
- Can’t come up with 25? Put the exercise aside and come back to it another time. Did you include water as a cleaning fluid? As something to put fires out with? Did you consider its weight? Did you think of using the bucket by itself?

- Did you come up with 25 in less than 4 minutes? A very respectable showing.

- Less than 3 minutes? Outstanding.

- Less than 2 minutes? Phenomenal.

Now can you come up with 40 things to do with two paper bags full of sand?
In a previous lecture, we used Aristotle’s syllogisms to emphasize the central concept of validity. Visualizing syllogisms in terms of three-circle Venn diagrams gave us a picture of validity in the strongest Aristotelian sense: airtight, ironclad validity. In this lecture, we will go beyond Aristotle to look at validity much more broadly. In general, validity refers to the degree of support between premises and conclusion. Does this fact support this claim? Do these premises render the conclusion highly plausible? Does this evidence give us good reason to believe the conclusion? The reasoning may not be airtight, but is it solid enough to act upon?

Beyond the Syllogism

- Here’s an example of the kind of argument we have to deal with every day. You’ll notice that it’s far more complicated than a syllogism:
  1. Spending $50 billion in annual economic aid to foreign countries would be justified only if we knew that either they would genuinely benefit from the exchange or we would.
  2. We don’t know that we would benefit from the exchange.
  3. It might further weaken our already vulnerable economy.
  4. We don’t know that they would genuinely benefit.
  5. The track record of foreign aid that has been misdirected, misappropriated, or lost in foreign corruption is all too familiar.
  6. With $50 billion a year, we could offer college scholarships to every high school graduate and major health benefits for every man, woman, and child in the United States.
7. Our obligations are first and foremost to our fellow citizens.

8. We should spend the $50 billion here rather than in foreign aid.

- There are lots of premises in that argument, far more than in Aristotle’s syllogisms, and lots of transition steps. Together, they are intended to drive us toward the conclusion. But how good is the argument? And how should we analyze an argument such as this? Trying to deconstruct it into syllogisms is nearly impossible. We will see how to analyze it later in this lecture.

**Flow Diagrams**

- The best way to analyze complex arguments is with a simple visualization: a **flow diagram**. Such a diagram will help us see the validity of a complex argument. Breaks, disconnects, and weak logical links can show us invalidity in an argument.

- The basic rule is simple: When one claim is intended to support another, we put an arrow from the first to the second. The problem, however, is that propositions don’t come labeled as premises or conclusions. They can function in either role. It all depends on the context, on the structure of the argument. Another difficulty is that the first sentence you hear doesn’t always come first logically.

- Consider this argument, for example: (1) If the governor is impeached, we might be no better off. (2) Impeaching the governor would require another election. (3) But there is always the chance that people would then vote for someone equally corrupt.

- The propositions are numbered in the order of presentation, but what we want to know is something different. We want to know where the reasoning flows from and where it flows to. It will help if we can identify the conclusion. Which is the conclusion: (1), (2), or (3)?
• It’s proposition (1) that is the conclusion, right at the beginning. Everything else is offered in support of that conclusion. The logic of the argument starts from (2), which leads to (3). And that leads to the conclusion: If the governor is impeached, we might be no better off.

\[2 \rightarrow 3 \rightarrow 1\]

Branching Flow Diagrams

• Of course, arguments get more complicated than that, so we need to add complications in the flow diagrams. First of all, arrows can branch. A set of propositions can lead to multiple conclusions or parallel conclusions.
  ○ Think about how to graph the following: (1) We can get only so much money from taxes; taxation resources have to be balanced among different social needs. (2) Taxation for prisons must, therefore, be balanced against taxation for education. (3) If we build more prisons, we’ll have less for education. (4) If we spend more on education, we’ll have less for the prisons we may need.

  ○ That argument has branching conclusions. The first proposition leads directly to the second. From (1) we graph an arrow to (2), but at that point, our arrows branch.

• Further, just as arrows can branch out, they can branch in. Sometimes several propositions lead to the same place.
  ○ Consider this example: (1) We are dangerously reliant on foreign energy sources. (2) Our oil comes primarily from the Middle East. (3) Most of our natural gas does, as well. (4) Even the elements in our batteries come from such places as Zambia, Nairobi, and China.
First, find the conclusion. The conclusion is the first sentence. Each of the other propositions is an argument for that conclusion in its own right. We graph it by having arrows converge on a single conclusion.

![Argument Diagram](image)

**Independent and Dependent Reasons**

- **Independent reasons** function independently. But sometimes reasons have to function together in order to lead to a conclusion. **Dependent reasons** only function together. In a case where all three propositions work together as dependent reasons, we can mark them like this:

![Dependent Reasons Diagram](image)

- How do we know whether propositions are working independently or dependently toward a conclusion? The answer is argument **stress testing**. If we have independent reasons and one of them fails, the argument should still go through. If we knock out a reason and the argument is still standing, it must be an independent reason. However, that won’t hold for dependent reasons.

**Graphing a Complex Argument**

- We now have the basic elements of any argument flow diagram. But when we start to graph real arguments, we can see how those elements can combine into an immense variety of structures.

- Consider the argument we started with above, about spending $50 billion in foreign aid or for college scholarships here. The conclusion is (8). How are the rest of the propositions meant to support that conclusion? Here’s a sketch of that argument.
The argument uses two dependent clusters, functioning independently of each other. One cluster uses (1), (2), and (4): Together, they offer an independent argument. If a stress test showed that the information about the college education in (6) and (7) was false, then (1), (2), and (3) together would still stand as an independent argument for the conclusion. Propositions (3) and (5) are backups for (2) and (4). All those together make one argument for the conclusion. Another argument comes from (6) and (7) working together.

Data and Warrants

- We can expand on flow diagrams for arguments using the work of the philosopher Stephen Toulmin. We have talked about premises that function together, representing them as numbers with a plus sign between them.

- What Toulmin points out is that there is often more structure just beneath that plus sign. Some of those numbers may stand for premises that function as what he calls data. Some function instead as what he calls warrants. They function together but in a very specific way.

- Any argument starts with some kind of data. But an argument often doesn’t just go from data to conclusion. It has a middle step—a step that says how the data are supposed to lead to the conclusion. Thus,
instead of a plus sign between premises (1) and (2), it might be better to represent them like this:

- Here’s an example: The CT scan shows a shadow on the lungs. When there is that kind of shadow, cancer is a real possibility and further tests are in order. We should do a biopsy. The fact that the CT scan shows a shadow on the lungs is the data in this case. The conclusion is that we should do a biopsy.

- The warrant is the general principle that takes us from the data to the conclusion. It is not so much an additional piece of information as an inference: Given these data, we should draw this conclusion. In this case, the warrant is “When there is that kind of shadow, cancer is a real possibility and further tests are in order.”

Different Kinds of Warrants

- Just as there are different kinds of arguments—scientific, legal, ethical—there are different kinds of warrants appropriate for those arguments.

- For some arguments—in scientific contexts, for example—a probability warrant is appropriate, which is a strong inductive or empirical support. Sometimes, the argument is over the use of a term; in that case, the warrant may be a definition. Sometimes, a legal principle or an ethical rule may serve as a warrant.

- Graphing arguments in terms of data and warrant can help in analyzing both individual arguments and the structure of debates. We all know that there are often two sides to an issue. The distinction between data and warrant lets us see that there are two different ways those sides may be in conflict.
Thinking Better

- Here’s a real challenge: Clip out a newspaper editorial. Isolate just one or two paragraphs and graph out the argument. If you do that, you’ll discover a number of things:
  - Just as we said, the conclusion doesn’t always come last. Sometimes, it takes real work to figure out what the conclusion is supposed to be.
  - You will find that the flow of reasons can be very complex. Often, two claims support a third, which is a subconclusion. Much like a subtotal, it is an intermediate conclusion, just part of the way toward a final conclusion.
  - You will find that some claims in an editorial are very important. Some will be minor. Some will be widely accepted. Some will be accepted only in very specific belief structures.
  - The most important thing you’ll notice is how much goes unsaid in normal arguments. What goes unsaid is often the most important part. An argument is often most vulnerable at the points at which major assumptions are made but are not made explicit.
  - In the next lecture, we’ll move to aspects of our reasoning that don’t need graphing: the conceptual heuristics that make us smart.

Terms to Know

**dependent reasons**: Premises that support the conclusion only when they are both present; propositions or claims that function together but are insufficient alone as support for the conclusion.

**flow diagram**: A systematic sketch of a train of thought illustrating the lines of support between premises and conclusions in a rational argument; when one claim is intended as support for a second claim, an arrow is drawn from the first to the second.
**independent reasons**: A group of premises, or reasons, that are given as support for a conclusion, each of which could support the conclusion on its own.

**stress test**: A technique used to examine the strength or stability of an entity under operational conditions that are more extreme than what is expected normally. In analyzing argument flow, a technique for detecting dependency between reasons by eliminating them individually in order to see whether the argument still goes through.

**warrant**: A general underlying principle that licenses an inference from data to a conclusion. In a probability warrant, the strength of the link between premise and conclusion is expressed in terms of probabilistic connection (e.g., 90 percent of the time, premise A is linked to conclusion B). In a definitional warrant, the premises are linked to conclusion as a matter of definition (e.g., whales are mammals by definition because they breathe air and give live birth). A legal warrant relies on a point of law as the link between the premise and conclusion (e.g., a contract requires a signature; thus, this unsigned document is unenforceable). An ethical warrant relies on an underlying ethical belief (e.g., if there is a shared belief that one should not deceive, then the conclusion that a deliberately deceitful act was wrong is warranted).

---

**Suggested Reading**

Kelley, *The Art of Reasoning*.

Toulmin, *The Uses of Argument*.

**Questions to Consider**

1. In Aristotle’s sense of validity, it is impossible for the conclusion to be false if the premises are true. This lecture uses a broader notion of validity: An argument is valid if the premises support the conclusion, even if they don’t make it impossible for the conclusion to be false. What do we stand to gain with this broader sense of validity? What do we stand to lose?
2. All things being equal, which kind of argument is stronger: one that relies on dependent reasons or one that relies on independent reasons?

3. Both Mr. Able and Mrs. Brightman agree that Colonel Mustard deliberately misled his stockholders about the imminent bankruptcy. Mr. Able argues that what Colonel Mustard did was permissible because the laws of contract do not demand full disclosure in such a case. Mrs. Brightman argues that what he did was impermissible because lying is always wrong. Do these people disagree on data in Toulmin’s sense or on warrant? What difference might that make as to where the argument can go from this point?

Exercises

Make up an argument of your own that follows this pattern of reasoning:

Pick a letter to the editor from the local newspaper.
(a) Draw a flow diagram of the argument. Note that you may need to add “hidden premises” or assumptions.

(b) Do the premises really support the conclusion? How solid is the reasoning from premises to conclusion?
A heuristic is a rule of thumb, such as “You should always have three to six months’ pay in the bank in case of emergency” or “To approximate a 15 percent tip, double the sales tax.” Heuristics are guides to action that are simpler than a complete calculation or a logical examination of all possibilities. In this lecture, we’ll learn about heuristics that can be simpler, cruder, faster, and just as good as or even better than a complete rational calculation. Simple heuristics come with a warning label, however. Although they’re often useful, they’re not guaranteed. Calculating a 15 percent tip by doubling the sales tax works only if the tax rate is 7.5 percent.

The Recognition Heuristic
- Two groups of students, one at the University of Chicago and one at the University of Munich were asked the following question: “Which city has a larger population: San Diego or San Antonio?” Which group do you think did better?

- The University of Chicago students are a smart bunch, and of course, they’re American. Sixty-two percent of them got the right answer (San Diego), but 100 percent of the German students answered correctly.

- In fact, the Germans got the answer right precisely because they know less. They seemed to have used this heuristic—this rule of thumb—“I’ve certainly heard of San Diego. I’m not sure I’ve heard of San Antonio. That must be because San Diego is a more prominent city, which probably means it’s larger.”

- Recognition is a fast-and-frugal heuristic. It’s frugal because it works with extremely limited information. In fact, it demands that your information be extremely limited.
Suppose you have three groups of people. One is extremely knowledgeable about an issue. One knows about half as much as the first, and one is totally ignorant. For some questions, because of the recognition heuristic, it’s the middle group that will do the best. Those who know absolutely nothing can’t use it—they don’t recognize any of the alternatives. Those who know too much can’t use it—they recognize them all. It is those in the middle, who recognize some but not all of the alternatives, who can put the recognition heuristic to work effectively.

The Wisdom of Crowds

When it comes to advice or data, what we need to be able to do is to tell the real experts—those who actually know—from the so-called experts. But it may be hard to pick those out. The wisdom-of-crowds heuristic indicates that we may not need to know who the real experts are.

Sir Francis Galton was a distinguished British scientist. He also happened to be a cousin of Charles Darwin. Galton is remembered positively for his studies of heredity and intelligence. He is remembered negatively for founding eugenics.

Galton was convinced that some people are inherently smarter than others. He thought we should breed the smart ones and prevent reproduction by those who were not as smart. He considered the intelligence of the general public to be deplorable. That fact about Galton makes this next case particularly interesting.

In the fall of 1906, the 85-year-old Galton attended the annual West of England Fat Stock and Poultry Exhibition, which had a weight-judging competition. A prize ox was put on display, and people put wagers on how much the ox would weigh after it had been butchered and dressed. There was a prize for the person whose guess was the closest.

Galton examined all the tickets and then took the median—the guess in the very middle, with half of the guesses greater and half
of the guesses less. The median guess was 1,207 pounds. The actual weight of the ox, when butchered and dressed, was 1,198 pounds. That’s an error of less than 1 percent. For the simple average of the guesses, the result was even more amazing. The average of the guesses for the weight of the ox was 1,197 pounds: an error of less 0.001 percent.

- What is the heuristic here? In this case it’s social. When in doubt, go with the mean judgment. When in doubt, rely on the wisdom of crowds.

Statistics Explains the Wisdom of Crowds
- The statistical phenomenon behind the wisdom of crowds is the fact that the ignorant answers randomize over all possibilities. When we go for the answer that got the most votes, the ignorant answers effectively cancel each other out. That leaves the knowledgeable answers to tip the scale.

- Statistics shows when you can rely on this heuristic and when you had better not.
  - In order for this to work, you must have a crowd in which some people really do know the answer. A crowd of schoolchildren, for example, won’t outguess Warren Buffett on stock choices.
  - The other condition is that the wrong answers must really be random. The heuristic won’t work if all the ignorant people
have the same kind of ignorance, for example, if they all give the answer “Mississippi” to all questions about rivers because it’s hard to spell. Lots of people with lots of different biases is fine—those biases will cancel each other out. What you don’t want is to have all the biases in the same direction.

The Heuristic of “Satisficing”

- In decision making, the ideally rational setup is one in which you examine all courses of action, calculate consequences, compare benefits and losses, and then decide. If you are making an important decision and you have the time, that’s what you should do. But the demands of action often preclude that kind of extended consideration.

- Often, the alternatives are too many and the demands are too immediate. If you are a firefighter deciding how to fight a basement fire, you don’t have time to be a “perfect maximizer.” You don’t have time to look up the blueprints, map out all possible ways of entry, analyze the probabilities that certain fire-retardant materials were used in construction, access medical records in order to determine the probability that a child on the third floor is asthmatic. Lives are at risk and property is going up in flames.

- The heuristic to use in situations where we don’t have time to be perfect maximizers is satisficing. Pick a realistic set of goals. Now, start looking for a course of action that will achieve those goals. Consider alternatives in any order they happen to occur to you. You’re not looking for perfection; you’re looking for good enough. You’re trying to be a “satisficer.” The heuristic: Take the first course of action that meets your goals and go.

- There is biological evidence that satisficing is built into a wide range of animal behavior. Believe it or not, satisficing is often used in computer programs, as well. Even computers don’t have time to consider all the options in every case.
Rational Calculation or “Go with Your Gut”?

- Heuristics are rational shortcuts. Given the selective pressure imposed by the environment, it is not surprising that animals should have some heuristic shortcuts built in. The animal has to do something to shake off the predator and has to do it now. If there is a way of doing that without all the time delay of careful logical calculation, evolution will select for it.

- People also have an evolutionary history. When we look at ourselves, we can expect to find some generally useful heuristics built in.

- Should you try to calculate things rationally, or should you “go with your gut”? The answer seems to be both. Neither emotions nor gut reactions should be ignored. They have the experience of our ancestors built in. When a decision is of the same type as those made by our ancestors, and it has to be made fast—such as getting out of the way of an approaching stampede—extended calculation is clearly a bad idea. Go with your gut.

- On the other hand, our rationality is also a major part of our evolutionary inheritance. If it hadn’t been good for anything, we wouldn’t have it. Rational calculation is particularly useful for long-range decisions. In that case, we should use all our resources, survey all available information, and weigh all options carefully. To rely on gut reactions and emotions alone might well be to make the wrong decision.

Cultivating Heuristics

- Not all heuristics are built in. Some can be deliberately acquired. Here’s a lesson on acquired heuristics from one of the great thinkers of our time: Neumann János Lajos.
  - Born in Hungary in 1903, he changed his name to John von Neumann by the time he came to America. He made major advances in set theory, number theory, and computer science and basically invented game theory. Von Neumann was a major mover in the Manhattan project, responsible for the atomic bomb in World War II, and in the later hydrogen
bomb, and he participated in the early years of the Atomic Energy Commission.

- Along with Einstein, he was one of the original members of the Institute for Advanced Studies at Princeton. All our computers embody what is known as the von Neumann architecture.

- The interesting thing about von Neumann is that he quite deliberately built his own kit of heuristics. And it was that kit of heuristics that let him cross fields so easily and productively, from physics and mathematics, to economics, to game theory.

- We’ve all seen technical textbooks with lists of formulae at the back. Although they may seem a little forbidding to us, those were von Neumann’s favorite part of any book. He would deliberately memorize the formulae in the back of textbooks: formulae in matrix arithmetic, in physics, in theoretical biology, in economics.

- He would then use those patterns as heuristics. Von Neumann’s strategy was to familiarize himself with those formal patterns in order to be able to see connections and analogies when he saw something like them somewhere else. He stocked up on mental patterns in one place so he could apply them elsewhere.

- In this lecture, we’ve emphasized the sunny side of conceptual patterns. But there’s a dark side, too—a dark side we can see in some of the mistakes we make. We’ll talk about those in the next lecture.

**Terms to Know**

**eugenics**: A social movement that advocates practices and direct interventions aimed at changing, or ostensibly “improving,” the genetic characteristics of a population.

**heuristics**: Simple guides to action or rules of thumb that allow us to act or make a decision without calculation or deliberation.
**mean**: The average of a set of numbers.

**median**: The midpoint of a set of numbers that has been ordered from lowest to highest; that point at which there are as many below as above.

**recognition heuristic**: A fast-and-frugal rule of thumb for decision making in which one picks the alternative or orders alternatives in terms of recognition; often most effective when one is working on very limited information about alternatives.

**satisficing**: A heuristic, or rule of thumb, for decision making in which one picks a realistic goal or set of goals, then selects the first course of action that meets the goal or goals. This heuristic emphasizes that one is not seeking perfection, just something that is “good enough.”

**wisdom-of-crowds heuristic**: A rule of thumb in which one bases one’s decision on the most popular answer selected by a random group of people. In order for this heuristic to be effective, certain conditions of randomness must be met. The reason it works can be explained statistically by the fact that in an appropriately random group of people, most will be ignorant about the topic, but a few will actually know the answer. The answers of the “ignorant” will distribute randomly and cancel each other out, allowing the answers of those who really know to tip the scale.

**Suggested Reading**

Gigerenzer, Todd, and the ABC Research Group, *Simple Heuristics That Make Us Smart*.

Surowiecki, *The Wisdom of Crowds*.

**Questions to Consider**

1. For what kinds of decisions are you a “satisficer”—taking the first option that satisfies your basic goals? For what kinds of decisions are you a “perfect maximize”? 
2. From your own experience, give a case where you “went with your gut” and ended up with a good decision. Give a case where you “went with your gut” and ended up with a bad decision. Was there a difference in the kind of decision at issue?

Exercise

Try the recognition heuristic on these:
(a) Which is the larger South American city, Bogota or Rio de Janeiro?

(b) List these people in order of wealth: Larry Ellison, Bill Gates, Christy Walton.

(c) Which is the larger African country, Libya or Sudan?

(d) List these pitchers in terms of the most career home runs: Wes Ferrell, Don Drysdale, Walter Johnson.

(See “Answers” section at the end of this guidebook.)
How many times have you berated yourself for a stupid mistake? A friend may try to make you feel better: “You’re only human. Everybody makes mistakes.” It turns out there’s a lot of truth in that. To err is human. Systematic error—not random but systematic and predictable—is built into some of the ways we think. Systematic error is even built into how we see things. In this lecture, we’ll look at the ways that we as humans systematically go wrong and how we can correct for them. If we are aware of cognitive biases, we can compensate for them.

Perceptual Bias

• We have emphasized the power of visualization in a number of lectures. Because of our evolutionary heritage, people are particularly good at fast information processing in visual contexts. But our evolutionary heritage comes with certain built-in biases, as well. There are perceptual heuristics built into the system that systematically mislead us. That’s what optical illusions are all about. A prime example is color vision.

• Although we may think we see shades of color the way they are, our perceptual system isn’t built to see pure colors. It isn’t built to detect the same wavelengths at different spots on a two-dimensional image; it’s built to detect differences in a three-dimensional world—a world with shadows.

• How long in our evolutionary history have we been dealing with three-dimensional contexts and real shadows? Millions of years. That’s what the system was built for. In contrast, we haven’t been dealing with two-dimensional representations for very long at all.

• The heuristics built into the perceptual system were built for one kind of context: a three-dimensional world with real shadows. If we have to judge things in a different context—for example, for an
artist who is judging color in a two-dimensional representation—we have to adjust for that built-in bias.

- We have similar built-in biases regarding judgments of length. For example, in an image of two tables, one seems to be long and narrow, while the other is closer to a square. Amazingly, though, the two shapes are precisely the same. The reason they don’t look the same is that we are judging them as if we are viewing a three-dimensional scene. Our perceptual system is made to detect the way things really are in the world, so it sees them that way.

- The lesson here is that in a two-dimensional context, our perceptual system continues to carry a three-dimensional bias. If we’re dealing with two-dimensional representations, we have to compensate for that.

- There is another surprising thing about these optical illusions. Even after we know the two shapes are the same, they still look different. People often talk about how belief influences perception—that we see what we want to see or what we expect to see. These optical illusions offer a counterexample.

- In cases like this, mere belief seems powerless in the face of our built-in perceptual biases. We know perfectly well that those two shapes are the same. But our built-in perceptual biases seem to ignore what we know.

**Change Blindness**

- Some of our mistakes are built into our information processing at a higher level. One source of error is attention bias. It is not true that we always see what we want to see or what we expect to see, and it’s a good thing that we don’t. But it is sometimes true that we don’t see what we don’t expect to see, particularly if we are concentrating on something else.

- There is a one-minute video from a famous experiment on attention bias that shows two groups of people—some in black shirts, some in
white shirts—interweaving in a small room as they pass basketballs to one another. Subjects were told to count the number of times the basketball is passed from one white-shirted team member to another.

- At about the 30-second mark in the video, a person in a gorilla costume enters the room. The gorilla pauses, thumps his chest, and leaves after about 5 seconds. All this time, the subjects have been trying to count the number of basketball passes. A few minutes later they are asked “Did you see anything unusual?”

- Amazingly, about half the experimental subjects failed to see the gorilla. They were concentrating on counting ball passes. Even when specifically asked, “Did you see anything unusual?” the answer was “No.” They were victims of what is called change blindness, an attention bias.

**Malleable Memory**

- Our memories are far from perfect. We think of our memories as photographs of real events filed away for future reference. But our memories, however vivid, are often backward reconstructions from later reports. You may have a memory of something you did when you were a child. However, it’s likely that what you remember is a reconstruction from family stories of that event, rather than the event itself.

- The influence of verbal framing on memory has also been studied by researchers. Elizabeth Loftus is one of the world’s foremost experts on eyewitness testimony.
  - In one of her classic studies, people were shown a film of a traffic accident and then were asked questions about it. The only difference was how the question was phrased. One group was asked, “About how fast were the cars going when they hit
each other?” The other group was asked, “About how fast were
the cars going when they smashed into each other?”

- One week later, the same people were asked a few more
  questions, such as “Did you see any broken glass?” There was,
in fact, no broken glass in the accident. But those who had
been asked about cars “smashing” into each other rather than
“hitting” each other were twice as likely to say that there had
been broken glass at the scene. The difference in a single word
created different memories.

**Availability Heuristic**

- Some of our patterns of reasoning come mistake-prone. The
  previous lecture discussed heuristics that make us smart—fast-and-
  frugal intuitive rules of thumb that often seem to pay off. But there
  are also heuristics that make us not so smart. Those are patterns
  that we use all the time and that don’t pay off—habitual patterns of
  thought or perception that systematically mislead us. One is called
  the **availability heuristic**.

- To illustrate this heuristic, answer this question: Are there more
  English words that start with the letter *K* or more English words
  that have the letter *K* in the third position?

  - If you’re like most people, you would say that there are more
    English words that start with the letter *K*. Why? It’s easy to
    rattle off words that start with *K*: *kite*, *kangaroo*, *kindergarten*,
    *Kool-Aid*. Those words are “available,” and we follow an
    availability heuristic: If it’s easier to think of, we tend to think
    it’s more common.

  - But in fact, there are about three times as many English words
    that have a *K* in the third position: *bike*, *ask*, *acknowledge*. It’s
    much harder to think of words in terms of a third letter rather
    than a first letter. But that has nothing to do with how many
    there are. The availability heuristic misleads us.
Anchor-and-Adjustment Heuristic

- Just as images can mislead us, so can our starting points. Those starting points are called anchors. Here, it’s an anchor-and-adjustment heuristic that can mislead us.

- The starting point—the anchor—can come from many places, such as how a question is phrased. Memory can be influenced by how a question is phrased. It turns out that the same holds for rational calculation.

- Consider this: “What percentage of African nations do you think are members of the United Nations? Is it more than 10 percent?” People who were asked the question that way tended to give answers averaging about 25 percent.

- Other people were asked the question this way: “What percentage of African nations do you think are members of the United Nations? Is it more than 65 percent?” In that case, the average answer was 45 percent.

- The theory is that the figure given in the question is taken as an anchor. “Is it more than 10 percent?” Yes, it must be more than that; thus, you adjust up from that anchor. You end up in the 25 percent range, perhaps. When the question is phrased as “Is it more than 65 percent?” people use 65 percent as the anchor and adjust down, ending up in the 45 percent range.

- Of course, the real probabilities have nothing to do with how the question is asked. Something like 93 percent of African states are members of the United Nations. The anchor-and-adjustment heuristic has led us astray.

Cultural Patterns of Thought

- Heuristics are rules of thumb for patterns of thought. But the conceptual patterns don’t have to be technical. Collecting patterns of thought across cultures can do the same thing, even if the patterns are something as simple as fairytales.
• Consider this puzzle: A treasure hunter is going to explore a cave on a hill near a beach. He suspects there might be many paths inside the cave and is afraid he might get lost. Obviously, he does not have a map of the cave; all he has with him are some common items, such as a flashlight and a bag. What can he do to make sure he does not get lost when trying to get back out of the cave? The answer is that he can take a bag full of sand from the beach and sprinkle it along his path.

• When that puzzle was given to American college students, about 75 percent gave the correct solution. But when the puzzle was given to Chinese students, only 25 percent solved it. The theory is that American students had a heuristic; they had a template of how to solve this kind of story. They had the story of Hansel and Gretel.

Terms to Know

**anchor-and-adjustment heuristic**: A common strategy used in calculating probabilities, but one that depends on how a question is phrased. Information given in the question is taken as a starting point, or anchor; individuals tend to adjust their responses upward if the anchor seems too low or downward if the anchor seems too high, arriving at an answer that is less extreme than the information given in the question but that may have little connection to the real answer.

**attention bias**: Overlooking the unexpected because we are attending to the expected.

**availability heuristic**: The tendency for individuals to assume that things that are easier to bring to mind must be more common or occur more frequently; the tendency to generalize from simple and vivid images generated by single or infrequent cases and to act as if these are representative.

**change blindness**: The propensity for individuals not to perceive unexpected changes, particularly when attention is focused on something else.
**perceptual bias**: A “hard-wired” tendency in our perceptual processing that forces us to perceive things in particular ways. Our color perception does not track pure wavelengths of light or actual lengths in a stimulus, for example, because our visual processing has evolved to interpret input immediately in terms of contextual cues regarding shadow and perspective.

**Suggested Reading**

Hallinan, *Why We Make Mistakes*.

Hoffman, *Visual Intelligence*.

**Questions to Consider**

1. Which line looks longer, the vertical or the horizontal?

   ![Line Illustration]

   Adolf Fick first studied this illusion in 1851. We seem to have a built-in propensity to exaggerate the vertical. The division of the lower line contributes to the illusion.

   Measure the two lines.

2. You now know they are the same length. Does the vertical line still look longer? What does that say about the claim that our beliefs determine what we see?

3. The first lecture used a “bird in the thorn bush” example. We are much better at integrating images when they are moving. Can you offer an evolutionary hypothesis for that? Can you offer an evolutionary hypothesis for our propensity to exaggerate the vertical?
Watch the attention bias experiment on-line (http://www.theinvisiblegorilla.com/gorilla_experiment.html). When instructed to count the number of passes made by people in white shirts, almost half of those watching the video failed to see the gorilla.

The degree of our change blindness is astounding. Watch some of the examples posted online by the Department of Psychology at the University of British Columbia (http://www2.psych.ubc.ca/~rensink/flicker/download/) or search for “Change Blindness Experiments” on YouTube.

Exercises
Rational Discussion in a Polarized Context  
Lecture 12

We seem to find ourselves in the midst of increased polarization: polarization of argument, of opinion, and of values. What is the role of rationality in a polarized context? On a more individual level: Is it possible to have a rational discussion with people who seem to disagree with you radically, on fundamental issues? Is there a way for rationality to reduce polarization? We’ll look at these questions in this lecture.

Partisan Polarization

- Before we look at the facts about polarization in America, we need to draw an important distinction. What we will call “partisan polarization” is polarization between political organizations, on the scale of either major political parties or smaller interest groups. What we will call “cultural polarization” is polarization within the general public on attitudes that may or may not be expressed politically.

- Partisan polarization is real. In 1980, 43 percent of Americans polled said that they thought there were important differences between the political parties. That figure is now 74 percent. In 1976, almost a third thought it didn’t even make a difference who was president; that figure is now cut in half.

- Internal uniformity within each party has grown just as distance between them has. But partisan polarization isn’t new. George Washington’s farewell address in 1796 emphasized the danger of “factions.” A year later, Thomas Jefferson complained that because of partisan polarization, “men who have been intimate all their lives cross the streets to avoid meeting.”

Cultural Polarization

- Although partisan polarization has increased over the last few decades, cultural polarization—polarization in general social attitudes among the general public—may not have.
• Polarization in public attitudes is limited to a few hot-button issues. On most issues, public polarization hasn’t increased between groups, regardless of what groups are being compared: the young and the old, men and women, the more and the less educated, different regions of the country, or different religious affiliations. On a number of fronts, polarization has, in fact, decreased.

• What then explains the polarization we do see? First, we’ll look at some answers from political scientists and sociologists. Then, we’ll explore the internal and logical dynamics of polarization, of particular interest to philosophers—the quasi-rational part of the picture.

• Political scientists concentrate on partisan polarization. They note historical changes in the parties. The Republican Party has drifted to the right; northern Democrats have become more liberal, and southern Democrats tend to be affiliated with Republicans.

• Sociologists add some observations regarding social and cultural polarization. Since the 1970s, we have sorted ourselves more residentially by income and social class. Our neighbors are increasingly like ourselves. Sociologists also note that America has a long history of fraternal and voluntary organizations that once brought people together with those who might not be exactly like them. But membership in those long-term organizations has waned, replaced with self-selecting short-term membership in groups organized for some single political purpose.

A Philosophical Analysis of Polarization

• When we think about it, it seems that a polarization of attitudes or opinions is to be expected. It may even follow something like rational lines.

• Suppose that Mr. Magoo is very much convinced of a particular opinion. Because he’s convinced that the position is right, he’s also convinced that any evidence offered against the position must be flawed in some way—bad research, perhaps, or an improper
interpretation. In the philosophical literature, Mr. Magoo is what is called a **Kripkean dogmatist** (based on a lecture by the contemporary philosopher Saul Kripke). An extreme case, Mr. Magoo thinks his current position is right and, on that ground, is prepared to reject any information to the contrary.

- Now, imagine two people who are strongly convinced of two opposing positions: Mr. Magoo and, say, Dr. Seuss. Both are Kripkean dogmatists. Both selectively reject evidence and argument against their positions. Both are selectively receptive to evidence and argument for their positions.

- Suppose these two are given a mixed bag of new evidence and arguments, just as we get on just about every topic every day. The new information includes a handful of evidence and arguments that seem to point in one direction, but also a handful that seem to point in the other.

- One might think—one might hope—that new evidence and argument would bring the two views closer together. But if we are dealing with Kripkean dogmatists, that is not what will happen. If attitudes start out already polarized, new evidence can make them even more so.

- Mr. Magoo and Dr. Seuss represent an extreme case, but the same dynamics will operate even when the case isn’t that extreme.

When shown images of the earth from space, a spokesman for the Flat Earth Society said, “It’s easy to see how a photograph like that could fool the untrained eye,” illustrating the Kripkean dogmatist’s selective rejection of evidence against his position.
A Case Study of Polarization

- A set of classic studies in social psychology shows precisely this effect. In these studies, individuals’ attitudes toward the death penalty were measured. Subjects who favored the death penalty were put in one group. Those who opposed it were put in another. Both groups were then given the same two pieces of evidence: results of one study supporting a claim of deterrence and results of another that opposed deterrence. Each group was also given further information on the two studies: experimental details, critiques in the literature, and the researchers’ response to those critiques.

- The result was just as you would expect from Mr. Magoo and Dr. Seuss. All participants found the studies supporting their positions to be better conducted and less subject to criticism than the counterevidence. Both groups ended up indicating stronger commitment to their original positions than when they began.

Selection of Information

- Another dynamic in polarization that hasn’t yet been strongly studied is selection of information sources.

- There was a time when America had essentially one source of television news: the nightly news on NBC, CBS, and ABC. Because of the canons of journalism these shows tried to avoid an overt political stance. Walter Cronkite was repeatedly cited as the most trusted man in America.

- Of course, we now have Fox News and MSNBC, which are clearly polarized politically. Instant access to any slanted source is immediately available on the Internet.

- Now, not only will Mr. Magoo and Dr. Seuss treat identical information differently, but they will seek out information differently. Mr. Magoo will trust information sources that reinforce his existing beliefs. Dr. Seuss will do the same on the other side. Because both will be able to find media reinforcing their existing beliefs, we can expect polarization to increase.
Warning Indicators of Polarization

- There is a good warning indicator of when two people might be falling prey to the dynamics of polarization we’ve been talking about. Suppose you and a friend agree on many things. On those issues, she seems as perfectly rational as you are. But her position on issue X seems radically wrong. How can she possibly think that?

- That’s the warning indicator. If it’s hard to believe that someone so rational in other areas could be so irrational in area X, that’s a sign that it’s time to lay out evidence and argument on both sides—time to recalibrate and reevaluate.

- There’s no guarantee that like-mindedness will result. But strong disagreement in a small area with a person you think is generally rational is an indicator that you may be playing Mr. Magoo to her Dr. Seuss.

Rationality in a Polarized Context

- In our attempts at rationality in a polarized context, the literature on negotiation strategies seems to fit surprisingly well. William Lury and Roger Fisher are members of the Harvard Negotiation Project and authors of the influential *Getting to Yes*. Following are some of their tips.

- Rationality deals in cool-headed evidence and argument. A major theme in the work of Lury and Fisher is removing an issue from its ties to emotion and ego, what they call “going to the balcony.” The idea is for both sides to try to view the issue as if they were not embedded in it. Try to see it from a distance—as if the positions were being advanced by someone entirely different.

- A correlate theme is ego-distancing. People identify with the views they hold. That’s part of the problem. But of course, neither truth nor rationality cares whether a position happens to be mine or yours. If an issue can be cast as one for joint resolution, through joint effort, that’s a step away from individual ego. And it’s clearly a step in the right direction.
- In **positional negotiation**, people identify themselves with specific positions from the start—positions to be won or lost. In positional negotiation, you may feel forced into one of two roles. You end up either a “soft” negotiator or a “hard” negotiator. But neither of those really works. The alternative is to change the game. If the context can be changed from a contest of wills to a mutual exploration and investigation of an issue, there’s a chance that the issue will be resolved.

- In leaving positional negotiation behind, Lury and Fisher urge people to look for a way to settle issues in terms of objective criteria. In the case of negotiation, that may mean an appeal to **independent standards** that allow us to look at the totality of evidence. The parties must decide in advance what kind of evidence will convince them of a particular position. If they can agree on that, they can look with new eyes at what the evidence actually is.

- Sometimes a dispute isn’t about data or evidence but about background principles. In those cases, the discussion must go deeper. The parties must ask what principles are at play and what kind of principles those are.
  - For example, proponents of gun control tend to emphasize statistics regarding harm. Opponents of gun control tend to emphasize constitutional rights.
  - But if we take the principles out of the context of debate, people on both sides recognize that the Constitution says something important about rights. People on both sides recognize that we should avoid unnecessary harm. If we can get from polarized people to joint recognition that the issue is one of conflicting principles, we’ve made progress.

- Is rhetoric at issue in such debates and negotiations? Absolutely. We’ll talk more about that subject in the next lecture.
independent standard: In negotiation or the attempt to reduce polarization, a deciding touchstone or court of appeal that is not open to manipulation and that can be agreed on by both parties in advance. In establishing a fair price for a house, for example, both parties might agree in advance to use the price that similar houses have recently sold for in the neighborhood.

Kripkean dogmatist: An individual who believes that his or her position is right and, on that ground alone, is prepared to reject any and all evidence to the contrary.

negotiation strategy: An approach to conflict resolution that may attempt to remove the “contest of wills” characteristic of positional negotiation. Among other techniques, negotiation strategies may include employing ego-distancing; talking about issues without identifying with a particular position; “going to the balcony,” that is, trying to put emotions aside and view the problem from a distance; appealing to independent standards; coming to some agreement about what kinds of objective criteria could help to clarify or settle the issue; and replacing debate with collaborative research on the topic.

dependency: Radical or extreme disagreement between groups with no apparent willingness to compromise and/or with few individuals representing a middle group between the extreme positions. Polarization normally implies a wide gap between positions and increased uniformity within positions. Political polarization refers to extreme positions taken by political organizations, either major political parties or smaller interest groups; cultural polarization refers to extreme differences in attitudes of the general public that may or may not be expressed politically.

positional negotiation: In conflict resolution scenarios, an approach in which people are ego-involved or identify with their specific positions. Those involved in positional negotiation often end up in one of two roles: as the “soft” negotiator, who tries to avoid conflict by giving in and winds up feeling exploited, or as the “hard” negotiator, who is out to win at all costs and, thus, starts with an absurd extreme, allowing room to make some concessions and still hit his or her initial target.
Suggested Reading

Brownstein, *The Second Civil War*.

Fiorina (with Abrams and Pope), *Culture War?*

Fisher and Lury, *Getting to Yes*.

Lord, Ross, and Lepper, “Biased Assimilation and Attitude Polarization.”

Questions to Consider

1. Many estimate that polarization in America is at a high point now. From your own experience, what periods do you remember as particularly polarized? As less polarized?

2. This lecture included some suggestions about how to reduce polarization, but it might be easier to prevent it. Can you suggest some approaches to discussion that might help avoid undue polarization from developing?

3. A general question in epistemology is as follows: It is irrational to discount new claims and new evidence simply because they conflict with what we already believe. On the other hand, we have to work from past experience: We can’t examine every one of our beliefs anew at every minute. What holds for us as individuals also holds for our society and our science. Can you think of any useful rules of thumb that can guide us in deciding what beliefs are worth reexamining and when?

Exercise

This is an exercise to work on together with a friend, a friend you talk to frequently, with whom you agree on many things, but with whom you disagree on some. Together, discuss the following questions:

(a) What topics are you two most in agreement on?

(b) What topics are you most polarized on?

(c) Why those topics?
Rhetoric has acquired a bad name, as in the expressions “mere rhetoric” or “empty rhetoric.” But that wasn’t always the case. At one time, rhetoric was highly respected as a required skill for effective speaking and presentation of ideas. What makes a presentation effective? According to Aristotle, the character of the speaker, which he called *ethos*, resonating with the emotions of the hearer, which he called *pathos*, and the logic, or rationality, of the argument, which he called *logos*. This lecture focuses on rhetoric versus rationality.

**The History of Rhetoric**

- At one time, rhetoric was highly respected as a required skill for effective speaking and presentation of ideas. Plato has a dialogue on the theme, the *Gorgias*, and Aristotle wrote a major work on the topic, *Rhetoric*.

- Classical education throughout the Middle Ages was based on the *trivium*, consisting of *grammar*, logic, and rhetoric. Those were the foundations of a liberal education, the areas everyone was expected to master.

- But the dark side of rhetoric also became clear early on. In Plato’s *Gorgias*, Socrates argues that rhetoric involves no real knowledge or sense of justice—that rhetoric alone, unalloyed with philosophy, is merely a form of deceptive advertising or flattery.

- In his work on the topic, Aristotle admits that rhetorical skills can be used for either good or bad. His task is to analyze persuasion. What is it that makes a presentation effective?
  - The persuasiveness of a presentation depends first of all on the character of the speaker. We are convinced by people we see as knowledgeable and wise, as effective leaders who are in the right. Aristotle calls that *ethos*.
○ Second, the persuasiveness of a presentation depends on the emotions of the hearer. A message that resonates emotionally is one that is more effective. Aristotle calls that pathos.

○ Third, persuasiveness depends on the logic of the argument. If you want to persuade, lay out the argument clearly and rationally, step by step. That is logos.

**Rhetoric in Lincoln’s Cooper Union Speech**

- The political scientist Frank Myers points out that all three elements of Aristotle’s *Rhetoric* are artfully managed in Lincoln’s famous Cooper Union speech. The question of the day was whether the federal government had the right to control the expansion of slavery into the western territories.

- With no flowery introduction, Lincoln takes on the question in terms of the Constitution. Despite his appearance, despite his frontier accent, Lincoln uses that part of the speech to establish himself as knowledgeable and, indeed, scholarly with regard to the questions at issue. In this, he fulfills the requirements of Aristotle’s quality of ethos.

- Another element of Aristotle’s *Rhetoric* is the logic of the argument. In that aspect Lincoln’s speech was accurate and well constructed, similar to a lawyer’s brief. Emotion, or pathos, appears in the very last part of Lincoln’s speech. Had he started with an emotional appeal, it is likely people would have seen him only as a ranter from the wilderness. What he needed first were character and logic.

**The Dark Side of Rhetoric: Schopenhauer’s Stratagems**

- Our contemporary and negative sense of rhetoric is firmly in place in the work of the 19th-century philosopher Arthur Schopenhauer. Schopenhauer was the author of *The Art of Controversy*, a user’s guide to rhetoric in the contemporary negative sense. He characterizes the topic as intellectual fencing used for the purpose of getting the best in a dispute. Chapter 3 of Schopenhauer’s book
is titled “Stratagems.” It’s a list of 38 rhetorical tricks. It might as well be called “How to Win an Argument by Any Means Possible.”

- Some of his tricks include explicitly manipulating emotion. Make your opponent angry, Schopenhauer advises. He won’t think straight. How should you do that? “By doing him repeated injustice, or practicing some kind of chicanery, and being generally insolent.”

- But emotion itself isn’t bad. Life and even rationality would be impossible without it. There are points, however, at which something more is required: careful weighing of alternatives, consideration of evidence, tracking of potential consequences, and, after due consideration, arriving at a decision. Appealing to emotion is intended to cross-circuit all that—pushing your buttons instead of appealing to your reason.

- **Appeal to emotion** is often by means of emotionally loaded words. Consider “Jones won the election,” as opposed to “Against all odds, Jones finally triumphed over the status quo.” The difference is in the emotionally loaded words.

- There are other classic fallacies that appear on Schopenhauer’s list of rhetorical tricks. The stratagem known as the **straw man fallacy** appears in a number of forms. Schopenhauer recommends exaggerating the claims made on the other side. That will, of course, make them easier to refute. He recommends twisting the words used on the other side so that they mean something different than intended. Ignore your opponent’s real position. Confuse the issue by imputing something different. The fallacy is called “straw man” because you are no longer arguing against the opponent’s real position.

- Among Schopenhauer’s recommendations is a standard debating trick: Shift the burden of proof to the opponent. One way to do this is simply to say, “This is my position. Go ahead. Prove me wrong.” That relieves you of the burden of proving yourself right. It’s your
opponent who is on the spot. But, of course, the fact that he may not be able prove you wrong doesn’t really prove that you’re right.

- Schopenhauer also recommends that you lump the opponent’s position into some odious category. “That’s fascism,” you can say, or “That’s socialism.” The emotionally loaded category is meant to do all the work by itself—and often does.

- If all else fails, Schopenhauer says, bluff. Or throw in so many unrelated charges that your opponent doesn’t have a chance to respond to them all. Or bewilder him or her with mere bombast.

- Schopenhauer’s rhetorical recommendations are ironic, of course. He didn’t want to be a victim of bombast any more than we do. But look at how far rhetoric in that sense is from what Aristotle meant. Look at how far rhetoric in that sense is from the presentational dynamics of Lincoln’s Cooper Union address. Unfortunately, it’s not so far from what we’ve become accustomed to in political argument.

The Ethics of Argument

- There is another lesson here regarding the ethics of argument. To most people, rhetorical tricks are intellectually repulsive. Schopenhauer’s stratagems have all the charm of a wax museum arcade of famous murderers. The best that mere rhetoric can offer is merely winning. Truth and genuine rationality demand a far higher standard than that.

- Schopenhauer says that the intellectual fencing he’s talking about has nothing to do with truth. That’s precisely the problem. Rational discussion, like rational argument, should be about the truth—a truth that may include the truth about people’s rights, about the best available policy, or about reasonable accommodation between different interests.

- We want to work toward the truth by means of rational argument. Reasoning with others, like reasoning with ourselves, comes with
a built-in goal: to work toward the truth. Persuasion alone isn’t a worthy goal. That’s mere rhetoric.

The Positive Side of Rhetoric

- Rationality isn’t an exclusively individual undertaking. Rationality is something that we can and should practice together. Rational discussion is just that—discussion. Discussion is a social endeavor, and rationality may actually demand a social dynamic. Often, it is only when we see ideas in a context of discussion, or even civil debate, that we can see them most clearly.

- That is certainly the theory in criminal law and civil suits. The adversarial system is based on the conviction that it is from the clash of arguments, in which each side presents its best case, subject to cross-examination, that “the truth will out.”

- Does the truth always come out in adversarial proceedings? We all know that a case may be carried by a lawyer’s rhetorical tricks—appeal to the emotion of the jury, for example—rather than by the force of evidence and argument. The consensus, however, seems to be that there is at least a better chance that the truth will out when we have arguments and evidence laid out on opposing sides, with a final verdict left to the collective deliberations of the jury of our peers.

Visualization: Graphing Logical Flow

- In a previous lecture, we used flow diagrams to graph the logical path of an argument. In that context, we were talking about an
argument as a one-sided affair. We can also use visualization to graph both sides of an argument, creating a flow diagram from a simple exchange between two people about a broken washing machine.

- When a philosophical discussion becomes complex, it helps to sketch it out in advance. Include not only exchanges you have heard but possible exchanges to come, that is, graphs with branches for different possibilities.

- The analysis of a rational discussion differs from that of a single argument in that there will be more than just premises and conclusions—more than even data, warrant, and conclusions. There will be challenges raised on one side that will force qualifications on the other, and that may well evoke refinement of the argument and further backup in response.

- Are such analyses complicated? Yes, yet in normal discourse, we manage to follow the train of thought, the logical flow of argument, all the time. What graphic visualization allows is the possibility of seeing that logical flow. What we’ve done here is to expand our graphing tools into the social context of discourse.

Terms to Know

**appeal-to-emotion fallacy**: A fallacy in which positive or negative emotional tone is substituted for rational or evidential support; an argument strategy intended to cross-circuit the ability of the listener to assess whether a rational link exists between premise and conclusion by “pushing emotional buttons.”

**ethos**: Character of the speaker; according to Aristotle, the first quality of a persuasive presentation is that the speaker must appear knowledgeable and wise.

**grammar**: The study of the proper structure of language in speech or writing; along with logic and rhetoric, an element of the classical medieval curriculum known as the *trivium*. 
logos: Logic; according to Aristotle, the third quality of a persuasive presentation is that it will lay out the argument clearly and rationally, step by step.

pathos: Emotion; according to Aristotle, the second quality of a persuasive presentation is that it resonates emotionally with the listener.

straw man fallacy: The rhetorical strategy of exaggerating or misrepresenting the claims of the opposition in order to more easily appear to refute those claims.

trivium: The basis of a classical education throughout the Middle Ages; a three-part curriculum composed of grammar, logic, and rhetoric.

Suggested Reading

Aristotle (Roberts, trans.), *Rhetoric*.

Schopenhauer, *The Art of Controversy*.

Questions to Consider

1. Aristotle considers rhetoric as the art of persuasion, and recognizes that it can be used to either or good or bad ends. From your own experience, can you think of an example of rhetoric used for a good end? For a bad end?

2. The negative side of rhetoric: If all else fails in “intellectual fencing,” Schopenhauer recommends these rhetorical tactics: (a) throw out so many unrelated charges that the opponent doesn’t have time to respond to them all and (b) become personal, insulting, and rude. Have you seen examples of those tactics in debates or arguments? Do you think they worked effectively?

3. The positive side of rhetoric: Can you think of a case in which you worked out a problem by exchanging ideas with a friend—a problem that neither of you would have solved independently?
Look up Shakespeare’s *Julius Caesar*, Act III, scene ii. This scene contains the whole of both Brutus’s and Mark Antony’s speeches. Analyze each in terms of Aristotle’s categories: *ethos* (speaker’s character), *pathos* (appeal to audience emotion), and *logos* (logic of the argument).

In the next political debate you see, pay attention to just the first two minutes of each speaker’s presentation or to the answers to the first two questions directed to each speaker. With paper and pencil, keep track of the rhetorical pose of each speaker, specific claims made, and rhetorical tricks used.
The first requirement for a good argument is validity—that the premises, if true, will lead us by strong logic to the conclusion. The second requirement is that the premises actually be true. A sound argument is one that is both valid and has premises that are indeed true. That is the kind of argument that will give us a conclusion we can rely on. This lecture presents bad arguments to enable you to recognize them and defuse them. The standard forms of bad arguments are called fallacies, and these appear in social and political argument all the time. In this lecture, we’ll look at a number of examples.

**Appeal to the Majority**
- Every parent is familiar with the refrain “but everybody else is doing it.” Arguments that treat majority opinion as if that alone constituted a reason for belief commit the appeal-to-majority fallacy.

- Here’s a simple example: “A majority of the population thinks we coddle criminals. It’s time we got tough on crime.” How much support does that premise give for that conclusion? Not much. That’s the fallacy of appeal to the majority.

- There is a simple fact that defuses a number of common fallacies, including appeal to the majority. It’s this: The fact that someone believes something doesn’t make it so. The fact that a whole bunch of people or a majority of people believe something doesn’t make it so either.
**Post Hoc Ergo Propter Hoc**

- *Post hoc ergo propter hoc* is a fallacy that just about everyone has committed. Here, two things happen in sequence, and we conclude that they must be causally connected. The name of the fallacy is a bit of Latin that means, essentially, “This came after this, so it must have been because of this.”

- Many superstitions are based on this fallacy. Professional baseball outfielder John White offers a typical example: “I was jogging out to centerfield when I picked up a scrap of paper. I got some good hits that night, and I guess I decided that the paper had something to do with it. The next night I picked up a gum wrapper and had another good night. I’ve been picking up paper every night since.”

- The thing to remember is that sometimes events are just coincidences. The fact that A happened and then B happened doesn’t prove that A caused B.

**Ad Hominem**

- An *ad hominem* (“against the person”) argument is one that attacks the opponent as a person, rather than his or her arguments.

- Here’s an example: “The Republicans say we should shrink the deficit and cut down on government spending. But these are the same people who cranked up the deficit by getting us into two wars simultaneously. These are the same people who cranked up the deficit by passing tax breaks for the rich. These are the people who gave us the Bush years of spending more massive than any preceding administration of either party.”

- What’s really at issue here is the claim that we should shrink the deficit and cut down on government spending, but that claim is never addressed at all. What we get instead is an attack on the people who put it forward.
• The point is that good arguments are good arguments even in the dark; they are good arguments even if you can’t see who is giving them. That’s the principle violated by the fallacy of *ad hominem*.

**Tu Quoque and Poisoning the Well**

• A specific form of the *ad hominem* fallacy is *tu quoque*, which is Latin for “you also,” or more colloquially, “You did it, too.”

• Here is an example of the *tu quoque* form of *ad hominem*: “The United States says that North Korea is stockpiling nuclear weapons and is, therefore, a danger to world peace. But surely both Israel and the United States have their own stockpiles of nuclear weapons!” Does this show that North Korea is not stockpiling nuclear weapons or that such action is not a danger to world peace? No, it simply states, “You did it, too.”

• A second form of *ad hominem* is the *poisoning-the-well* fallacy. Like all *ad hominems*, it involves an attack on the person rather than the argument, but it is specifically an attack on someone’s motives for his or her statements.

• In 2004, when campaigning for the presidency, John Kerry published a misery index intended to show that things had gotten worse under the George W. Bush administration. The other side fired back: “John Kerry talks about the economy because he thinks it will benefit him politically.” True, perhaps, but irrelevant. That’s just poisoning the well.

**False Alternative**

• The trick of a *false alternative*, or false dilemma, is to set things up as if the opponent must choose either A or B. It’s a false alternative because there may be many more options open.

• Here’s an example: “You either marry Julie and have a wonderful life and family, or you live the rest of your life as a lonely bachelor. Go ahead. Make your choice.”
That’s clearly a false alternative. Are those your only options? Of course not. You might live the life of a gregarious, fun-loving bachelor instead of a lonely one. You might end up marrying Julie and not having a wonderful life. You could move to Alaska, start a commune, marry someone else, and live happily ever after.

**The Complex Question**

- The complex-question fallacy takes the nefarious strategy of the false alternative a little farther.

- The classic example is a lawyer who is prosecuting someone for wife beating. The accused is put on the stand and told to answer merely yes or no to the questions put to him. The prosecutor says, “Tell me, Mr. Smith, have you stopped beating your wife yet?”

- Mr. Smith must say either yes or no. But if he says yes, the prosecutor says “Aha! You say you have stopped beating your wife. So you admit you beat her in the past. I rest my case.” If he says no, the prosecutor says “What? You say you haven’t stopped beating your wife? You admit you’re still beating her. I rest my case.”

- It’s a trick question, a trap. The defendant is forced to say yes or no. Either answer is legitimate only on a certain assumption—that he has beaten his wife. He’s forced into a false alternative, each side of which begs the question against him.

**Hasty Generalization**

- The hasty generalization fallacy is also known as jumping to conclusions. If you know something about a small number of things, you then jump to a conclusion about all of them.

- Here’s an example: “The first person I met on Long Island was short and rude. The second person I met on Long Island was short and rude. I guess everybody on Long Island must be short and rude.”
• Racist and sexist generalizations are usually hasty generalizations: “All the women I’ve ever known were big tippers; therefore, all women are big tippers.”

**Appeal to Ignorance**

• Similar to the fallacy of the appeal to emotion, the **appeal to ignorance** is equally illegitimate. Here, the fallacy is to act as if our ignorance is something more than it is—as if ignorance alone gives us some kind of positive evidence.

• Here’s an example: “Many ships and planes have utterly disappeared in the regions of the Bermuda Triangle, without explanation and often without a physical trace. Some mysterious force must be at work beneath the waves.”

• Just as believing something doesn’t make it so, sometimes ignorance is just ignorance. Ignorance isn’t by itself evidence for something—a mysterious force beneath the waves, for example.

**A Closer Look**

• The aim of this lecture was to vaccinate you against some standard forms of bogus arguments, to inoculate you against some logical fallacies. How does that fit in with some of our earlier work? That question might be posed as a set of challenges:
  ○ One of the other lectures introduced the wisdom of crowds as a positive heuristic. Isn’t that just the fallacy of the appeal to the majority?

  ○ What about the example in which students were asked which city is larger, San Antonio or San Diego? The German students made a correct inference from the fact that they had heard of San Diego and not San Antonio. Isn’t that just the fallacy of an appeal to ignorance?

  ○ The conclusion of another lecture was that in short-range decisions, it may be rational to go with your gut. Isn’t that the fallacy of hasty generalization?
• Take a closer look at each of these questions, and you’ll see that there are important differences. The wisdom of crowds doesn’t just recommend that you think what everybody else thinks. It’s a strategy for extracting knowledge from a distribution of opinions in which some people really do know the answer. It draws a clever inference, though a fallible one, from the distribution of beliefs.

• In the San Diego/San Antonio case, the German students did better than Americans in picking the larger city precisely because they knew something about one and were largely ignorant of the other. That’s a clever inference, though admittedly a risky one. It’s not a simple jump from “I don’t know how ancient monuments were built” to “It must have been ancient astronauts.”

• There is also an important difference of context. When there is no time for careful investigation—when you have to get out of the way of an oncoming train—fast and frugal heuristics are precisely what you need. Better to go with your gut and jump to a conclusion, even if it might be wrong, than to wait on the tracks trying to calculate the velocity of an approaching locomotive.

• In the case of explicit argument, the standards for rationality are higher. There, we want more than a handy heuristic. When someone is trying to convince us of a conclusion, and when we have time to really think about it, it’s important to test how tight the logic really is. Do the premises really support the conclusion? How strongly? Does the conclusion really follow? Or is this a bogus argument, built on fallacious reasoning?

Terms to Know

*ad hominen*: A fallacy that depends on an attack against the person making a claim instead of the claim that is being made.

*appeal-to-ignorance fallacy*: A fallacy in which absence of information supporting a conclusion is taken as evidence of an alternative conclusion.
This fallacy acts as if ignorance alone represents some kind of positive evidence.

**appeal-to-majority fallacy**: An argument that treats majority opinion as if that alone constituted evidence supporting a conclusion or gave a reason for belief. This fallacy ignores the fact that people, even large numbers of people, are fallible.

**complex-question fallacy**: A “trick question” presenting a false dilemma, or forced-choice alternative, presented in such a way that any answer is incriminating. For example: “Answer yes or no: Have you stopped beating your wife?” If you say yes, you have essentially admitted that at one time, you did beat your wife; if you say no, you have admitted that you are still beating her.

**false alternative**: A fallacy in which a problem is presented as an either/or choice between two alternatives when, in fact, those are not the only options. Also called a “false dilemma.”

**hasty generalization fallacy**: Also known as jumping to conclusions. This fallacy occurs when one jumps to a conclusion about “all” things from what is known in a small number of individual cases. Racism and sexism often take the form of hasty generalizations.

**poisoning-the-well fallacy**: A fallacy that depends on an attack against a person’s motives for saying something rather than a refutation of the claims being made; a subtype of *ad hominem*.

**post hoc ergo propter hoc**: “After it, therefore because of it”; a fallacy based on the claim that because something followed another thing, it must have been because of that other thing. This fallacy overlooks the possibility of coincidental occurrence. Abbreviated as *post hoc*.

**tu quoque**: A fallacy in reasoning that tries to defuse an argument by claiming, “You did it, too.”
1. Here is a standard form for some magic tricks: The magician seems to make a coin “travel” from point A to point B. The trick is actually done by “disappearing” one coin at point A and producing a different but similar coin at point B. In order to create the illusion of “transfer,” the magician relies on the audience to commit a logical fallacy. What fallacy is that? (See “Answers” section at the end of this guidebook.)

2. Sometimes fallacious forms of argument are just a short distance from practical conceptual strategies. What is the difference between these?
   (a) We have to choose between the options available. How does that differ from false alternative?

   (b) We need to work from past experience. How does that differ from hasty generalization?

   (c) We need to consider the plausibility of an information source. How does that differ from *ad hominem*?

   (d) We often decide policy by putting the question to a vote. How does that differ from appeal to the majority?

**Exercise**

I once tried to give a midterm by showing only the commercials in a half-hour of commercial television and having students write down all the fallacies they saw committed. This was a good plan, but it didn’t work because there were just too many fallacies.
The next time you watch a half-hour of television, pay attention to the commercials. Within three months of an election, give yourself 10 points for each instance of the following fallacies you find:

- Appeal to the majority
- Hasty generalization
- *Post hoc ergo propter hoc*
- False alternative
- Appeal to ignorance
- *Ad hominem* (more common in political than in commercial advertising)
- Appeal to dubious authority (coming in a later lecture)

If you’re trying this exercise outside of an election year, give yourself 15 points for each instance. It should be easy to get 100 points in a half-hour of television; 200 points is a good, solid job; 300 points is outstanding; and beyond that, give yourself credit for having a touch of genius.
In this lecture, we’ll analyze a “great debate” between two speakers, Mr. McFirst and Ms. O’Second, on the topic of democracy. The purpose of all debates is to mold opinion. As Lincoln said in the first of a series of debates with his rival, Stephen Douglas, “Public sentiment is everything. With public sentiment, nothing can fail; without it, nothing can succeed.” The debate we’ll look at in this lecture won’t rise to the level of Lincoln and Douglas, but you can use it to hone your skills in argument analysis.

McFirst Opens
- McFirst opens the debate by stating that democracy is the “wave of the future.” As evidence, he cites the Arab Spring, as well as some data from Freedom House, an organization that conducts research on the topic of freedom around the world.

- McFirst’s opening argument has two conclusions: (1) “What we have witnessed is the increasing democratization of the world” and (2) “Democracy is the wave of the future.”

- His support for the first conclusion includes some vague images (dictatorships are falling), in addition to the data from Freedom House. His argument isn’t bad, at least if we can trust the data and the source.

- In support of the second conclusion, McFirst says that if we project the Freedom House numbers into the future, we will have worldwide democracy. But we don’t really have any further reason to believe that. In his emotional language, McFirst also makes use of the fallacy of appeal to emotion.

- In considering the statements of both speakers, we should ask ourselves three questions: (1) What conclusion or conclusions are at issue? (2) What argument or arguments are given? How does
the reasoning go? How good are those arguments? (3) Have any fallacies been committed?

O’Second Responds

- Ms. O’Second responds to McFirst’s conclusions by noting that not all players in the events of the Arab Spring were advocates of democracy and questioning the objectivity of Freedom House. She also states that there is no reason to believe that the data from Freedom House can be extended into the future to predict worldwide democracy.

- What is O’Second’s conclusion? That her opponent is wrong, of course. Which of McFirst’s two claims was O’Second attacking? The answer is both. Part of her response was directed at his claim that we’ve seen democracy increasing. The other part of her response was directed at the claim that democracy is the wave of the future. But she used different arguments in those two parts.

- In addressing the claim that democracy is increasing, O’Second had two arguments. Her first is that we can’t know what we have seen until we see how it eventually pans out. But that argument doesn’t really prove that McFirst is wrong. It attacks not so much the claim as McFirst’s confidence in it.

- In her second argument, O’Second attempts to knock down the one data-driven argument McFirst gives by impugning the source Freedom House. But that argument doesn’t show that the data are wrong. Like the first argument, its goal is to raise doubts.

- What about O’Second’s challenge to McFirst’s claim that democracy is the wave of the future? At that point, she uses an argument from analogy: the multiple-blade razor analogy. But the analogy doesn’t really work here. At best, it simply shows that we can’t always project current trends into the future. It doesn’t give us any particular reason to believe that this is one of those times.
• O’Second uses an *ad hominem* fallacy, poisoning the well: “It is surely significant that Freedom House is located in Washington, DC.” She adds to that fallacy with an appeal to ignorance. She doesn’t really give us any solid reason to think Freedom House should be discredited.

**McFirst Uses Schopenhauer’s Strategems**

• O’Second challenged McFirst’s reasons for believing that democracy is on the rise, but instead of addressing her challenge, McFirst, in his response, charges O’Second with arguing that democracy is not a good thing. In other words, he attacks a position that she never held—a straw man.

• At the same time, McFirst gives himself an easier position to argue for—not that democracy is historically on the rise or that we can predict it will spread but simply that it’s a good thing. And if we’re looking for fallacies, McFirst’s quote from Churchill might count as an emotion-loaded *appeal to authority*.

• Given a chance to improve his argument, McFirst focuses on the real points of disagreement. He backs up the data claim for his first conclusion—that democracy is increasing worldwide—countering O’Second’s appeal to ignorance with real information.

• He also seems to have backpedaled on his second conclusion a bit. He replaces his initial confident claim that the trend will continue with the more modest claim that there’s no reason to think it won’t.

**O’Second Casts Doubt**

• O’Second responds by noting that the American democracy arose from a particular historical background and set of precedents. Without that history, she says, democracy may not come as easily to other societies or it may come in a different form.

• This response by O’Second may be a little more subtle and a little harder to analyze. She begins with an Aristotelian syllogism:
“Democracy... real democracy... is important. Important things take real work. So democracy is going to take real work.”

- With that reasoning, she gives her initial challenge a different spin than we saw before. She argued earlier that O’First’s projection into the future was unjustified. Here, she says that we shouldn’t assume the inevitability of democracy because democratization will take real work.

- O’Second then uses a brief sketch of the American Revolution as the backdrop for another argument.
  - She starts with the claim that our democracy would not have the form it does without particular aspects of our history, such as the Protestant Reformation or our existence as colonies.
  
  - From that, she draws the inference that we should, therefore, not expect other paths to democracy in other cultures to look like ours. We should not expect the outcome to be the same.

In analyzing the statements of any speaker in a debate, ask yourself: (1) What conclusions are at issue? (2) What arguments are given? (3) Have any fallacies been committed?
○ How good is that reasoning? Although it’s suggestive, it’s not solid proof. She doesn’t prove that our institutions demanded those particular historical precedents, nor does she prove that other cultures with other histories couldn’t reach the same outcome.

○ If we look at her language carefully, she seems to recognize that this line of thought is merely suggestive. For example, she doesn’t say that democracy is bound to come in different forms; she says that it “may.”

○ What O’Second has given us, then, are some skeptical doubts and some suggestive counter-suggestions. In urging us not to immediately accept her opponent’s stirring forecast of a predictable rise in democratization worldwide, however, that may be enough.

Term to Know

appeal-to-authority fallacy: A fallacy in which the opinion of some prominent person is substituted for rational or evidential support; often used in advertising by linking a product to a celebrity “expert” rather than providing rational or evidential support for a claim about the product.

Suggested Reading

Fineman, The Thirteen American Arguments.
The Lincoln-Douglas Debates.

Questions to Consider

1. Who do you think won the debate: Mr. McFirst or Ms. O’Second? If that questions is too simple, explain why.
2. Consider how you think through an issue by yourself, and consider how an issue is addressed in a debate. In what way are those the same? In what ways do they differ?

3. Our adversarial system of justice has each side present its best case, with the opportunity to cross-examine the other side. The decision is ultimately made by some neutral party—a judge or jury. What are the pros of such a system with regard to finding the truth? What are the cons?

**Exercises**

You’ve undoubtedly watched a presidential debate live. Look for a video of a recent debate online (http://elections.nytimes.com [select the Video tab, then search for “Presidential Debates”] or search on YouTube). Watch the video, pausing after each contribution to analyze its strengths and weaknesses. Ask the same questions that we asked in this lecture:

(a) What precisely is the position?

(b) What precisely is the argument given?

(c) How good is that argument?

(d) Did the speaker commit any fallacies?
Advertising is ubiquitous. The average American sees 2 million television commercials in a lifetime, to say nothing of billboards, newspapers, magazines, and much more. The advertiser’s job is to get you to buy something or to vote for someone, not to encourage you to reason logically. Although philosophers sometimes analyze advertising in terms of logical fallacies, it is often more effective to think of advertising in terms of psychological techniques. Those techniques fall into two categories: The first relies not on information or content but on emotion, and the second does rely on content and does use information, but that information is used in misleadingly persuasive ways.

Attractiveness Attracts

- The people in television commercials are a strange breed. They’re far more attractive than the average person next door. Ads use good-looking people because attractiveness attracts. Regardless of what product is being sold or its quality, an attractive person will be used to sell it. The “attractiveness attracts” advertising strategy is used to appeal to your nonrational psychology rather than your logical rationality.

- To process advertising rationally, we have to be aware of, and compensate for, the fact that attractiveness attracts. In judging whether to buy what a commercial wants us to buy or to vote how a commercial wants us to vote, we need to filter out that aspect of persuasion that relies merely on an irrelevant attractiveness factor.

- Of course, it’s not just the people in advertisements who are at their attractive best. The settings are also pristine. People drink beer on gorgeous beaches that have no crowds or speak to you from clean, spacious kitchens. Car ads feature single cars gliding through isolated and beautiful surroundings. And next time you go to a fast food restaurant, look at the image pictured on the overhead menu.
Hold up what you purchase and see how representative that image is of what you just bought.

- Precisely because attractiveness attracts is a standard strategy, some advertising campaigns catch our attention by deliberately flaunting the strategy. For example, Frank Perdue was a skinny, bald man who reminded people of the chickens he sold. His very unattractiveness was used to convey sincerity and concern for quality.

- A related form of emotional manipulation is the appeal to prestige. Because an association with Europe can carry that prestige, in some products, we see foreign branding.

- Attractiveness attracts works not only for products but also for politics. Candidates for office are often filmed with dramatic lighting, positioned behind a prestigious podium and with an American flag in the background.

**Appeal to Emotion**

- The appeal to emotion is a classic fallacy. One of the places it shows up is in political advertising. There, the emotions most often exploited are pride and fear. A political commercial might show an American flag rippling in the breeze or pan across amber waves of grain with “America the Beautiful” playing on the soundtrack. Whoever the candidate and whatever the message, remember that what you’re seeing is a fallacious appeal to emotion.

- The other emotion exploited in political advertising is fear. A classic example is the television commercial used by Lyndon Johnson against Barry Goldwater in 1964. It opens with a little girl happily counting petals on a daisy. The count is then taken over by a male voiceover and becomes a countdown, culminating in the flash of a nuclear explosion. The implication was that Goldwater was a reckless man ready to push the nuclear trigger. This is an example of the fallacy of false alternative, but it’s also clearly an appeal to emotion: fear.
Better Thinking: Detecting Fallacies in Advertising

• Here is a simple philosophical tool for detecting the kinds of fallacies we’ve been talking about.

• Choose any advertisement. Read it, or listen to it, or watch it with care. Then sit down with a piece of paper and list the reasons given for buying the product; next, list the reasons given for buying this product rather than a competitor’s. A rational decision is one based on reasons.

• If the advertisement is relying merely on an attractive image, it may turn out that you can find no reasons to list at all. Sometimes you can list reasons, but once examined, it becomes clear that they aren’t very good reasons.

Appeal to Dubious Authority

• In the appeal to dubious authority, an instantly recognizable celebrity touts the virtues of a product, recommends it, or says that he or she uses it. What makes it an appeal to dubious authority is that you’re given no reason to believe that the celebrity knows what he or she is talking about.

• For example, a professional athlete may know a good deal about sports but may not know much about cars, watches, or brokerage companies.

Laws against False Advertising

• Most of the advertisements we’ve talked about so far rely on psychological manipulation rather than on claims or statements of fact. Some commercials do make claims, however, either explicitly or by implication.

• There are laws against false advertising. The core piece of federal legislation in this regard is the Lanham Act, which prohibits “any commercial advertising or promotion [that] misrepresents the nature, characteristics, qualities, or geographic origin of goods, services or commercial activities.” The Federal Trade Commission
is empowered to enforce regulations against “unfair and deceptive practices in commerce.”

- The law also allows consumers and rival companies to pursue civil action. The suit must prove a false statement of fact that could be expected to deceive, that it was likely to affect purchasing decisions, and that the plaintiff has suffered injury of some kind. Short of a class-action case, that kind of litigation is a tall and expensive order for any consumer. And, of course, these legal restrictions extend only to commercial advertising. Political advertising isn’t included.

**Not the Whole Truth**

- One way of misleading people is to tell them the truth but to be careful not to tell them the whole truth.

- Consider just a few of the phrases that tend to show up in advertising: (1) “none proven more effective” (true even if all competing brands have been shown to be equally effective); (2) “faster acting than the leading brand” (it is not obvious what the leading brand is, and it could be slower acting than any of the other contenders); and (3) “recommended by doctors” (it would be hard to find a pharmaceutical that wasn’t recommended by some doctors, but any implication of general medical consensus is a false implication).
Say the Magic Word

- Here’s a final category of advertising tricks: the magic word. There are a whole set of magic words that reappear in advertising. Some may not mean what you think they mean. Some may not mean much of anything at all. “As seen on TV” is a perfect example.

- Consider, for instance, foods labeled “low fat,” “reduced fat,” “light,” “98% fat free,” “low calorie,” “reduced calorie,” “lower sodium,” or “lite.” All of those labels sound great, but what do the words mean?
  ○ When you check the Food and Drug Administration guidelines, some of those terms carry absolute standards. A main dish counts as “low fat” if it contains less than 3 percent fat and not more than 30 percent of its calories come from fat.
  ○ But some of those terms are relative. A potato chip can boast of “reduced sodium” if it has at least 25 percent less sodium than a standard potato chip. The standard is higher for “light”: A can of beer, for example, must have 50 percent fewer calories from fat than the standard.

- Such standards often have no legal “teeth”; they are merely non-binding industry recommendations.

- There is also another problem here. If a product is labeled “low fat,” that means it contains a certain percentage of fat. It doesn’t mean it won’t make you fat. A product can be labeled “low fat” despite the fact that it is packed with calories, sugar, carbohydrates, and salt. The label that says “low salt” may speak the truth, but what it labels may still be unhealthful in other ways.

Test Your Skills

- As you look at advertising and try to identify what may be wrong or suspicious about ads you see, ask yourself these questions: What psychological buttons are being pushed? What is the precise claim that is being made? What relevant information may have been left out?
• Consider an imaginary ad urging voters to reject a gun control measure. Gun control is a difficult issue, worthy of rational discussion, but this ad merely stimulates fear.

• Another imaginary ad, for an energy drink, appeals to dubious authority by using a woman in a lab coat as the spokesperson and referring to what physicians say about fatigue and energy.

• In order to outwit the advertiser, we have to see through the attractiveness attracts appeal; the appeals to prestige, emotion, and dubious authority; misleading half-truths; and magic words. In other words, we have to learn to see through the spin. In the next lecture, we’ll talk about spin again, not in advertising but in putting a spin on statistics.

Terms to Know

appeal-to-prestige advertising strategy: Linking a product to status symbols in order to enhance the product’s desirability. Foreign branding is a specific form of the appeal-to-prestige strategy.

“attractiveness attracts” advertising strategy: Enhancing the desirability of a product by using spokespeople who are more attractive than average, by showing the product in beautiful settings, or by doctoring the product’s image or the product itself so that it is more photogenic. The flip side, “unattractiveness detracts,” is common in political campaign literature, which often seeks unflattering images of the opponent.

foreign branding: In advertising, associating a product with a foreign country in order to increase its desirability; a particular type of prestige advertising.

Suggested Reading

Packard, The Hidden Persuaders.

University of Illinois Advertising Exhibit.
Questions to Consider

1. You can probably think of an advertisement that influenced you to purchase the product. What was it about the advertisement that influenced you? Were you influenced on the basis of information and argument or something else?

2. Think of an advertisement that you think definitely didn’t influence you to purchase the product—that perhaps convinced you not to. What was it about the advertisement that had that effect?

3. What legal limits should be put on advertising? Should advertising be unregulated? Should there be a restriction on “truth in advertising,” penalizing anything other than the literal truth? Should there be restrictions beyond that, on implication or suggestion? How far would be going too far?

Exercises

Next time you are in a fast food restaurant, look first at the image of the food portrayed on the menu board. When your food comes, take it out of its wrapping and hold it up for comparison. How do the two differ?

The Clio awards are the Academy Awards of advertising. You can find samples and Clio winners at http://www.clioawards.com/. Take a look at last year’s winners—choose the radio ads, perhaps, and focus on ads you are not already familiar with. Construct your own ratings by writing down answers to each of these questions:

(a) How entertaining is the clip?

(b) How much information does it give you about the product?

(c) The reasons given for buying the product were:

(d) Does the advertisement offer reasons to believe that this product is superior to the competition?

(e) Does the advertisement commit any fallacies?
P eople love statistics, and well they should. Statistics give us the straight facts, in numbers and graphs. We want our information manageable and easy to understand. But while statistics can give us the facts we need in ways that are easy to understand, they can also give us the facts in ways that are prone to mislead and can be manipulated. We should all be a bit more critical of statistics on all sides of an issue, our own included.

Statistics versus Facts

• The appeal of statistics is they are hard-edged and factual. Information conveyed numerically is, however, just like any other information: It may or may not be true. The unfortunate psychological fact is that people seem to be more inclined to believe a claim simply because it comes with precise numbers. That’s why numerical information spreads fast. But often the numbers get even more impressive—and even farther from the truth—as the information spreads.

• Consider this statistic: Every year since 1950, twice as many American children have been killed with guns. Joel Best is a sociologist at the University of Delaware and the author of a book on misleading statistics. He nominates this as the most inaccurate social statistic ever.

• Suppose it were true that child deaths due to guns had doubled each year. Start with just 1 child killed with a gun in 1950. If the number doubles each year, that means there would have been 2 children killed in 1951, 4 in 1952, 8 in 1953. By 1960, the number would have been 1,024. It would have passed 1 million in the year 1970 alone. By 1980, there would have been 1 billion children killed with guns—more children killed than there were people living in the United States in 1980.
• Best first encountered this claim in a dissertation prospectus. But the dissertation cited a source, and he found the statistic in the source, too. It came from a 1994 report by the Children’s Defense Fund. But what the source actually said was: “Since 1950, the number of American children gunned down every year has doubled.”

• Consider those two sentences closely. Do you see the difference? What the original report actually said was that the number of gun-related child deaths doubled between 1950 and 1994.

• Between 1950 and 1994, the U.S. population almost doubled, so the change in gun-related child deaths per capita wasn’t particularly dramatic. Just by a different placement of the phrase “every year,” that statistic mutated into the outrageous claim that the number of deaths doubled every year.

Evaluating Statistics Critically: Sampling Size, Survey Bias

• Statistics represent some kind of counting, and counting is about as objective as things get. But in evaluating statistics, you have to pay attention to who did the counting and how. It’s also important to pay attention to the sample and where it came from. The perennial problem is that the basic sample may turn out to be unrepresentative. The bigger a percentage in the sample, the more confident you will be that the sampling reflects the true reality.

• The classic case here was Al Gore versus George W. Bush in the election of 2000. Around 8 pm on election night, most stations had called Florida for Gore. Given that only a small portion of the voting precincts had officially reported by then, how was that
early projection made? The answer is exit polling. In exit polling, pollsters pick a certain percentage of precincts to sample. Out of tens of thousands of voting precincts in the United States, at best, a few thousand are sampled, with perhaps 100 voters per precinct.

- We can see how that could go wrong. By 9:30 pm, with more precincts in, using their own exit polling, the Bush team challenged the network results. By 9:55, all the networks had changed their verdict from “Gore wins” to “too close to call.”

- A more important problem, however, was that the sampling space for the exit surveys might have been unintentionally biased. Pollsters didn’t sample all precincts in Florida, and the choice of which precincts to sample may have slanted the results. Asking for volunteers may itself have slanted the results.

- Questionnaires often come with their own biases. A major issue is the leading question. Consider the way these questions are phrased: “Do you agree with most Americans that it’s time we got a handle on decay in our inner cities?” “Do you think it’s important to back our troops by giving them the support they need?”

- Think what it would take to answer those questions in the negative. If the statistics we’re dealing with are taken from that kind of questionnaire, the data are biased simply by the way the question was asked. And yet, when results are reported, you may have no way of knowing that the original questions were loaded.

**Test Your Skills**

- In a graph of the number of drivers in fatal crashes by age group for a typical year, we see there were 6,000 fatal crashes for 16- to 19-year-olds. For 20- to 24-year-olds, the number jumps to almost 9,000. From there, it tapers down by five-year intervals. Those over 79 counted for fewer than 1,000 fatal crashes. These data suggest that people between 20 and 24 are a major hazard on the road.
• The data in the chart may be entirely accurate, but the conclusions we’re tempted to draw from them are not. There aren’t very many people over age 79 who drive, and thus, there are fewer drivers of that age to be involved in fatal crashes. The same thing is true at the bottom of the scale. Because not all 16- to 19-year-olds drive, there are fewer of them on the road. The sampling, then, is uneven.

• A better graph is the number of fatal crashes per miles driven by people in different age categories. Those numbers give us a very different graph. It becomes clear that people on the extremes of the age spectrum aren’t the safest.

Mean, Median, and Mode

• In order to condense data into a single number, some information is inevitably left out. But what information to emphasize and what information to leave out is always a choice. The difference between mean and median is a good example.

• The mean of a series of numbers is what most people think of as the average. Statistics are often reported in terms of the mean. But that can be misleading.
  ○ Consider the salaries of 1985 graduates of the University of North Carolina immediately upon graduation. Which graduates could boast the highest salaries, those in computer science or business?

  ○ Actually, the graduates with the highest mean salary were those in cultural geography. The fact is that there were very few people who graduated with a degree in cultural geography, but one of them was the basketball star Michael Jordan.

  ○ A major problem with using average or mean as a measure is that it is vulnerable to the problem of outliers. One enormous salary in a group can lift the mean in misleading ways.

  ○ Here’s a simple example: Imagine that there are nine people in a company. Four of them earn $20,000 a year; one makes
$30,000; and three earn $40,000. The boss gives himself a salary of $200,000. In this sample, the mode is $20,000, the mean is almost $48,000, and the median is $30,000.

- Means, modes, and medians may give you different numbers, and people on different sides of an issue, pushing a different line, will make different choices about what numbers to use.

**Statistical Distribution**

- The truth is that anything that compresses a complex social reality into a single number loses information. If you want more information, you want to see how the numbers distribute across the sample. That demands something more like a graph. But here again, there are possibilities for misrepresentation.

- Consider two graphs that claim to show the changes in violent crime between 1993 and 2010. One shows a slight decrease over time, and the other shows a dramatic decline. Which one is right? The answer is that they’re both right—they portray precisely the same data—but the difference is in how they portray it. The choice of scale on a graph can exaggerate or minimize an effect.

**Correlation**

- Sometimes statistics concern not just one thing but a relationship or correlation between two. Consider a graph with two lines, one representing the number of homicides per 100,000 people between the years 1960 and 1970 and one representing the number of people executed for murder in that period. The graph shows that as the number of executions between 1960 and 1970 steadily falls, the number of homicides steadily rises.

- The two lines go in opposite directions. It is precisely as the execution rate falls that the homicide rate increases. In one of the earliest sophisticated statistical studies of deterrence, Isaac Ehrlich used data on executions and the homicide rate to conclude “an additional execution per year…may have resulted in seven or eight fewer murders.”
But this is where an important distinction comes in. What the graph shows is correlation. Ehrlich’s conclusion, on the other hand, is phrased in terms of **causation**: Murder rates go up because execution rates go down.

Correlation and causation are not the same. At best, such statistics yield a correlation. That is important in judging causality, but correlation alone never provides causal information.

### Terms to Know

**causation**: The act of producing an effect. In a cause-and-effect relationship, the presence of one variable, the effect, can be established as a direct result of the other variable, the cause.

**correlation**: A direct or inverse relationship between two variables. In a direct correlation, the strength or frequency of both variables increases proportionately to each other; in an inverse correlation, the strength or frequency of one variable increases as the strength or frequency of the other decreases. However, correlation does not establish that one variable causes the other; one variable is not necessarily the cause and the other the effect.

**mode**: The most frequently occurring value in a set of numbers.

**outlier**: An extreme value; something that is out of the ordinary. In statistics, a number that is extremely divergent from all the others in a set of numbers; outliers have misleading impact on the mean, or average, of a set of numbers.

### Suggested Reading

Best, *Damned Lies and Statistics*.

———, *More Damned Lies and Statistics*.

Huff, *How to Lie with Statistics*. 
Questions to Consider

1. From your own experience, is there a case where your views were particularly shaped by statistics? What were the crucial statistics, and how did they influence you?

2. What precisely is the difference between these two sentences: (1) Every year since 1950, the number of American children gunned down has doubled. (2) Since 1950, the number of American children gunned down every year has doubled.

3. Why do you think it is that an appeal to statistics seems so persuasive? What aspect of the power of statistics is deserved? What aspect of their power is not?

Exercises

Eleven children take a quiz, with scores between 1 and 10. Construct a series of scores so that each of these is importantly different:

(a) The mean score (the average)

(b) The median score (there are as many scores above as below)

(c) The mode (the score that most children got)

(See “Answers” section at the end of this guidebook.)

“If you take the drug Doomleva, your chances of catching Doomlitis are cut in half. A very small portion of the population has been shown to carry a genetic allergy to Doomleva, which can cause sudden death.” What further statistics do you need in order to know whether you should take Doomleva?
We gamble all the time. Although we may not call it gambling—we reserve that term for high-risk games in casinos—we are always playing the odds. For example, we play the odds as individuals when we decide what job to take, or where to live, or whether to marry a particular person. As a culture, we gamble all the time, too. Science stands as one of the greatest intellectual achievements of all time. But outside of mathematics, none of our scientific conclusions is a certainty. Our confidence in those predictions is all based on a batch of evidence—we hope good enough—and a best guess on the basis of that evidence.

**Being Rational about Probability**

- As long as we live and breathe, we’re forced to play the odds, to gamble. Our concern is whether or not we gamble rationally, and we often don’t. **Probability** is another name for what we’re talking about here. We often don’t handle probabilities rationally. For example, which is more dangerous, a handgun or a swimming pool? Say that you have one neighbor with a swimming pool and another neighbor who keeps a gun in the house. Which is the safer place for your child to play?

- If we take the statistics seriously, swimming pools are far more dangerous. About 550 children drown in swimming pools each year—1 per every 11,000 residential pools. About 175 children are killed by guns each year—1 for every 1 million guns in the United States. The probability of a child drowning in a swimming pool is more than three times greater than the probability of a child being killed with a gun.

- Why do we tend to think otherwise? Imagine kids playing in a swimming pool: That’s a wonderfully positive image. The image of kids with guns, on the other hand, is frightening. When it comes to probabilities, we tend to think in terms of images, and it’s those images that often lead us astray.
Law of Large Numbers

• Calculating probability works like this: Imagine you have a quarter in your hand. As we know, the chance that it will come up heads when you flip it is 50/50, or a probability of 1/2. The probability is the ratio of the number of possible outcomes in your chosen category (heads) over the total number of possible outcomes.

• The basic principle of probability, however, demands equiprobable outcomes. That works well for standard games of chance. Cards, dice, and roulette wheels are made for easily calculated equiprobable outcomes. But we also want to use probability in real life, and real life isn’t so straightforward.

• Here’s an illustration with coin flips: We know that the probability of a quarter coming up heads is 1/2. If you flip a coin 20 times, how many heads will you expect to get? If you said about 10, you’re right. You shouldn’t expect to get exactly 10 heads. The perfect division of half and half is unlikely with small numbers. But as the number of flips gets larger, the probability increases that the result will be close to a 50/50 split.

• This is known as the law of large numbers. In the long run, frequencies of heads and tails will tend to even out.

Linking Probability and Frequency

• The law of large numbers matters, especially in real life, because it links probability with frequency. If the probability of heads is 1/2, then in the long run, we can expect about half the flips to come
In other words: If the probability is 1/2, the long-range frequency will approach 1/2.

- In real life, we estimate probabilities by reasoning in the other direction—from frequencies to estimated probabilities. For example, what’s the probability that a hurricane will hit New Orleans this year? By one reckoning, New Orleans has been hit by hurricanes 61 times in the last 139 years. That’s a frequency of 61/139, or about 43/100. Reading our probability off that frequency, we’d say the probability of a hurricane hitting New Orleans this year is about 40 percent.

Combined Probabilities
- After an airplane crash, the FAA wants to know not just the probability of hydraulic system failure in a plane, not just the probability of landing gear problems, not just the probability of pilot error, but the probability of any of those things. For that kind of case, we need to be able to calculate combined probabilities.

- Suppose you’re playing stud poker. You have to stick with the five cards you’re dealt. If we know the probability that you’ll be dealt a straight of some kind, and we know the probability that you’ll be dealt two pairs, we calculate the probability that you’ll be dealt one or the other by simply adding those probabilities together. The probability of a straight is a little more than 0.39 percent. The probability of being dealt two pairs is significantly better—just shy of 5 percent. Thus, the probability of getting one or the other is just shy of 5.39 percent.

- There is an important proviso to the simple addition rule, however. Adding probabilities this way works only if the two events are mutually exclusive—only if we know they can’t both happen.

- That’s the simple rule for “or.” There’s an equally simple rule for “and.” If we know the probability of drawing a red card from deck 1, and we know the probability of drawing a black card from deck 2, can we calculate the probability of doing both? In that case, we
multiply. The probability of drawing a red card from deck 1 is 1/2; the same is true for drawing a black card from deck 2; thus, the probability of doing both is $1/2 \times 1/2 = 1/4$.

- The simple rule for “and” comes with a proviso, just like the simple rule for “or.” You can figure the probabilities for two things happening by multiplying their individual probabilities as long as they are independent events.
  - What does it mean to say that two events are independent? They are probabilistically isolated. There are no strings of influence between them. The fact that one happens doesn’t affect whether the other will.

  - For example, successive flips of a coin are independent. If you flip a coin once, the chance of heads is 1/2. If you flip it a second time, the chance of heads is again 1/2. What comes out on the second flip doesn’t depend on what came out on the first flip. The two flips are independent events. The quarter doesn’t remember what it did last time. As far as it knows, each flip is a new day—a clean slate.

**Gambler’s Fallacy**

- To gamble rationally, you must know whether or not you’re betting on independent events. But we know that people often don’t bet rationally. Treating events that are independent as if they aren’t is called the gambler’s fallacy.

- For example, Mr. and Mrs. Smith have five children, all boys. They would like to have a girl, and they believe that with five boys in a row, their next child is bound to be a girl. Of course, that’s not true. The reproductive system doesn’t have a memory any more than the coin does. The gender of a couple’s next child is an independent event.

- Sometimes, people offer a little more reasoning: With five boys in a row, the next one’s bound to be a girl. What are the chances the Smiths would get six boys in a row?
• It’s true that the chances of having six boys in a row are small. After all, if the odds are 50/50 between having a boy and having a girl, the chance of having six boys is \( \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{64} \).

• But that’s not the question. The question is what are the chances that the next child will be a boy? And the chances of that are precisely the same as the chance that the first one was going to be a boy. Each birth is an independent event. Mr. and Mrs. Smith have committed the gambler’s fallacy, treating independent events as if they aren’t independent.

A Classic Example

• If you use the full canon of the more complicated laws of probability, it is possible to calculate combined probabilities even when events are not mutually exclusive, or not independent. The following is a classic example.

• You find yourself with 30 or 40 other people in a restaurant, office, or gym. What are the chances that two of those people were born on the same day of the year? In fact, it’s very probable you would have two people born on the same day of the year.

• The calculation might go like this: We pick person 1, who was born on October 15. We then pick person 2. The probability that person 2 was born on a different day is fairly high: \( \frac{364}{365} \). With person 3, the odds are down to \( \frac{363}{365} \). As we add people, we keep multiplying: \( \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \times \frac{361}{365} \ldots \) and so on.

• With just 23 people, the multiplication boils down to 50/50 odds that those people will all have different birthdays. In a room with only 23 people—the size of a small classroom—the odds are 50/50 that two people share the same birthday. At 30 people, the probability reaches 70 percent that two birthdays match. At 35, it becomes an 80 percent probability; with 50 people, the probability of a match hits 97 percent.
• Life is full of decisions, and rational decision making demands knowing the odds. But good decisions also demand a lot more than that. In the next lecture, we’ll explore some of the lessons and some of the limits of decision theory.

Terms to Know

**combined probability**: The rules for calculating the odds that two or more events will happen. The probability of either one or another of two mutually exclusive events happening can be calculated by adding their individual probabilities. The probability of two independent events both happening can be calculated by multiplying their individual probabilities. Other rules apply for dependent events.

**equiprobable**: Having an equal mathematical or logical probability of occurrence.

**gambler’s fallacy**: Treating independent events as if they were dependent events. Someone who thinks black is “bound to come up” in roulette because of the appearance of a string of blacks (or reds) has committed the gambler’s fallacy.

**independent events**: Events that are isolated probabilistically; the fact that one event happens is not linked to and will not affect the fact that the other event happens.

**law of large numbers**: In probability theory, the fact that as the number of trials increases, the outcome will approach the mathematically expected value. For example, in a situation where there are two equiprobable outcomes, such as heads or tails when flipping a coin, the longer the run, the more likely the outcome will be a 50/50 split.

**mutually exclusive events**: Events are mutually exclusive if the presence of one categorically excludes the presence of the other; both cannot occur.

**probability**: The ratio calculated by dividing the number of possible outcomes in a particular category by the total number of possible outcomes.
Aczel, *Chance*.

David, *Games, Gods and Gambling*.

1. You draw a card from a standard deck. What is the probability that it is both red and has an even number on it? Convince yourself that these two approaches give you the same answer:
   (a) Count the cards that fit the bill: 2 of hearts, 4 of hearts, and so on. Put that total over 52.
   (b) The probability of drawing a red card is 1/2. The probability of drawing a card with an even number is 5/13. Thus, the probability of drawing a card that is both red and has an even number is $1/2 \times 5/13 = 5/26$.

2. Do you know anyone who has committed the gambler’s fallacy? Have you?

3. What do you think that John Maynard Keynes meant by “In the long run, we’re all dead”?

### Exercises

Your cat just had 4 kittens. Which of the following do you think is more likely?
(a) All are of the same sex.
(b) Half are of each sex.
(c) Three are of one sex, and one is of the other.
First, give your intuitive answer. Then, calculate the answer by laying out all 16 possibilities. We can lay them out like this, left to right in terms of the first kittens to appear:

<table>
<thead>
<tr>
<th>MMMM</th>
<th>FMMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMMF</td>
<td>FMMF</td>
</tr>
<tr>
<td>MMFM</td>
<td>FMFM</td>
</tr>
<tr>
<td>MMFF</td>
<td>FMFF</td>
</tr>
<tr>
<td>MFMM</td>
<td>FFMM</td>
</tr>
<tr>
<td>MFMF</td>
<td>FFMF</td>
</tr>
<tr>
<td>MFFM</td>
<td>FFFM</td>
</tr>
<tr>
<td>MFFF</td>
<td>FFFF</td>
</tr>
</tbody>
</table>

In how many of those cases are the kittens all of the same sex? In how many cases are there half of each sex? In how many cases are there three of one sex and one of the other?
S

ome times there are just too many options, as in the cereal aisle of
the supermarket. Choice is a good thing, but more options may not
necessarily make us happier. We obsess over a perfect or optimal
solution and get analysis paralysis. When more important decisions are at
stake, ignorance is a major problem. What will happen if we take this or that
course of action? Are we making things better or worse? If the context didn’t
start out emotional, it may end up that way. This lecture is about how we can
make decisions more rationally.

**Decision Theory**

- **Decision theory** builds on the probabilities we discussed in the last
  lecture. Here’s an example: As a gift, someone has given you your
  choice of tickets in two local lotteries. Lottery A offers a 2 percent
  chance of winning $100. Lottery B offers a 1 percent chance of
  winning $150. Which one should you take?

  - Decision theory tells you to take lottery A. In lottery A, you
    multiply the 2 percent probability of winning by the $100 you might
    win: $0.02 \times$100 = $2. That $2 is called the **expected utility** of
    picking lottery A. In lottery B, you multiply your 1 percent chance
    of winning by the $150 you might win: $0.01 \times$150 gives you an
    expected utility of $1.50.

  - Decision theory comes with a slogan: Maximize expected utility.
    An expected utility of $2 beats an expected utility of $1.50; thus,
    you should choose lottery A.

**Desirability**

- According to decision theory, the rational way to make decisions
  is on the basis of expected utility, calculated by multiplying
  probability by desirability. However, in making decisions, what we
consider desirable is also influenced by tags or labels that make us think we should consider certain things more desirable.

• In a study by researchers at Stanford and the California Institute of Technology, volunteers were asked to taste a series of wines. The volunteers weren’t wine experts but moderate wine drinkers. The wines were labeled by price: $5, $10, $35, $45, and $90 per bottle. But because this was a psychological experiment, the wine in the bottle labeled $10 and in the bottle labeled $90 was exactly the same.

• Desirability followed the labels rather than the wine itself: Subjects preferred a wine when it came with a more expensive label. In fact, in brain scans of the participants, pleasure centers in the brain showed a stronger response to the wine labeled $90 than to the wine labeled $10, even though they were exactly the same wine.

Decision Theory by the Numbers
• Let’s look at a more complicated example, in which the values aren’t purely monetary, and our options have downsides, as well as potential winnings. Imagine that you’re invited to someone’s house for dinner. You volunteer to bring the wine, but should you bring white or red?

• According to decision theory, you need to multiply probability by desirability. We’ll do that with a matrix, working with desirability first.

<table>
<thead>
<tr>
<th>Type of Wine</th>
<th>Fish or Chicken</th>
<th>Beef or Pork</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>Right wine</td>
<td>Wrong wine</td>
</tr>
<tr>
<td>Red</td>
<td>Weird choice</td>
<td>Right wine</td>
</tr>
</tbody>
</table>

• In order to apply decision theory, we need to translate those values into numbers. The right wine will be worth +10 points; the wrong wine will be –10. We’ll score “weird choice” in the middle, at 0. If we plug in those numbers, we have a desirability matrix:
Decision theory demands that we now add probabilities. What is the probability that your hosts will serve fish or chicken versus beef or pork? Perhaps you know enough about your friends to know that they’re health-conscious and might favor fish or chicken for that reason. But then again, it’s a special occasion; they might serve beef. You end up guessing that the probability is 70 percent that your hosts will serve fish or chicken.

The probability of fish or chicken is the same, no matter what kind of wine you buy. Thus, we multiply both entries in that column by 0.70. You estimated only a 30 percent probability of beef or pork; thus we multiply both entries in that column by 0.30. Remember, decision theory demands that we multiply probability by desirability. The results give us a matrix of values from +7 to –3.

For the final stage, we look at the problem from your perspective. You don’t have any control over what’s for dinner. You have control only over the choice of the wine. What we want to do is combine your potential gains and losses for each of your options.

If you chose white wine, you add together the +7 points from the case of fish or chicken with the –3 points for the case of beef or pork. Your final expected utility for the white wine option is 4 points.
• If you choose red wine, you add the 0 for fish or chicken with +3 for beef or pork, giving you a final expected utility of 3 points.

• Again, the goal is to maximize expected utility, and 4 beats 3. Decision theory says to go with the white wine.

**Tversky and Kahneman**

• Some of the case studies in this lecture come from the work of Daniel Kahneman and Amos Tversky, for which Kahneman received the Nobel Memorial Prize in Economic Sciences for his work on decision theory (Tversky had died by the time of the award). Interestingly, both Kahneman and Tversky were psychologists, not economists.

• Why do people make the decisions they do? Tversky and Kahneman put it this way: Choices involving gains are often **risk averse**. Choices involving losses tend to be **risk taking**. That holds true even though the only difference is in how a question is phrased. Phrase the question in terms of gains, and we decide one way. Phrase it in terms of losses, and we decide a different way.

• Kahneman received the Nobel Prize for **prospect theory**. Decision theory is all about how people should make decisions. It’s a normative theory: It tells you what you rationally should do. Prospect theory is a descriptive theory: It’s about how people actually make decisions.
  ○ Decision theory dictates a straight line on a graph, going through the 0 point at 45 degrees. Losses and gains on this graph are perfectly symmetrical; a gain of 2 perfectly counterbalances a loss of 2, for example.

  ○ With prospect theory, the same graph has a curvy line, intended to track how people actually view gains and losses. Losses count negatively more than correlate gains count positively. The graph indicates that this is particularly true when smaller rather than larger gains are at stake.
• Sometimes the decisions we make are irrational, for precisely the reasons that prospect theory points up. Sometimes we miscalculate probabilities. Sometimes we think in images instead of taking statistics seriously. Sometimes context leads us astray about what we really value.

• But decision theory, at least as we know it, may also have important limits. Tversky and Kahneman are very careful not to say that the gain-loss differences in valuation are necessarily irrational.

**Diminishing Marginal Utility**

• Economists have long recognized the concept of *diminishing marginal utility*. Additions of one unit are not always additions of one unit of value. As the units mount up, their individual value may diminish. One dollar is probably not worth as much to a billionaire as to a homeless person.

• Diminishing marginal utility doesn’t threaten the foundations of decision theory. All the calculations we’ve given are written in terms of utilities: probabilities multiplied by what is really valued, to the extent it is valued. If dollars have diminishing marginal utility, we won’t always be able to use a straight monetary scale as if it were a straight value scale.

• The lesson of Tversky and Kahneman may be that we have to do something similar for gains and losses. In at least some contexts, it may be that losses are rationally feared more than gains measured in the same units.

**Pascal’s Wager**

• Decision theory goes in a surprising direction envisaged by one of its founders, Blaise Pascal, who provided an argument for the existence of God. Pascal’s argument is interesting because it isn’t really an argument that God exists; rather, it’s a decision-theoretic argument. Whether God does or does not exist, Pascal argues, it is rational to believe in him. The argument is known as *Pascal’s wager*. 
• Pascal’s argument can be framed as a matrix, exactly like the ones we used before. Across the top, we have two possibilities—that God exists and that he doesn’t. Down the side, we have two other possibilities—that you believe or that you don’t. That gives us four spaces to fill in. What will happen in each of those combinations?

• Suppose you don’t believe that God exists, and you’re right. We put a value of 100 in the lower right box. Suppose that you believe that God exists, and you’re wrong. We enter a value of –10 in the upper right box. Now suppose that God exists, and you believe in him. You stand to gain infinite bliss, so we put an infinity sign in the upper left box. Suppose that God exists, and you don’t believe in him. We enter a negative infinity sign in the lower left box.

<table>
<thead>
<tr>
<th>Personal Belief</th>
<th>God Exists</th>
<th>God Doesn’t Exist</th>
</tr>
</thead>
<tbody>
<tr>
<td>I believe</td>
<td>$+\infty$</td>
<td>$-10$</td>
</tr>
<tr>
<td>I don’t believe</td>
<td>$-\infty$</td>
<td>$+100$</td>
</tr>
</tbody>
</table>

• The expected utility of the “I don’t believe” option is $-\infty + 100$, which still equals $-\infty$. That’s the expected utility of disbelief. The expected value of the “I believe” option is $\infty - 10$, which still equals $\infty$. That’s the expected utility of belief.

• If you are rational, Pascal surmises, you will maximize expected utility. In this case, you have an infinite everything to gain in the case of belief and an infinite everything to lose in the case of disbelief. The rational option is belief.

**Terms to Know**

**decision theory**: A theory of how to make rational decisions by maximizing expected utility.

**diminishing marginal utility**: The concept that units of value may not amass in equal increments; additions of one unit may not always be the same
as additions of one unit of value. For example, one dollar may not be worth as much to a billionaire as to an individual who has no savings at all.

**expected utility**: In economics, calculated by multiplying the potential benefit of an outcome by its probability.

**Pascal’s wager**: A decision-theoretic argument offered by Blaise Pascal in the 17th century with the conclusion that it is rational to believe in God. If you believe in God and he exists, your payoff is eternal bliss; if you believe in God and he does not exist, your payoff is being wrong, a small loss; if you do not believe in God and he does not exist, you have the pleasure of being right, a small gain; if you do not believe in God and he does exist, your payoff is eternal damnation. Given the potential gains and losses, expected utility dictates that the most rational thing to do is believe in God.

**prospect theory**: In the work of Daniel Kahneman and Amos Tversky, a descriptive theory concerning how people make decisions in terms of prospective loss and gain rather than final outcome alone.

**risk aversion**: In economics, an investor is considered risk averse if he or she avoids risk as opposed to being risk taking. People’s choices involving possible losses are often more risk averse than similar choices regarding possible gains.

**risk taking**: In economics, an investor is considered risk taking if he or she is not opposed to taking risks. Choices involving possible gains are often more risk taking than similar choices regarding possible losses.

**Suggested Reading**

Kahneman and Tversky, “Prospect Theory.”


Resnik, *Choices: An Introduction to Decision Theory.*
Questions to Consider

1. Suppose you really were invited to someone’s house for dinner and asked to bring the wine. But you don’t know what will be served, can’t contact your host by phone, and don’t have enough money for two bottles. Should you bring red or white? Use your own estimate of probabilities and desirabilities in order to explain your answer.

2. How good would you feel if you found a $100 bill on the street? How bad would you feel if you found that a $100 bill had accidently slipped out of your wallet? It’s $100 either way, but people tend to rate potential gains differently from potential losses. Do you think it’s rational or not to have different reactions in these two cases?

3. Here is the matrix for Pascal’s wager. Do you think that it should convince someone to believe in God?

<table>
<thead>
<tr>
<th>Personal Belief</th>
<th>God Exists</th>
<th>God Doesn’t Exist</th>
<th>Expected Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>I believe</td>
<td>$\infty$</td>
<td>$-10$</td>
<td>$+\infty$</td>
</tr>
<tr>
<td>I don’t believe</td>
<td>$-\infty$</td>
<td>$+100$</td>
<td>$-\infty$</td>
</tr>
</tbody>
</table>

Exercise

Your company has the option of acquiring two new lines out of three options. Each will involve a $100,000 investment, and the company can afford only two of the three. You have to choose which two to take:

Option 1: You estimate that this option has a 20 percent chance of success and an 80 percent chance of failure but a $1 million payoff over your investment if it succeeds.

Option 2: You estimate that this option has a 40 percent chance of success, a 60 percent chance of failure, and a $500,000 payoff over your investment if it succeeds.
Option 3: You estimate that this option has a 90 percent chance of success and only a 10 percent chance of failure but will return only $150,000 over your investment if it succeeds.

Which two options would you take? Explain your reasoning in terms of probabilities multiplied by desirabilities, plus any other considerations you think are relevant. (See “Answers” section at the end of this guidebook.)
Why is it important to think scientifically? Thinking scientifically means putting claims rigorously and systematically to the test. It means holding beliefs about the world up to the world in order to see whether those beliefs are true or not. Just because someone calls a discipline a science doesn’t make it so. If it’s real science, it’s science because of the procedures of rational test.

Science versus Pseudoscience

- We can easily differentiate between science—physics, chemistry, microbiology, neurophysiology—and pseudoscience—astrology, UFOlogy, phrenology. In certain cases, however, people disagree about these categories; an example is parapsychology. What precisely is the difference between science and pseudoscience?

- The 20th-century philosopher most responsible for concentrating on this problem—and for presenting one of the most durable and well-known answers to it—was Karl Popper. Popper wrote, “I wished to distinguish between science and pseudoscience; knowing very well that science often errs, and that pseudoscience may happen to stumble on the truth.”

- How do we distinguish between science and pseudoscience? There were two answers to Popper’s question that were in circulation at the time: an old answer and a new one. The old answer was that what distinguishes science is that it is inductive; it proceeds from experience to theoretical generalizations. That does seem to be true of science: It builds from individual experiences to general theories. But it’s not clear that that is what distinguishes science from pseudoscience.

- Popper was dissatisfied with the old answer. But there was also a new answer current in Vienna during his time, an answer associated
with a group of scientists and philosophers called the Vienna Circle. The young mathematician Kurt Gödel was a member, and Wittgenstein’s *Tractatus Logico-Philosophicus* was a highly respected source. Their answer to Popper’s question was that the mark of science was **verifiability**.

### Falsifiability Is the Key
- Popper didn’t think that that answer was any better. His objection was that it seemed all too easy to verify a theory. Popper’s conclusion was that this new answer wasn’t adequate to distinguish real science from its pretenders either. Popper proposed an alternative: What a theory needed in order to be scientific wasn’t verifiability but **falsifiability**.

- A genuinely scientific theory, according to Popper, is one that takes risks. A theory that nothing would disconfirm is pseudoscientific. A genuinely scientific theory isn’t like that. It sticks its neck out. Here, Popper’s favorite candidate is Einstein’s theory of relativity.
  - Popper was working at just about the time of the Eddington expedition, intended to test Einstein’s theory. Einstein predicted that light would be bent by gravity. If the theory was true, the position of stars photographed during an eclipse should differ from their positions when photographed in the night sky, without the gravitational effect of the sun.

The history of the pseudoscience astrology demonstrates that the issue is not always falsification of theories but how people use theories—in this case, to make money.
If the Eddington expedition came back with photographs that didn’t show the effect, Einstein’s theory was dead in the water: falsified. It was precisely because Einstein’s theory was so clearly falsifiable that Popper took it as the paradigm of genuine science.

Unfalsifiable: The Opposite of Science

• Suppose you want to create a pseudoscience, one that is protected against falsifications precisely because it is unfalsifiable. Perhaps you claim that there is a very strange thing hovering in the distance: a “cereotopic IV (invisible) sphere.” There is nothing anyone can say or do to refute the theory. Unfalsifiability is built right in.

• Another way to create a pseudoscience is to make your theory a moving target. Every time someone tries to refute it, you say, “Oh, that wasn’t exactly my theory. More precisely, it is…,” and you change the theory a bit.
  ◦ That was the history of Ptolemaic astronomy. In simple form, the Ptolemaic theory held that the planets and the sun circle the earth. If that were true, all the planets would move uniformly across the night sky. But, in fact, they don’t. If you plot the positions of Mars against the constellations night by night, you’ll find that Mars moves in the same direction for a while but then starts moving backwards and then forward again. This movement is called retrograde motion.

  ◦ Thus, the theory was modified a bit; what is circling the earth aren’t planets but invisible pivots around which the planets turn. What you see in retrograde motion is the planet going backwards around that invisible point, called an “epicycle.”

  ◦ Such additions to a theory as epicycles are called post hoc modifications—modifications after the fact. If you add those often enough, you make your theory unfalsifiable by making it a constantly revised moving target.

• Yet another way to make a theory unfalsifiable is to build on ambiguous phenomena. Use something that is hard to see or
that can be interpreted in any way you want. In this category fall Percival Lowell’s reports of canals on Mars.

- Another tip for building a pseudoscience: Build in a stipulation that only the properly trained or the truly initiated can see at work.

- The point of these examples is to alert you to what pseudoscience is so that you will recognize it. Pseudoscience isn’t a problem only outside the scientific establishment; it is something to guard against even within the hallowed halls of established science—perhaps especially within the hallowed halls. Even the best of scientists can fool themselves by falling for the traps of pseudoscience.

**Putting Theories to the Test**

- When it comes to influential bits of 20th-century philosophy of science, Popper’s falsifiability criterion may top the list. Scientists themselves use it as a mark for good science. No theory of any breadth, Popper emphasizes, can ever be conclusively confirmed. There is no point at which we say “confirmed,” close the book, and abandon all future testing. We should continue to look for falsifications. We should continue to put our theories to the test. That’s the only way we’ll find better ones.

- The core demand of scientific rationality is that we test theories. How precisely do we do that? Let’s look at a simple case in what’s known as the Wason selection task: Imagine that you have four cards in front of you, labeled E, K, 4, and 7. Here’s the theory you want to test: If a card has a vowel on one side, it has an even number on the other. What is the minimum number of cards you have to turn over to test this theory?

- The most popular answer is to start with the E card. And, of course, that’s right. If we turn over the E and see an even number there, that’s in accord with the theory. If there isn’t an even number on the E card, the theory is refuted: falsification. That’s a good test. If it comes out one way, we get a confirmation. If it comes out another, we get a falsification.
• Are there any other cards that would test the theory? The second most popular answer is the 4 card. But that is definitely not right. Suppose we turn over the 4 card and find a vowel. Then we have a card that fits the theory. But what if we don’t find a vowel? If we turn over the 4 and don’t find a vowel, we still haven’t falsified the theory. All the 4 card can give us is confirmation. Turning over the 4 card can’t give us falsification; thus, it can’t really test the theory.

• The second card we should turn over is the 7 card. Suppose we find a vowel on the other side. Then we have a vowel on one side, an odd number on the other, and we’ve falsified the theory. That gives us a real test.

The Limits of Experimentation
• In this lecture, we’ve emphasized “putting it to the test,” or experimentation, as the core concept of scientific rationality, with an emphasis on the importance of seeking falsifications of a theory. There are a few warnings and provisos regarding the role of experimentation in science, however. There are many different sciences, and the role of experiment differs in each one.

• If an organic chemist forms a hypothesis regarding the composition of a specific protein, he or she can design an experiment and test it. But if a paleoanthropologist forms a hypothesis regarding the origins of Australopithecus afarensis, there is no clear and obvious experiment that would confirm or refute it. There are rival scientific theories about the formation of the moon, but there is no clear moon- forming experiment we can perform to decide between them.

• The role of experiment in science is crucial, but it is often unavailable. Sometimes, it takes years to invent an experiment that would confirm or disconfirm aspects of a theory. This was true both for parts of Einstein’s relativity and for quantum mechanics.

• When we do have experimental evidence for a theory, it is often indirect. We cannot afford, either economically or ethically, to perform wide-scale social experiments that would answer some
of our questions. The experimental evidence we have regarding such questions must be extrapolated from social psychological experiments involving small groups of people in relatively artificial circumstances.

- Scientific rationality demands that we put hypotheses to the test—any theory that waives away empirical test as irrelevant would clearly be an unscientific theory. But the role of experiment is often more elusive and indirect than it might seem.

**Terms to Know**

**falsifiability**: Karl Popper’s criterion for distinguishing real science from pseudoscience was that real scientific theories can be shown to be false. Real science takes risks in that theories are formulated so that they are open to disconfirmation by empirical testing or empirical data.

**pseudoscience**: A set of practices or tenets that is made to resemble science and claims to be scientific but violates scientific standards of evidence, rigor, openness, refutation, or the like.

**Ptolemaic astronomy**: The view of astronomy put forward by Ptolemy (A.D. 90–168); based on a geocentric (earth-centered) model of the universe.

**retrograde motion**: In astronomy, the apparent movement of a body opposite to that of other comparison bodies. Plotted against constellations night by night, Mars moves in one direction for a while but then appears to move backwards before it again seems to reverse directions. Retrograde motion does not fit with the predictions of a simple geocentric theory, which holds that the sun and planets circle around the earth.

**verifiability**: The criterion for what distinguishes science from non-science, proposed by a group of 20th-century scientists and philosophers called the Vienna Circle. Science, they said, is that which can be confirmed by empirical or experiential data.
**Wason selection task:** A psychological task used to examine the psychology of reasoning. If shown four cards, one displaying the letter E, one the letter K, one the number 4, and one the number 7, using a minimum number of cards, which cards do you turn over to test the theory “If a card has a vowel on one side, it has an even number on the other”? It has been demonstrated that people tend to seek confirmation when working with unfamiliar or abstract concepts, such as “vowels” and “even numbers,” but will seek falsifiability in similar tasks using familiar situations, such as underage drinking.

**Suggested Reading**

Gardner, *Fads and Fallacies in the Name of Science*.

Grim, ed., *Philosophy of Science and the Occult*.

Shermer, *Why People Believe Weird Things*.

**Questions to Consider**

1. You certainly don’t want your theory to be false. Why then should you want it to be falsifiable?

2. One of the points made in the lecture is that people make different decisions in different forms of the Wason selection task.  
   (a) Which cards should you turn over to put this theory to the test: If a card has a vowel on one side, it has an even number on the other?  
   
   ![Card Options](E K 4 7)

   (b) Which cards should you turn over to put this theory to the test: If you are drinking alcohol, you must be 18?  
   
   ![Card Options](Drinking beer Drinking Coke 18 16)

   In the first case, people tend to get the answer wrong; they say. “Turn over the E and the 4.” In the second case, they tend to get the answer right: “Turn over ‘Drinking beer’ and 16.”
The cases are logically identical. Why do you think people get the answer right in one case and wrong in the other?

3. Karl Popper’s criterion of falsifiability states that a theory is scientific only if it is falsifiable—only if there is some experience that would prove it wrong. Is Popper’s theory falsifiable? Is the criterion of falsifiability itself falsifiable? What conclusion should we draw if the answer is no?

Exercise

Below are some examples of theories. From your own familiarity, or with a bit of research, decide which of these you think belongs in which category:

(a) A theory that is falsifiable but hasn’t yet been falsified.

(b) A theory that is falsifiable and has, in fact, been falsified.

(c) A theory that is unfalsifiable.

- Erich Von Daniken’s theory that a small group of aliens was stranded on Easter Island in ancient times and taught the natives to make robot-like statues.

- Thor Heyerdahl’s theory that the Polynesian islands were populated not from Asia to the west but from ancient Peru to the east.

- The theory in Aristotle and earlier Greeks that everything is made of four essential elements: earth, air, fire, and water.

- The Ptolemaic theory that the sun orbits around the earth.

- The Lamarckian theory of evolution that an animal can pass on to its offspring characteristics that it has acquired during its lifetime.

- Erich Von Daniken’s theory that the stone from which the Maoi statues on Easter Island are made is not found on the island.
A n understanding of scientific experimentation, and of its limits, is an important part of the contemporary philosopher’s toolkit. This understanding is crucial both in backing your own claims and in evaluating the claims of others. In this lecture, we’ll talk about the structure of experiments—putting questions to the test.

A Beautiful Experiment

- On January 28, 1986, the Space Shuttle *Challenger* lifted off at the Kennedy Space Center, but 73 seconds into the flight, the *Challenger* disintegrated in the skies over the Atlantic, instantly killing all aboard. What caused the crash?

- To answer that question, President Ronald Reagan set up the Rogers Commission, which included Charles Yeager, Neil Armstrong, Sally Ride, and the physicist Richard Feynman. The most famous moment in the commission’s work occurred during a televised hearing, when Feynman performed a simple experiment to show that the O-rings sealing a joint on the right solid rocket booster could have failed under conditions of extreme cold.

The Well-Designed Experiment

- In order to establish an empirical claim, the gold standard is a well-designed experiment. The key, however, is the qualifier “well-designed.”
  - Consider again the structure of Feynman’s experiment. The commission was looking for a chain of causal events: A causes B, B causes C, and C caused the disaster.
  - It became clear that a blowby of pressurized hot gas and flame could have hit an adjacent external tank, leading to the structural failure. The manufacturer insisted that the O-rings
couldn’t have been what allowed the blowby. But Feynman’s experiment with ice water showed that indeed they could have.

- Note three things about Feynman’s experiment: the risk of the experiment, its power, and its limitations. Karl Popper insisted that every good scientific hypothesis takes a risk—a risk of being proven wrong. Feynman took that risk. His hypothesis about the O-rings might have been disproven in the public eye.

- But note also the limitations of the experiment. All that was shown is that the O-rings could have failed. The experiment didn’t show that something else might not also have been a factor—even a major factor. The experiment addressed only one link in the causal story that the Rogers Commission finally settled on: the link from O-rings to blowby. Feynman’s experiment, like all experiments, functioned only against a broader theoretical background—within a wider theoretical context.

**The Controlled Experiment**

- All experimentation is about an “if…then” statement. What any experiment really shows is “If this, then that.” Feynman’s experiment shows that if you put a C-clamped O-ring of this material in ice water, then its resilience fails.

- We need to know “If $X$, then $Y$.” But we also need to know “If not $X$, then not $Y$.” We need to know that $X$ is what makes the difference as to whether or not $Y$ happens. The core of any test of causality is, therefore, a double test. We need to have two cases side by side, one with $X$ and one without $X$. That’s the whole idea of a control and a controlled experiment.

**Randomized Controlled Trial**

- Another beautiful experiment was conducted in 1747 by James Lind, surgeon on H.M.S. *Salisbury*, in investigating the problem of scurvy. Lind tested the numerous proposed cures for the disease by selecting 12 men from his ship, all suffering from scurvy and as
similar as he could choose them. He then divided the men randomly into six pairs, giving each pair one of the cures.

- The important points here are that Lind chose men as similar as possible and that he randomized them into pairs. As it turned out, all the men showed some improvement. But the two who were given fruit recovered dramatically.

- Lind’s is an early example of what we now call a randomized controlled trial. In constructing a solid experiment with complex subjects, such as people, you want to make sure you have a random enough selection procedure within a large enough group so that your results are unlikely to have happened by chance alone.

- Complex experimental subjects—again, such as people—make randomized testing a necessity. There are also other ways in which
people make things complicated. These call for other constraints in experimental design.

Philosophical Understanding of Scientific Procedure

- The philosophical understanding of scientific procedure developed hand in hand with its application. We’ll look at three high points of this development: the work of Francis Bacon, John Stuart Mill, and R. A. Fisher.

- In 1620, Francis Bacon published his *Novum Organum*, intending his work to be a new organon, a new philosopher’s toolkit. The core of that new organon was an emphasis on “if…then” experimentation.
  - Bacon contrasts experimentation with other ways—what he takes to be erroneous or fallacious ways—of forming beliefs.
  - He pits empirical experiment against appeal to authority, against vague theorizing, and against appeals to experience without experimental manipulation.

- In 1843, the philosopher John Stuart Mill published his *System of Logic*, intended to offer a logic appropriate to scientific investigation. Mill gave a series of “methods,” all of which focus on the kind of double “if…then” test we have emphasized for deciding whether \( X \) causes \( Y \).
  - **Mill’s method of difference** is precisely that double test. If \( X \) is the only difference between two cases, with \( Y \) a result in one and not in the other, we have an indication that \( X \) causes \( Y \).
  - In **Mills method of agreement**, if two cases are alike only in \( X \), and \( Y \) occurs in both, we again have evidence that \( X \) and \( Y \) are causally linked.
  - In **Mill’s joint method of agreement and difference**, \( Y \) appears whenever we have \( X \) and disappears when we don’t.
  - We have talked as if \( X \) and \( Y \) were simply on or off, present or not. Mill also generalizes that to a case in which things might
be more or less $X$ or $Y$. This is **Mill’s method of concomitant variation**: If $Y$ varies as $X$ varies, we again have indication of a causal link.

- Beyond Mill, the most important improvements in our understanding of scientific procedure have been in the development of statistical measures for randomized testing. Here, much of current practice traces back to the English biologist R. A. Fisher.
  - Fisher’s work was motivated by a beautiful experiment of his own, known as the Lady Tasting Tea example. The topic was trivial, but Fisher’s experimental design and analysis are not.
  - Fisher’s experiment is essentially what all controlled randomized experimentation looks like today, from political polling to medical experimentation.

**The Double-Blind Experiment**

- Suppose you want to know whether Coke or Pepsi tastes better. You do a straightforward test: You buy a can of Coke and a can of Pepsi and offer people a sample from each. In order to improve your experiment better, you blind the taste testers so that they are unable to see the cola cans.

- There are cases, however, that call for an even more scrupulous experimental design. When a researcher strongly believes in a hypothesis, there is always the chance that he or she will influence the results by giving unconscious clues to the experimental subjects. In order to guard against that, **double-blind experiments** are used. In a double-blind experiment, neither the investigator nor the subject knows until later what test condition the subject is in.

- In evaluating experimental evidence, you want to know the following: What was the control? If complex subjects are involved, how was **randomization** assured? What is the probability of getting the result by chance? For some questions, you may want to know whether the experiment was blinded or double-blinded.
Scientific Method and Scientific Discovery

• The principles of experimental design we have been discussing are all parts of the toolkit of empirical investigation. The scientific method is as follows: Define a question. Formulate an explicit hypothesis. Design a test with measurable outcome. Perform the test and analyze the results.

• Scientific research, however, also often requires a little bit of luck. It has been estimated that half of all scientific discoveries have been stumbled on. This was the case, for example, with the discovery of penicillin by Alexander Fleming in 1928.

• Scientific discovery can be chancy. Philosophers of science have long distinguished between a context of discovery and a context of justification. There’s no recipe, no algorithm, no scientific method of discovery. Once a discovery is made, however, the process of demonstrating the fact to others can follow a very specific method—the method of repeatable experiment.

Limitations to the Scientific Method

• Thus far, we’ve seen aspects of experimental design that are important both in presenting and in evaluating experimental evidence, but there are limitations in some aspects of experimental design.

• For example, experimental results can be faked and sometimes have been, both in the fringes of pseudoscience and in core areas of testing.

• Experimental results can be selectively chosen to support the favored case. This is called cherry-picking. There are perennial worries that companies choose to report those experiments that make their products look good and choose not to report those that don’t. What we really need to know are the results of all the relevant experiments.

• Not every experimental result is submitted for publication, and not every result submitted for publication is accepted. This is called the file-drawer problem. When you survey the literature to find out
whether most experiments support or fail to support a particular effect, you may be surveying a misrepresentative sampling of all the tests that have actually been run. The tests that show a less dramatic or less popular result may be buried, unseen, in someone’s file drawer.

Consider the Context

- Experimentation is never pure; it always takes place in a context. All effective experimentation tests one belief against a background of other beliefs. For the moment, at least, those other beliefs are simply assumed.

- One of the earliest cases cited in the history of experimentation is Eratosthenes’s calculation of the circumference of the earth.
  - The basic observation was this: On June 21, the summer solstice, in the Egyptian town of Syene, the sun appeared directly overhead. Looking down a well, a man’s head would block the sun’s reflection.
  - Back in Alexandria, Eratosthenes knew that on that day, the elevation of the sun was 1/50 of a circle off from being directly overhead—a little over 7 degrees, and he knew that Syene was directly south of Alexandria. Thanks to the royal surveyors, he also knew the distance between the two towns.
  - Assuming the earth to be a sphere and the sun to be so far away that its rays could be thought of as essentially parallel, Eratosthenes was able to use those two facts and a little mathematics to conclude that the circumference of the earth was a little more than 25,000 miles—very close to our contemporary estimate.
  - But note the assumptions here. If the earth were not a sphere, the calculation would not hold, and the experiment would not show what it was taken to show. If the sun were not so far away that its rays could be treated as parallel, the calculation would not hold.
○ Because no experimental result is established independently of a set of background assumptions, what every experiment really establishes is an alternative: either the claimed result holds or one or more of our background assumptions is wrong.

**Terms to Know**

**context of discovery**: Used by philosophers of science to label the process by which a hypothesis first occurs to scientists, as opposed to the context of justification, the process by which a hypothesis is tested. The consensus is that context of justification follows clear, rational rules, but context of discovery need not do so.

**context of justification**: Used by philosophers of science to label the process by which a hypothesis is tested or established, as opposed to how it is first invented or imagined. The consensus is that context of justification follows clear, rational rules, but context of discovery need not do so.

**controlled experiment**: A test that demonstrates a causal connection between a variable \(X\) and an outcome \(Y\) by establishing both “if \(X\), then \(Y\)” and “if not \(X\), then not \(Y\).” In essence, a controlled experiment is a double test of two cases side by side, one with and one without \(X\).

**double-blind experiment**: In experimentation, a way of controlling the possibility that unconscious bias of subjects and investigators may influence the results by assigning subjects so that the experimental condition is hidden from them. In a double-blind experiment, neither the subjects nor the investigators who interact with them know which experimental condition is being tested until after the experiment is complete.

**Mill’s joint method of agreement and difference**: In John Stuart Mill’s *System of Logic*, evidence for causal linkage between \(X\) and \(Y\) is stronger when conditions for both Mill’s method of agreement and Mill’s method of difference are present; that is, \(Y\) appears whenever \(X\) is present and disappears when it isn’t.
Mill’s method of agreement: In John Stuart Mill’s *System of Logic*, if two cases are alike only in $X$, and $Y$ occurs in both, we have evidence that $X$ and $Y$ are causally linked.

Mill’s method of concomitant variation: In John Stuart Mill’s *System of Logic*, applies to cases in which there is a range of values for $X$ and $Y$. If the amount or strength of $Y$ varies as the amount or strength of $X$ varies, there is evidence for a causal link between $X$ and $Y$.

Mill’s method of difference: In John Stuart Mill’s *System of Logic*, if the presence or absence of $X$ is the only difference between two cases, with $Y$ a result in the first case and not in the second, we have an indication that $X$ causes $Y$.

randomization: In experimentation, a procedure for assigning groups so that the differences between individuals within groups distribute with equiprobability across experimental conditions.

### Suggested Reading

Crease, *The Prism and the Pendulum*.

Peirce, “The Fixation of Belief.”

Salsburg, *The Lady Drinking Tea*.

### Questions to Consider

1. Do you think that dogs and cats make experiments? Draw on your own experience to explain why or why not.

2. This lecture outlines several ways that an experiment can fail to be well designed. Here are examples of two ways an experiment can fail:
   (a) For a taste test, people are asked to choose whether Post Toasties or Kellogg’s Corn Flakes taste the best, with a full view of the product boxes.
   
   The experiment is not blinded.
(b) You put a deck of cards between you and a friend. You perform an experiment in clairvoyance by picking cards one by one and looking at them with their backs to your friend. Your friend’s job is to use extrasensory perception to say whether the card is red or black.

The experiment is not double-blinded.

Now supply your own examples for an experiment that would fail in each of these ways:
(a) The experiment shows that when factor $X$ is present, factor $Y$ is present, too. But it doesn’t show that when factor $X$ is not present, factor $Y$ is not. It is not a controlled experiment.

(b) The experiment exposes one group of people to factor $X$. They develop factor $Y$. The experiment includes another group of people who are not exposed to factor $X$. They do not develop $Y$. But people were not assigned randomly to the two groups. It is not a randomized controlled experiment.

3. The Greeks, for all their brilliance, never developed a full science involving experimentation. Why do you think that was? How might history have been different if they had?

**Exercises**

Can your friends tell expensive bottled water from inexpensive bottled water and each of those from tap water? Expand R. A. Fisher’s Lady Drinking Tea experiment by preparing four samples of each and giving them to a friend “blind,” in random order.

How well did your friend do in categorizing each sample, compared with what one might expect by chance alone?
In a previous lecture, we talked about decision theory. In a nutshell, we said, when choosing between two options, calculate the probability and value of possible outcomes. Then pick the option that has the highest probability multiplied by value. Maximize expected utility. Decision theory, however, is made for contexts in which there is just one decision maker: you. Where decision theory leaves off, game theory begins. Game theory is designed to address, in an abstract way, rational decision making in the social context—the context in which multiple players are making interlocking decisions. An understanding of game theory and an appreciation of its limitations should both be part of the philosopher’s toolkit.

Game Theory

- **Game theory** originated with John von Neumann, who was also instrumental in the development of computers and the atomic and hydrogen bombs.

- What does von Neumann’s involvement with nuclear weapons have to do with game theory? Not long after World War II, the Soviet Union acquired nuclear capability. The United States could not navigate in that world as if it was the only decision maker that mattered. It was necessary to think through chains of action and response, and multiple decisions on both sides mattered significantly. With the possibility of mutual annihilation looming, the stakes could hardly be higher.

- It was in anticipating a world with two nuclear superpowers that von Neumann developed game theory. In 1944, von Neumann and Oskar Morgenstern published *Theory of Games and Economic Behavior*. The subject sounds innocent enough: games. But the real target was rational strategies for conducting a cold war, rather than a hot one.
Game theory is designed to address, in an abstract way, rational decision in the social context—the context in which multiple players are making interlocking decisions.

Classic Game Theory: Prisoner’s Dilemma

- Games in von Neumann’s sense aren’t just checkers and chess. They include every situation involving cooperation and competition in which the outcome depends on the multiple actions of multiple players. The game that has become the central exhibit of game theory is called the Prisoner’s Dilemma. That particular game—that particular tradeoff between potential competitors or cooperators—has been the most studied game in game theory.

- The story behind the game is one of two bank robbers, Al and Bruiser, caught and held in separate cells. The district attorney admits that the evidence in the case is pretty thin. If both bank robbers refuse to cooperate, they will both get off with light sentences. But if Bruiser confesses and Al doesn’t, Bruiser’s evidence will be used against Al; Bruiser will go free; and Al will be put away for years. The same is true if Al confesses and Bruiser doesn’t. If both confess, they will both go to jail, but they may receive some time off because they pled guilty.
The question is: Should one accomplice give evidence against the other? The important point is that the outcome doesn’t all depend on one prisoner or the other. It depends on how the decisions by both prisoners play out. That’s game theory.

As in previous lectures, we can use a matrix to represent possible outcomes, and we see that there are four ways this game could play out, reflected in the four quadrants of the matrix: Both players cooperate with each other, both defect against each other, A cooperates and B defects, or A defects and B cooperates. (Note that “cooperate” here means to cooperate with the other prisoner, not the police.) In the matrix, payoffs for A are listed in bold and italics in each box, and those for B are listed in standard type.

\[
\begin{array}{c|cc}
 & \text{Cooperate} & \text{Defect} \\
\hline
\text{Cooperate} & 3, 3 & 0, 5 \\
\text{Defect} & 5, 0 & 1, 1 \\
\end{array}
\]

Game theory is about how decisions interact. The matrix shows potential gains and losses for each side given different combined decisions. If one prisoner cooperates with the other, while the second player defects against the first, the defector walks away with five points—a full five years of freedom. The cooperator gets no years of freedom. If both players choose to cooperate with each other, both end up with three points—three years of freedom. If both defect, they both end up with only one year of freedom.

The rational thing to do is to defect. This is a first major result in game theory. In a one-shot Prisoner’s Dilemma, defection is strongly dominant.

**Tit for Tat**

In real life, however, interactions are almost never a one-shot deal. Social life involves repeated interactions with the same people and the same organizations. When we step up to this level, we’re dealing
with what is called an iterated game. In an iterated Prisoner’s Dilemma, the payoffs on each round stay the same, but there is more than one round and the players remember what happened in previous rounds.

- In the iterated Prisoner’s Dilemma, there’s one strategy that’s of particular interest: **tit for tat**. With this strategy, one player mirrors the play made by the opponent in the previous round; the player cooperates if the opponent cooperated in the previous round but defects if the opponent defected.

- Robert Axelrod, a political scientist at the University of Michigan, tried to understand the iterated game by running a computer-based Prisoner’s Dilemma tournament. He set it up so that every submitted strategy played 200 rounds against every other strategy. He also made sure that every strategy played against itself and against an opponent that cooperated and defected at random. Axelrod added up the total points that each strategy gained in all those competitions. The winner was the strategy that got the most points over all.

- The strategy that came out with the most points—the most rational strategy to play against that whole field of competitors—was the simple strategy tit for tat.

- In the one-shot Prisoner’s Dilemma, defection is the rational choice. But in repeated games played with many different players—as in Axelrod’s tournament and more like real life—the strategy that does best is one that starts by cooperating and returns like for like.

- There’s an important lesson to be learned from the tit for tat result: Cooperation may triumph even in situations where we might think it impossible.

**The Limitations of Game Theory**

- The limitations of game theory have lessons to teach us about social rationality.
• The Prisoner’s Dilemma is clearly an artificial setup—artificial enough to be captured in a matrix. That in itself isn’t bad—
  **simplification** is one of our most important conceptual tools—but there are several artificialities in that setup of which we should be particularly wary.
  ○ The first is that everybody knows the rules, everybody knows the payoffs, and it’s always clear to both sides whether an action was a cooperation or a defection. Such a situation is called a game of perfect information. But real life is often not like that.
  ○ Second, note that all moves are either full cooperation or full defection. There’s no halfway here.
  ○ Third, also note that players have to play on each round. They can’t just opt out.

• For real social rationality, then, it’s not enough to act as one would in game theory. For real social rationality, one also has to keep in mind that actions of cooperation and competition may not always be clear and don’t have to be all or nothing; further, opting out may be an important possibility.

**Behavioral Economics**

• Decision theory and game theory have one core characteristic in common. Both gauge rationality in terms of self-interest. In game theory, each player is calculating how to maximize his or her own gain on the assumption that other players are doing the same. But do people actually make decisions in the way that game theorists—and economists in general—regard as rational?

• The answer is that they don’t. That’s a major finding of the developing field of behavioral economics, which puts the standard assumptions to the test. The truth is that people systematically violate the economist’s canonical assumptions of rationality.

• Consider the **Ultimatum Game**. Two people, the proposer and the responder, meeting for the first time and have to split an amount
of money. The proposer gets to propose how the money is to be split—50/50, perhaps, or 90 percent to the proposer and 10 percent to the responder. The responder then gets to choose whether to accept that division or not. If the responder accepts, the money gets divided as proposed. But if the responder declines the division, neither side gets any money at all.

- Here is what economic rationality says that you should do: If you’re the responder, you should take any division that gives you any money at all. Why? Because something is better than nothing.

- What is the rational thing to do as a proposer? If the responder is rational in the way just outlined and you’re just out for the money, you should propose as much as possible for you and as little as possible for the responder. Because the responder is rational, he or she will take the money offered. Because you’re rational, you’ll walk away with a bundle.

- But in the real world (at least in industrialized societies), proposers tend to offer about 44 percent to the other side. In sampling across a range of societies, across all continents, virtually no proposer offers less than 25 percent to the other side. Interestingly, the more that a culture incorporates a market economy, the higher the offer is to the other side.

Social Rationality

- As of now, the best formal tool we have for studying social rationality is game theory. It at least recognizes the issue that decisions made jointly may have a different rationality than those made individually. Game theory supplies some simple models. It even offers some hope for emergence of cooperation in egoistic situations, such as the Prisoner’s Dilemma.

- Its limitations when applied to real life carry lessons as well. It may not always be clear whether our real-life “moves” are cooperations, defections, or something in between. The “game” we’re playing in real life—the payoffs to each side—may be far
from obvious. Indeed, the game itself may be open to negotiation. The motivations of the multiple players won’t be as predictable as purely rational models might lead us to believe and might well depend on differences in cultural background.

Terms to Know

**game theory**: The mathematically idealized study of rational interaction in competitive and cooperative situations.

**Prisoner’s Dilemma**: In game theory, a two-person game in which the value of mutual cooperation is greater than the value of joint defection, the value of defecting against a cooperator is greater than the value of mutual cooperation, but the value of cooperating against a defector is lower than the value of mutual defection, as illustrated in the following matrix:

<table>
<thead>
<tr>
<th></th>
<th>Cooperate</th>
<th>Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperate</td>
<td>3, 3</td>
<td>0, 5</td>
</tr>
<tr>
<td>Defect</td>
<td>5, 0</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

**Prisoner’s Dilemma Matrix**

In the Prisoner’s Dilemma, the real loser is cooperation in the face of defection.

**simplification**: Conceptual replacement of a complex concept, event, or phenomenon with a stripped-down version that is easier to understand or to manipulate mentally.

**tit for tat**: In game theory, a strategy for playing an iterated game in which a player mirrors the play made by the opponent in the previous round; the player cooperates if the opponent cooperated in the previous round but defects if the opponent defected in the previous round.
**Ultimatum Game:** In game theory, a game in which two participants are given the task of splitting a payoff, for example, a certain amount of money. One participant is randomly selected to be the Proposer of how to make the split, and the other, the Responder, must decide whether to accept the offer. If the Responder accepts, the payoff is split between them as proposed; if the Responder rejects the offer, no one gets a payoff.

**Suggested Reading**


Poundstone, *Prisoner’s Dilemma.*

**Questions to Consider**

1. From your own experience, can you think of a situation in which you tried to produce one result but something very different happened, not because of what you did but because of what someone else did independently?

2. Suppose you really were involved in the Prisoner’s Dilemma. You and your partner in crime are in separate cells and cannot communicate. If you both stonewall, you’ll both get off with just one year in prison. If you both confess, you’ll both get two years in prison. If you confess and your partner doesn’t, you’ll go free and he’ll do five years in prison. If he confesses and you don’t, he’ll go free and you’ll do the five years. If this were a real case, which option would you take? Would you confess or stonewall?

3. The empirical evidence shows that people from different cultures play even artificial games, such as the Prisoner’s Dilemma, differently. Does that show that cultural influence can override rationality or that rationality is somehow culture-relative?
Spend a little Monopoly money to find out what people actually do in the Prisoner’s Dilemma.

(a) For the one-shot game, explain the payoffs to two friends, emphasizing that you want them to think rationally and that this game will be a one-shot deal: Each is to choose independently whether to cooperate or defect. If they both cooperate, they get $3 each. If they both defect, they get $1 each. If one defects and the other cooperates, the defector gets $5 while the cooperator gets nothing.

(b) For the iterated game, play 10 rounds with the same payoffs. Who got the most money over all? A question for further thought: Was the play different on the last round?

Try the same experiment with payoffs from one of the other games, Chicken or Stag Hunt. Do you get different results in those cases?

<table>
<thead>
<tr>
<th></th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Cooperate</td>
<td>Defect</td>
</tr>
<tr>
<td>Cooperate</td>
<td>3, 3</td>
</tr>
<tr>
<td>Defect</td>
<td>5, 1</td>
</tr>
</tbody>
</table>

Chicken Matrix

<table>
<thead>
<tr>
<th></th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Cooperate</td>
<td>Defect</td>
</tr>
<tr>
<td>Cooperate</td>
<td>5, 5</td>
</tr>
<tr>
<td>Defect</td>
<td>1, 0</td>
</tr>
</tbody>
</table>

Stag Hunt Matrix
This lecture covers an important conceptual tool—thinking with models—that incorporates the problem-solving strategies we have been discussing, such as visualization, simplification, and thought experiments. It has a long and distinguished history, but with the advent of computers, it will have an even more significant future.

Schelling’s Model: Visualize, Simplify

- Let’s start with an example of thinking with models, drawn from the work of the economist and political scientist Thomas C. Schelling.
  - To examine residential segregation, Schelling constructed a model with a checkerboard grid and used dimes and pennies to represent two ethnic groups. He distributed these randomly across the checkerboard, making sure to leave some empty spaces.
  - What Schelling thought of as the “neighbors” of a penny or dime were those in the immediate squares around it. The neighborhood for any square included those eight squares touching it on each side and on the diagonals.
  - Schelling then gave his pennies and dimes some low-level preferences. The dimes wanted to live in a neighborhood where one-third of the people around them were of the same ethnicity, but they were perfectly happy to have two-thirds of their neighbors be different. The same was true of the pennies.
  - What happens as we play those preferences out? Even with those low preference levels, distributions on the board start to form neighborhoods. There are small patches that are almost all pennies and small patches that are almost all dimes.
• One important point to draw from Schelling’s model is that thinking is something you can do with your eyes. That is the strategy of visualization.

• A second and obvious characteristic of Schelling’s model is that it uses the strategy of simplification.
  ○ Schelling was out to understand something very complicated—social patterns of housing by ethnicity. Those patterns are the result of innumerable complex decisions by thousands of interacting individuals, influenced by multiple factors.
  ○ But note that Schelling’s model includes none of that detail. It doesn’t even really have individuals who belong to ethnic groups—just pennies and dimes. His strategy here is obvious: to try to understand something complicated by understanding something related but much simpler.

Three-Part Model Structure: Input, Mechanism, Output

• In examining the structure of Schelling’s thinking, we see that his model works with an input and generates an output by way of some mechanism. The input is the original random distribution of pennies and dimes. The mechanism here is Schelling’s rule: Pick a penny or dime at random. If at least one-third of its neighbors are like it, it stays. If not, it moves. The output of the model is the final arrangement of pennies and dimes.

• Although we can build a three-stage model, we are really only confident of two of the three stages. On the basis of the two pieces of information we have, we use the three-stage model to give us the third piece of information we don’t have—the missing piece of the puzzle.

• Prediction uses information about input and mechanism to forecast new information about the future. Retrodiction uses information about mechanism and output to generate information about the past. Explanation uses information about input and output to generate
possible mechanisms for linking the two. In each case, we build in what we know at two stages of the three-stage model and read off the information we need at the third.

**Using Prediction**
- As we said, prediction uses what we know at two stages to get new information about the future. Weather prediction uses models in precisely this way. We know what the weather readings are across the United States today. We know what they’ve been for the past week. Those are our input conditions.

- Over decades, we have developed certain principles that we think apply to the weather; for example, we know what happens when a dry cold front meets a wet warm front; we know something about wind and the dynamics of the atmosphere. All of these are part of the mechanism of our models. In prediction, we know the input conditions and are fairly confident of the mechanism. We read the new information we’re after at the output stage: an estimate of what the weather is going to be like tomorrow.

- If we use an X for the information we have and an O for the information we’re trying to get, the case of prediction looks like this:

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Mechanism</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
</tbody>
</table>

**Using Retrodiction**
- Retrodiction, on the other hand, uses two different stages to give us information about the past. For example, there is still a great deal of debate about how the moon was formed. One theory is that the centrifugal force of the spinning earth ejected it as a chunk early in the history of the solar system. Another theory holds that the moon was a separate body captured in the earth’s gravitational field. According to a third theory, the moon was formed when some other large body hit the early earth, blasting out material that became the moon.
We can build a model to decide among those theories. We build into our mechanism a representation of physical laws: centrifugal force relative to rotation, the physics of velocity and impact, and gravity. The output is a given: a moon rotating around the earth in the way we know it does, with the composition we know it has.

If any of those inputs produces our given output with the mechanism of natural law, we have a point in favor of that particular theory. If we can’t get the model to work with a particular input, we have a point against it. In prediction, we enter input and mechanism and read off the information we’re after from the output. In retrodiction, we assume mechanism and output and read off the information we’re after from what input produces that output with that mechanism.

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Mechanism</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>Retrodiction</td>
<td>O</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Using Explanation

Explanation uses the three stages in yet a different way. What mechanism might take us from just about any input—a fairly random one, for example—to the patterns of residential segregation we see around us?

In Schelling’s case we assume an input, observe an output, and ask what mechanism would take us from one to another. Put beside the others, the structure of explanation looks like this:

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Mechanism</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>Retrodiction</td>
<td>O</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Explanation</td>
<td>X</td>
<td>O</td>
<td>X</td>
</tr>
</tbody>
</table>
Limitations to Thinking with Models

- Thinking with models is a way of thinking through “if...then” connections in the real world. In order to be useful, models must be simpler than the reality that we’re trying to understand. But they also have to match reality in relevant ways.

- If we want a reliable prediction of tomorrow’s weather, our input must be correct, and our mechanisms must work the way the weather really does. That’s a challenge that meteorologists have been dealing with in computer models for years.

- The case of retrodiction is perhaps even more challenging. We can probably rely on our observations of the moon, but here, too, we must be confident that our mechanism captures the relevant physical laws. Another problem with retrodiction is that there might be more than one set of inputs that are sufficient to produce the outcome.

- The same issue shows up in explanation. The fact that a particular mechanism leads from a particular input to an observed output doesn’t tell us that it’s the only mechanism that could produce that outcome. It also doesn’t tell us that it’s the mechanism that actually produced that result in reality.

Models and Scientific Experimentation

- Scientific experiments are also about constructing “if...then” scenarios. There is, in fact, a close connection between model building and experimentation. The models we’ve talked about so far are conceptual models or conceptual models instantiated in computer programs. But the history of physical models shows the same three-stage structure. There, too, the three stages can also be used for prediction, retrodiction, or explanation.

- In 1900, Orville and Wilbur Wright built a small wind tunnel to test models of miniature wings in preparation for building a flying machine. Over the course of several months, they tested more than 200 wing designs to enable them to predict which ones might actually work.
But remember the limitations of models. At this point, the Wright brothers were working only with a model. They assumed that the forces in their miniature wind tunnel would scale up to the real world. But in physical models, as in conceptual models and computer models, that correspondence with reality is always a risk.

**Hobbes’s *Leviathan***

- Conceptual models have sometimes served merely as metaphors, as in Plato’s image of the cave in *The Republic*, for example. But conceptual models have often served as something more in the history of philosophy.

- The philosopher Thomas Hobbes witnessed firsthand the ravages of the English civil war that raged from 1642 to 1651. What he wanted to understand were prospects for stability and peace. We could phrase Hobbes’s question this way: Given what we know of human nature as input, what mechanism would produce stable peace as an outcome?

- Hobbes’s view of human nature was that individuals were out for their own interests alone. That was his input, what he called a “state of nature.” In a state of nature, life was “solitary, poor, nasty, brutish and short.” What mechanism could lead us from that to a stable peace?

- The name of Hobbes’s book describing his proposal for the necessary mechanism was *Leviathan*. In order to get from the state of nature to a state of stable peace, people would have to contract together. They’d have to give up their individual liberties to a central state—a leviathan (more specifically, to a monarchy).

- Hobbes’s book exhibits thinking with a three-stage model, but like all models, it has risks. We need to know how real all three stages of the model are. Are people as egoistic as Hobbes treats them in his input condition? Would a social contract really give the desired output, or would it result in a centralized tyranny? Even if Hobbes’s
social contract mechanism would produce the outcome, is it the only way to do so? Is it the most desirable way to do so?

- Three-stage models are a way to put strategies of visualization and simplification to work in figuring out “if...then” scenarios. It often helps to think of things in terms of simplified input, mechanism, and output, whether that structure is instantiated physically, computationally, or conceptually.

- Thinking in models can be used for many purposes: to guide data collection, to explore the core dynamics of a process, to suggest analogies, to illuminate core uncertainties, to point up overlooked details, to offer options for change or intervention, to show the simple core of complex processes and the complexities behind simple ones, and perhaps most importantly, to suggest new questions.

### Terms to Know

**explanation**: In a three-stage model, information about input(s) and output(s) is used to generate possible mechanisms for linking the two.

**input**: The first structural stage of a three-stage input-mechanism-output model: the initial setup, conditions, or changes on which the model operates.

**mechanism**: The second structural stage of a three-stage input-mechanism-output model: the operations that run on input in order to produce the output.

**output**: The third structural stage of a three-stage input-mechanism-output model: the result of a model run given a particular input.

**prediction**: In a three-stage model, information about input(s) and mechanism is used to forecast the future, that is, to generate information about probable output(s).

**retrodiction**: In a three-stage model, information about the mechanism and output(s) is used to generate information about the past (inputs).
Suggested Reading


Cummings, Chakravarty, Singha, et al., Toward a Containment Strategy for Smallpox Bioterror.

Epstein, “Why Model?”

Questions to Consider

1. Models are now commonly used in weather forecasting (what the weather will be like tomorrow and next week), political forecasting (who is likely to win an election), and financial forecasting of economic and stock market trends. These models are often instantiated as computer programs. From your own experience, how well do you think those computer models currently do in making predictions in those three fields?

2. Schelling arranges dimes, pennies, and empty spaces randomly on a checkerboard grid. He then picks a coin at random. If the occupied spaces around it contain 30 percent or more of its own kind, it is “happy” and stays. If the occupied spaces contain less than 30 percent of its own kind, it is “unhappy” and moves to an empty spot close by. As the rules are progressively applied, segregated suburbs of pennies and dimes develop. What can legitimately be concluded from that model about de facto racial segregation? Just as important: What cannot be concluded from that model?

3. In response to people who ask why they should work with models, the computational modeler Josh Epstein likes to say: “You already model. You’re working with a model in your head. The real question is just what assumptions you are making and how good that model is.” Do you think Epstein is right about that, or not?
NetLogo is a simple programming environment designed and authored by Uri Wilensky, director of Northwestern University’s Center for Connected Learning and Computer-Based Modeling. It comes with a wonderful Models Library of examples ready to run and easy to tweak. You can download it for free at http://ccl.northwestern.edu/netlogo/download.shtml and will find Schelling in the Models Library under “Social Science” and “Segregation.”

Start by setting %similar-wanted at different levels. What happens with segregation if neighbors are “happy” with just 25 percent of their own kind? What happens if they are “happy” only with 50 percent of their own kind?
Exmaining the philosopher’s toolkit has been the aim of these lectures, that is, learning a set of conceptual techniques and strategies, tips, methodologies, and rules of thumb for thinking more clearly, with greater focus, more effectively, and more accurately and for building stronger and more rational patterns of inference and argument. In this lecture, we’ll talk about the philosopher’s toolkit in terms of some of the great thinkers who have used it and added to it. What was it that made these thinkers great?

Plato

- The influence of Plato has been enormous. Alfred North Whitehead said that all of Western philosophy was “a series of footnotes to Plato.” But what made Plato so great?

- First and foremost, Plato saw the power of abstraction. Abstraction gives us the power of generality. In abstraction, we simplify away the differences of different cases—the sand of triangles drawn in sand, the chalk of triangles in chalk—and think simply in terms of the shape things have in common. In order to think abstractly, you don’t add on; you subtract—you simplify.

- But Plato didn’t talk about abstractions as removed from reality. He thought that the realm of abstraction was the true reality. For Plato, the world we experience around us is just an illusory, imperfect, and distant echo of the true reality of abstraction.

- Another element that made Plato great is that his works are dialogues, not monologues. Crucial to the dialogues is the interaction between multiple points of view, multiple opinions, and multiple people. The emphasis is on social rationality—ideas as something we work through together.
Aristotle

- Aristotle founded logic by seeing things with a very particular vision. Deep in the complicated and messy process that is thinking, deliberation, discussion, and argument, Aristotle saw something amazingly simple: structure. The ability to see that structure gave him the idea of systematizing thought.

- If we could systematize thought, we might be able to avoid some of our biases. We could make our thinking faster, more accurate, and more effective. That vision we owe to Aristotle, and it’s a vision that extends even into contemporary computers.

- Aristotle is also a great thinker because, in contrast to Plato, he insisted on the lessons of experience. He was a real-world observer. Although there aren’t any true experiments in Aristotle—that came much later—he believed that real knowledge of the world demands looking at what the real world is really like.

- This contrast between Plato and Aristotle is captured in Raphael’s painting The School of Athens. At the center, Raphael shows Plato and Aristotle in deep discussion. Plato, the great teacher, is pointing upward—toward the abstract. Aristotle, his rebellious student, is gesturing with a level hand toward the concrete reality before him.

Galileo

- Early experimentation was often used to question the authority of Aristotle. That was certainly true of Galileo, who has to rank as one of the great thinkers on anyone’s list. The major lesson from the great experimenters is: Don’t trust authority—not even Aristotle’s. Don’t rely on what most people think. Design an experiment and put it to the test.

- Galileo was an aggressive and indefatigable observer. We’ve mentioned Galileo’s legendary experiment, dropping a musket ball and a cannonball from the top of the Leaning Tower of Pisa. Whatever Aristotle thought, whatever anybody thought, Galileo’s experiment showed that things fall at the same rate regardless of weight.
But precisely how do things fall? In order to investigate, Galileo built a model. What he really wanted to know was how things accelerate in free fall. But his tools weren’t precise enough to measure that. Instead, he rolled balls down an inclined plane and measured the speed with which objects passed successive marks on the plane. He found principles that applied whatever the angle of the ramp and generalized from those to answer the question about acceleration in free fall.

Descartes
- If Galileo is the founding figure in modern science, René Descartes may be the founding figure in both modern philosophy and modern psychology.
- One of his major works is the *Discourse on the Method of Rightly Conducting One’s Reason and of Seeking Truth in the Sciences*. His messages:
  - “Divide each of the difficulties under examination into smaller parts.”
  - “Starting with what is simplest and easiest to know, ascend little by little to the more complex.”
  - “Never… accept anything as true which [you do] not clearly know to be such… avoid precipitancy and prejudice.”
  - “To be possessed of a vigorous mind is not enough; the prime requisite is rightly to apply it.”

Newton
- Sir Isaac Newton is known as a great formalizer. But he was a great experimentalist, as well. When the plague hit England in 1664, Newton went home to his mother’s house in Lincolnshire and began experimenting with prisms.
- The prevailing view was that a prism stained light into different colors. But Newton was able to show that what the prism...
did was bend different colors of light differently—what we call different wavelengths.

- We know from his notebooks that Newton didn’t follow any rigid scientific method. He simply tried many things, inventively and on the fly, to see what happened. His example carries an important lesson about the two things effective thinking requires.
  - One requirement is creative exploration. Good thinking—whether experimental or not—requires that we take time to play with ideas. Genuine creative thinking cannot be regimented. Children know this intuitively.
  - But creative conceptual play isn’t enough. Good thinking also requires a second phase in which one squints critically at the creative results. That is the stage at which Newton asks what precisely has been shown. Can we formulate those results more clearly? How exactly does the experiment refute established doctrine?

Thinking Better: Lessons from the Great Thinkers

We can sum up the lessons from the great thinkers in seven maxims.

- First, simplify.
  - We’ve seen this lesson in Newton and Descartes: Break a complex problem into simple parts; start with the simple, build up to the complex. We emphasized that strategy early on in the Tower of Hanoi problem, in going to extremes, and in the strategy of thought experiments.
  - If you can’t solve a hard problem, use a simpler model. That is what Galileo did. Simplification is what model building is all about, from the Wright brothers’ wind tunnel to the Mississippi River basin model to Schelling’s segregation model.
  - Plato’s abstraction represents the power of generality you can get not by adding detail but by subtracting it.
New thinking, in science as well as anywhere else, must be based in creative play—we need creative thinking to give us material worth working with.

- Second, look for patterns, but watch out for them, too.
  - Visualization allows us to harness visual pattern recognition. It lets us think faster, more intuitively, and more powerfully. Aristotle knew the power of visualization. He used it in his square of opposition. Einstein was one of the great visualizers. The heuristics that make us smart are often forms of pattern recognition.

  - But the biases that lead us astray are often patterns, as well. Our vivid images can mislead us as to true probabilities. We miss the gorilla in the room because we’re paying attention to some other pattern. Patterns are extremely powerful. That’s why it’s important to put them to work and equally important to handle them with care.

- Third, think systematically.
○ To systematize thought is the Aristotelian vision. Thinking systematically is what gave Newton his edge. We have developed a number of tools for systematic thought: tools in statistics, probability, and decision theory.

○ One simple lesson in thinking systematically goes all the way back to the bat and ball example in the first lecture. Want to think better? Check your work.

• Fourth, put it to the test.
  ○ Respect for reality runs through the whole history of experiment. Charles Sanders Peirce thought of science as the one truly self-correcting technique for fixation of belief. Karl Popper emphasized that genuine science demands risk—the risk of falsifiability. Galileo is the patron saint of experimentation. Simple and beautiful experiments go all the way from Newton’s prism to Richard Feynman’s O-rings.

  ○ Do you really want to know what reality entails? Ask. Put your thinking about reality to the test.

• Fifth, think socially.
  ○ Many of the complexities we face are social complexities. They require thinking in terms of multiple perspectives and multiple players. Thomas Hobbes knew the importance of thinking socially, as did Jean Jacques Rousseau. That kind of thinking is the whole goal of contemporary game theory.

  ○ Inevitably, there are social threats to good thinking, as well. We have to watch out for mere rhetoric, standard fallacies, and the tricks of advertisers. Plato’s dialogues are the first great works in philosophy: They emphasize this social side to thinking.

• Sixth, think with both sides of your brain.
  ○ Newton combined the power of creative play with systematic critique. Both parts need to be present: the discipline and the imagination.
We emphasized early on that emotion need not be the enemy of rationality if used in the right domain. We couldn’t do without fast-and-frugal heuristics, but for long-range decisions, we need careful calculation, as well. Remember Daniel Kahneman’s system 1 and system 2 thinking.

- Finally, give yourself time to think.
  - Descartes must have been a phenomenally quick thinker, yet he said, “Those who travel slowly may yet make far greater progress than those who run.”
  - Google’s engineers are given one day a week to work on something entirely of their own. It’s unfortunate, but these days, our last bastions for free-range thinking seem to be the shower and the commute to work. Real thinking demands conceptual exploration, and that just takes time.

**More Great Thinkers to Come**

- Each of these great thinkers worked with a conceptual toolkit, but none of them built it alone. All inherited a set of conceptual tools from their predecessors. Newton said, “If I have seen farther, it is by standing on the shoulders of giants.” All added to the toolkit and then passed it on.

- That cultural heritage has now been passed to you: the philosopher’s toolkit—to use, to apply, to add to, to improve, to pass on. It’s wonderful to peer into the future and realize that there are more great thinkers to come.

**Suggested Reading**

Baggini and Stangroom, *Great Thinkers A–Z.*

Boorstein, *The Discoverers.*

Bronowski, *The Ascent of Man.*
Questions to Consider

1. One of the lessons drawn from the great thinkers in this lecture—with Newton as an example—was that good thinking requires both (a) creative conceptual play and (b) systematic and critical review. Which of those do you think you are better at? Do you have a friend who is better at the other?

2. In Raphael’s School of Athens (http://mv.vatican.va/4_ES/pages/z-Patrons/MV_Patrons_04_01.html), Plato points heavenward, emphasizing abstract conceptualization. Aristotle spreads his hand ahead, emphasizing attention to the real world. Which of those do you think is most important in good thinking? Would it be possible to have good thinking with just one and not the other?

3. What do you think we will learn from the next great thinker in our history? What do you think the next conceptual breakthrough will be?

Exercises

Sit down with a piece of paper and make your own list of great thinkers—the 10 greatest, perhaps. Who do you think have been the greatest in the history of Western thought?

The second question is harder, but more important: What exactly made those figures great?
Glossary

**abstraction**: The thought process that allows us to derive general concepts, qualities, or characteristics from specific instances or examples.

**ad hominen**: A fallacy that depends on an attack against the person making a claim instead of the claim that is being made.

**amygdala**: A small region deep within the brain that is associated with emotion.

**anchor-and-adjustment heuristic**: A common strategy used in calculating probabilities, but one that depends on how a question is phrased. Information given in the question is taken as a starting point, or anchor; individuals tend to adjust their responses upward if the anchor seems too low or downward if the anchor seems too high, arriving at an answer that is less extreme than the information given in the question but that may have little connection to the real answer.

**appeal-to-authority fallacy**: A fallacy in which the opinion of some prominent person is substituted for rational or evidential support; often used in advertising by linking a product to a celebrity “expert” rather than providing rational or evidential support for a claim about the product.

**appeal-to-emotion fallacy**: A fallacy in which positive or negative emotional tone is substituted for rational or evidential support; an argument strategy intended to cross-circuit the ability of the listener to assess whether a rational link exists between premise and conclusion by “pushing emotional buttons.”

**appeal-to-honesty advertising strategy**: Use of a spokesperson who appears credible and trustworthy.
appeal-to-ignorance fallacy: A fallacy in which absence of information supporting a conclusion is taken as evidence of an alternative conclusion. This fallacy acts as if ignorance alone represents some kind of positive evidence.

appeal-to-majority fallacy: An argument that treats majority opinion as if that alone constituted evidence supporting a conclusion or gave a reason for belief. This fallacy ignores the fact that people, even large numbers of people, are fallible.

appeal-to-prestige advertising strategy: Linking a product to status symbols in order to enhance the product’s desirability. Foreign branding is a specific form of the appeal-to-prestige strategy.

Aristotelian logic: Aristotle’s attempt to systematize thought by outlining a set of formal relations between concepts and propositions. These relations can be visualized by his square of opposition and his treatment of arguments as syllogisms. See square of opposition, proposition.

attention bias: Overlooking the unexpected because we are attending to the expected.

“attractiveness attracts” advertising strategy: Enhancing the desirability of a product by using spokespeople who are more attractive than average, by showing the product in beautiful settings, or by doctoring the product’s image or the product itself so that it is more photogenic. The flip side, “unattractiveness detracts,” is common in political campaign literature, which often seeks unflattering images of the opponent.

availability heuristic: The tendency for individuals to assume that things that are easier to bring to mind must be more common or occur more frequently; the tendency to generalize from simple and vivid images generated by single or infrequent cases and to act as if these are representative.
axiom: A claim accepted as a premise without proof and from which other claims are derived as theorems. Euclid’s geometry is a familiar example, in which theorems are derived from a small number of initial axioms.

axiomatic: Organized in the form of axioms and derivations from them. Euclidean geometry is an example of an axiomatic system. See axiom.

Brownian motion: The random motion of particles in a gas or liquid, such as motes in sunlight or grains of pollen in water.

Cartesian coordinates: The position of a point on a plane as indicated in units of distance from two fixed perpendicular lines, the x (horizontal) axis and the y (vertical) axis.

categorical proposition: In Aristotelian logic, a simple proposition that combines two categories using “all are,” “none are,” “some are,” or “some are not.” Categorical propositions can be visualized using two circles in a Venn diagram.

category: Any group of related things; the grouping is based on what are perceived as important similarities between those things.

causation: The act of producing an effect. In a cause-and-effect relationship, the presence of one variable, the effect, can be established as a direct result of the other variable, the cause. Opposed to correlation.

change blindness: The propensity for individuals not to perceive unexpected changes, particularly when attention is focused on something else.

Chicken: In game theory, a two-person matrix game in which the value of mutual cooperation is greater than the value of joint defection, the value of defecting against a cooperator is greater than the value of mutual cooperation, but the value of cooperating against a defector is higher than the value of mutual defection, as illustrated in the following matrix:
In Chicken, the real loser is mutual defection. For alternative payoff scenarios, see *Prisoner’s Dilemma* and *Stag Hunt*.

**combined probability**: The rules for calculating the odds that two or more events will happen. The probability of either one or another of two mutually exclusive events happening can be calculated by adding their individual probabilities. The probability of two independent events both happening can be calculated by multiplying their individual probabilities. Other rules apply for dependent events; see *independent events, mutually exclusive events*.

**complement**: In logic, given any category, the complement comprises all those things that do not fall in that category. For example, “senators” and “non-senators” are complements.

**complex-question fallacy**: A “trick question” presenting a false dilemma, or forced-choice alternative, presented in such a way that any answer is incriminating. For example: “Answer yes or no: Have you stopped beating your wife?” If you say yes, you have essentially admitted that at one time, you did beat your wife; if you say no, you have admitted that you are still beating her.

**concepts**: Ideas, the basic elements or “atoms” of thought, as distinct from the words that represent those ideas.

**concept tree**: The hierarchical structure that visualizes the relationships within a set of related concepts.
**conclusion**: The endpoint of an argument; in a logical argument, the claim to which the reasoning flows is the conclusion. See also **premise**.

**connotation**: The emotional tone or “flavor” associated with the ideas or things that words label.

**context of discovery**: Used by philosophers of science to label the process by which a hypothesis first occurs to scientists, as opposed to the context of justification, the process by which a hypothesis is tested. The consensus is that context of justification follows clear, rational rules, but context of discovery need not do so.

**context of justification**: Used by philosophers of science to label the process by which a hypothesis is tested or established, as opposed to how it is first invented or imagined. The consensus is that context of justification follows clear, rational rules, but context of discovery need not do so.

**contradiction**: A statement that both asserts and denies some proposition, P, often represented in the form “P and not-P.” If either part of a contradiction is true, the other cannot be true, and thus, a contradiction P and not-P is treated as universally false.

**contradictories**: The relationship between propositions on the diagonals of Aristotle’s square of opposition. It is a contradiction for both propositions on a diagonal to be true; if one proposition of the diagonal is true, the other must be false. See **square of opposition**.

**contrapositive**: A way of transforming categorical propositions by switching subject and predicate and replacing each with its complement. For some categorical propositions, the result is an immediate inference: the truth or falsity of the proposition is not altered. The contrapositive transformation preserves equivalence only for propositions in the upper left and lower right on the square of opposition: the universal positive (“All S are P”) and particular negative (“Some S are not P”). See **square of opposition**, **universal proposition**, **particular proposition**, **complement**.
contraries: The relationship between propositions at the top left (“All S are P”) and right (“No S are P”) of Aristotle’s square of opposition. If two propositions are contraries, it is not possible for both propositions to be true, but it is possible for both propositions to be false. See subcontraries, square of opposition.

controlled experiment: A test that demonstrates a causal connection between a variable (X) and an outcome (Y) by establishing both “if X, then Y” and “if not X, then not Y.” In essence, a controlled experiment is a double test of two cases side by side, one with and one without X.

converse: A way of transforming categorical propositions by switching subject and predicate. For some categorical propositions, the result is an immediate inference: the truth or falsity of the proposition is not altered. The converse preserves equivalence only for propositions in the upper right and lower left on the square of opposition: the universal negative (“No S are P”) and the particular positive (“Some S are P”). See square of opposition, universal proposition, particular proposition.

correlation: A direct or inverse relationship between two variables. In a direct correlation, the strength or frequency of both variables increases proportionately to each other; in an inverse correlation, the strength or frequency of one variable increases as the strength or frequency of the other decreases. However, correlation does not establish that one variable causes the other; one variable is not necessarily the cause and the other the effect. See also causation.

decision theory: A theory of how to make rational decisions by maximizing expected utility.

deductive validity: A deductively valid argument is one in which it is logically impossible for all premises to be true and the conclusion to be false.

deferred gratification: The ability to restrain oneself from taking an immediate payoff in order to obtain a larger payoff later.
demarcation: The philosophical problem of precisely defining the differentiation between two concepts. Specifically, the problem of precisely differentiating science from pseudoscience.

denotation: The things that a concept or word applies to.

dependent reasons: Premises that support the conclusion only when they are both present; propositions or claims that function together but are insufficient alone as support for the conclusion.

descriptive: Used to designate a claim that merely reports a factual state of affairs rather than evaluating or recommending a course of action. Opposed to normative.

diminishing marginal utility: The concept that units of value may not amass in equal increments; additions of one unit may not always be the same as additions of one unit of value. For example, one dollar may not be worth as much to a billionaire as to an individual who has no savings at all.

diversion fallacy: Also known as a “red herring”; the diversion fallacy arrives at a conclusion after diverting the listener’s attention from relevant considerations to the contrary.

double-blind experiment: In experimentation, a way of controlling the possibility that unconscious bias of subjects and investigators may influence the results by assigning subjects so that the experimental condition is hidden from them. In a double-blind experiment, neither the subjects nor the investigators who interact with them know which experimental condition is being tested until after the experiment is complete.

epidemiology: The study of the distribution, patterns, dynamics, and causes of health-related events in a population.

epithumia: Appetite or passion; according to Plato, the second element of the soul.
**equiprobable**: Having an equal mathematical or logical probability of occurrence.

**ethics**: The field of philosophy that focuses on moral issues: ethically good actions, ethically right actions, rights, and obligations.

**ethos**: Character of the speaker; according to Aristotle, the first quality of a persuasive presentation is that the speaker must appear knowledgeable and wise.

**eugenics**: A social movement that advocates practices and direct interventions aimed at changing, or ostensibly “improving,” the genetic characteristics of a population.

**expected utility**: In economics, calculated by multiplying the potential benefit of an outcome by its probability.

**explanation**: In a three-stage model, information about input(s) and output(s) is used to generate possible mechanisms for linking the two. See input, mechanism, output.

**extension**: In philosophy, the set or totality of things that a concept applies to.

**fallacy**: A form of argument in which the premises appear to support a conclusion but, in fact, do not; the term is often used to refer to familiar types of logical mistakes that may be used to trick or mislead.

**false alternative**: A fallacy in which a problem is presented as an either/or choice between two alternatives when, in fact, those are not the only options. Also called a “false dilemma.”

**falsifiability**: Karl Popper’s criterion for distinguishing real science from pseudoscience was that real scientific theories can be shown to be false. Real science takes risks in that theories are formulated so that they are open to disconfirmation by empirical testing or empirical data. See also verifiability.
**flow diagram**: A systematic sketch of a train of thought illustrating the lines of support between premises and conclusions in a rational argument; when one claim is intended as support for a second claim, an arrow is drawn from the first to the second. See *premise, conclusion*.

**foreign branding**: In advertising, associating a product with a foreign country in order to increase its desirability; a particular type of prestige advertising.

**formalization (systematization)**: The process of characterizing abstract relations, physical processes, or chains of thought in terms of explicit axioms, principles, or rules. Euclidean geometry systematizes spatial relations in a certain way; logic formalizes patterns of rational inference and valid argument.

**gambler’s fallacy**: Treating independent events as if they were dependent events. Someone who thinks black is “bound to come up” in roulette because of the appearance of a string of blacks (or reds) has committed the gambler’s fallacy.

**game theory**: The mathematically idealized study of rational interaction in competitive and cooperative situations.

**grammar**: The study of the proper structure of language in speech or writing; along with logic and rhetoric, an element of the classical medieval curriculum known as the *trivium*.

**hasty generalization fallacy**: Also known as jumping to conclusions. This fallacy occurs when one jumps to a conclusion about “all” things from what is known in a small number of individual cases. Racism and sexism often take the form of hasty generalizations.

**heuristics**: Simple guides to action or rules of thumb that allow us to act or make a decision without calculation or deliberation.

**Homo economicus**: The name given to a “species” of agents who would act rationally in the way economic game theorists say that individuals should act in order to maximize their own gain.
**implication**: In Aristotelian logic, the relationship moving from the top to the bottom left corners or the top to the bottom right corners of the square of opposition. If the proposition on the top left corner is true, then the proposition on the bottom left corner is also true; if the proposition on the top right corner is true, then the proposition on the bottom right corner is also true. Expressed as “if all S are P, then some S are P” for the left side of the square of opposition and “if no S are P, then some S are not P” for the right side of the square of opposition. See **square of opposition**.

**independent events**: Events that are isolated probabilistically; the fact that one event happens is not linked to and will not affect the fact that the other event happens.

**independent reasons**: A group of premises, or reasons, that are given as support for a conclusion, each of which could support the conclusion on its own.

**independent standard**: In negotiation or the attempt to reduce polarization, a deciding touchstone or court of appeal that is not open to manipulation and that can be agreed on by both parties in advance. In establishing a fair price for a house, for example, both parties might agree in advance to use the price that similar houses have recently sold for in the neighborhood.

**induction (inductive reasoning)**: Thinking that proceeds from individual experiences to theoretical generalizations.

**inference**: In logic, the derivation of a conclusion from information contained in the premises.

**input**: The first structural stage of a three-stage input-mechanism-output model: the initial setup, conditions, or changes on which the model operates.

**intuitive approach**: An approach to problem solving that focuses on the general outline of a problem, attending to a few critical issues and generating a quick decision after consideration of relatively few alternative solutions. Opposed to an analytic approach, which examines a wide range of specifics and generates an exhaustive list of alternative solutions.
Kripkean dogmatist: An individual who believes that his or her position is right and, on that ground alone, is prepared to reject any and all evidence to the contrary.

law of large numbers: In probability theory, the fact that as the number of trials increases, the outcome will approach the mathematically expected value. For example, in a situation where there are two equiprobable outcomes, such as heads or tails when flipping a coin, the longer the run, the more likely the outcome will be a 50/50 split.

logic: The study of patterns of rational inference and valid argument.

logical fallacy: See fallacy.

logos: Logic; according to Aristotle, the third quality of a persuasive presentation is that it will lay out the argument clearly and rationally, step by step.

matrix: A rectangular array; a visualization technique using a checkerboard to represent how two variables align. One set of values is represented by rows in the matrix; a second set of values is represented by columns.

mean: The average of a set of numbers.

means-end reasoning: Describes the practical problem-solving process that links actions to expected ends or results; means-end reasoning does not help in evaluating which ends or results are desirable.

mechanism: The second structural stage of a three stage input-mechanism-output model: the operations that run on input in order to produce the output.

median: The midpoint of a set of numbers that has been ordered from lowest to highest; that point at which there are as many below as above.

mental set: In perception, the background expectation that may influence what is perceived; that is, when one is expecting a normal deck of cards, mental set may lead one to ignore altered cards, such as a red ace of spades.
Mill’s joint method of agreement and difference: In John Stuart Mill’s *System of Logic*, evidence for causal linkage between $X$ and $Y$ is stronger when conditions for both Mill’s method of agreement and Mill’s method of difference are present; that is, $Y$ appears whenever $X$ is present and disappears when it isn’t.

Mill’s method of agreement: In John Stuart Mill’s *System of Logic*, if two cases are alike only in $X$, and $Y$ occurs in both, we have evidence that $X$ and $Y$ are causally linked.

Mill’s method of concomitant variation: In John Stuart Mill’s *System of Logic*, applies to cases in which there is a range of values for $X$ and $Y$. If the amount or strength of $Y$ varies as the amount or strength of $X$ varies, there is evidence for a causal link between $X$ and $Y$.

Mill’s method of difference: In John Stuart Mill’s *System of Logic*, if the presence or absence of $X$ is the only difference between two cases, with $Y$ a result in the first case and not in the second, we have an indication that $X$ causes $Y$.

mode: The most frequently occurring value in a set of numbers.

mutually exclusive events: Events are mutually exclusive if the presence of one categorically excludes the presence of the other; both cannot occur.

mysticism: A claim to an immediate, intuitive, and nonexperiential knowledge of reality.

necessary condition: $X$ is a necessary condition for $Y$ if one must have $X$ in order to have $Y$; one cannot have $Y$ without $X$. See also sufficient condition.

negotiation strategy: An approach to conflict resolution that may attempt to remove the “contest of wills” characteristic of positional negotiation. Among other techniques, negotiation strategies may include employing ego-distancing; talking about issues without identifying with a particular position; “going to the balcony,” that is, trying to put emotions aside and view the problem from a distance; appealing to independent standards;
coming to some agreement about what kinds of objective criteria could help to clarify or settle the issue; and replacing debate with collaborative research on the topic.

**normative**: Used to designate a claim that is evaluative in nature or recommends a course of action, as opposed to descriptive.

**nous**: Reason; according to Plato, the first element of the soul and that which should rule.

**null hypothesis**: In experimental design, the prediction that there will be no detectable differences between the experimental conditions.

**open file**: In chess, a file is a vertical row with no pawns of either color on it, allowing a clear route into enemy territory for the queen or a rook.

**outlier**: An extreme value; something that is out of the ordinary. In statistics, a number that is extremely divergent from all the others in a set of numbers; outliers have misleading impact on the mean, or average, of a set of numbers.

**output**: The third structural stage of a three stage input-mechanism-output model: the result of a model run given a particular input.

**overconfidence heuristic**: The psychological tendency for people to overestimate their own abilities, also known as the Lake Wobegon effect.

**particular proposition**: In logic, a proposition about “some” rather than “all”: “Some S are P” (a particular positive, e.g., some cleaning products are poisons) or “Some S are not P” (a particular negative, e.g., some cleaning products are not poisons). The particular positive occupies the lower-left corner of Aristotle’s square of opposition; the particular negative occupies the lower-right corner.

**Pascal’s wager**: A decision-theoretic argument offered by Blaise Pascal in the 17th century with the conclusion that it is rational to believe in God. If you believe in God and he exists, your payoff is eternal bliss; if you believe in God and he does not exist, your payoff is being wrong, a small loss; if
you do not believe in God and he does not exist, you have the pleasure of being right, a small gain; if you do not believe in God and he does exist, your payoff is eternal damnation. Given the potential gains and losses, expected utility dictates that the most rational thing to do is believe in God.

pathos: Emotion; according to Aristotle, the second quality of a persuasive presentation is that it resonates emotionally with the listener.

pattern recognition: The ability to recognize a set of stimuli arranged in specific configurations or arrays, for example, to recognize faces as faces rather than patches of color or melodies as melodies rather than merely sequences of notes.

perceptual bias: A “hard-wired” tendency in our perceptual processing that forces us to perceive things in particular ways. Our color perception does not track pure wavelengths of light or actual lengths in a stimulus, for example, because our visual processing has evolved to interpret input immediately in terms of contextual cues regarding shadow and perspective.

goisoning-the-well fallacy: A fallacy that depends on an attack against a person’s motives for saying something rather than a refutation of the claims being made; a subtype of ad hominem.

polarization: Radical or extreme disagreement between groups with no apparent willingness to compromise and/or with few individuals representing a middle group between the extreme positions. Polarization normally implies a wide gap between positions and increased uniformity within positions. Political polarization refers to extreme positions taken by political organizations, either major political parties or smaller interest groups; cultural polarization refers to extreme differences in attitudes of the general public that may or may not be expressed politically.

positional negotiation: In conflict resolution scenarios, an approach in which people are ego-involved or identify with their specific positions. Those involved in positional negotiation often end up in one of two roles: as the “soft” negotiator, who tries to avoid conflict by giving in and winds up feeling exploited, or as the “hard” negotiator, who is out to win at all
costs and, thus, starts with an absurd extreme, allowing room to make some concessions and still hit his or her initial target.

**post hoc ergo propter hoc**: “After it, therefore because of it”; a fallacy based on the claim that because something followed another thing, it must have been because of that other thing. This fallacy overlooks the possibility of coincidental occurrence. Abbreviated as *post hoc*.

**prediction**: In a three-stage model, information about input(s) and mechanism is used to forecast the future, that is, to generate information about probable output(s). See **input**, **mechanism**, **output**.

**prefrontal cortex**: The anterior portion of the brain, which lies just behind the forehead, heavily involved in complex planning and decision making.

**premise(s)**: The proposition(s) or claims that are given as support for a conclusion; in a rational argument, the reasoning flows from the premises to the conclusions. See **conclusion**.

**Prisoner’s Dilemma**: In game theory, a two-person game in which the value of mutual cooperation is greater than the value of joint defection, the value of defecting against a cooperator is greater than the value of mutual cooperation, but the value of cooperating against a defector is lower than the value of mutual defection, as illustrated in the following matrix:

<table>
<thead>
<tr>
<th></th>
<th>Cooperate</th>
<th>Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperate</td>
<td>3, 3</td>
<td>0, 5</td>
</tr>
<tr>
<td>Defect</td>
<td>5, 0</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

*Prisoner’s Dilemma Matrix*

In the Prisoner’s Dilemma, the real loser is cooperation in the face of defection; for alternative payoff scenarios, see **Chicken** and **Stag Hunt**.

**probability**: The ratio calculated by dividing the number of possible outcomes in a particular category by the total number of possible outcomes.
**proposition**: A claim, statement, or assertion; the message or meaning behind the words in a written or spoken sentence; the information a sentence expresses or conveys.

**prospect theory**: In the work of Daniel Kahneman and Amos Tversky, a descriptive theory concerning how people make decisions in terms of prospective loss and gain rather than final outcome alone.

**pseudoscience**: A set of practices or tenets that is made to resemble science and claims to be scientific but violates scientific standards of evidence, rigor, openness, refutation, or the like.

**Ptolemaic astronomy**: The view of astronomy put forward by Ptolemy (A.D. 90–168); based on a geocentric (earth-centered) model of the universe.

**Pythagorean theorem**: In Euclidean geometry, the relation of the three sides of a right triangle, represented by the formula $a^2 + b^2 = c^2$. The area of the square whose side is the hypotenuse of the right triangle (the side opposite the right angle) is equal to the sum of the area of the squares of the other two sides.

**quantum mechanics**: Developed early in the 20th century, quantum mechanics is a sophisticated theory of physics at the subatomic scale; a mathematically elegant, empirically established but philosophically puzzling theory of the very small.

**questionnaire bias**: A way of phrasing a question that influences the way people answer the question. Depending upon how the question is phrased, respondents may be more likely to respond in a negative or a positive direction than they would if the question were phrased in an unbiased, or neutral, way.

**randomization**: In experimentation, a procedure for assigning groups so that the differences between individuals within groups distribute with equiprobability across experimental conditions.

**rational argument**: A way of presenting and supporting claims that relies on logical transition from premises to conclusion.
**rationality**: Exercising reason, that is, analytical or logical thought, as opposed to emotionality.

**recognition heuristic**: A fast-and-frugal rule of thumb for decision making in which one picks the alternative or orders alternatives in terms of recognition; often most effective when one is working on very limited information about alternatives.

**relativity theory**: Proposed as an alternative to Newtonian physics, Einstein’s theory of relativity (general and special) holds that matter and energy are interchangeable, that time moves at a rate relative to one’s rate of speed, and that space itself can be curved by gravity. In large part, relativity theory is a series of deductions from the assumption that some things are not relative. For example, assuming that the speed of light is a constant, Einstein showed that observable simultaneity will be relative to the comparative motion of two observers.

**retrodiction**: In a three-stage model, information about the mechanism and output(s) is used to generate information about the past (inputs). See **input**, **mechanism**, **output**.

**retrograde motion**: In astronomy, the apparent movement of a body opposite to that of other comparison bodies. Plotted against constellations night by night, Mars moves in one direction for a while but then appears to move backwards before it again seems to reverse directions. Retrograde motion does not fit with the predictions of a simple geocentric theory, which holds that the sun and planets circle around the earth.

**rhetoric**: The skills of effective speaking and presentation of ideas; also, the techniques of persuasion, fair or foul, for either good ends or bad.

**risk aversion**: In economics, an investor is considered risk averse if he or she avoids risk as opposed to being risk taking. People’s choices involving possible losses are often more risk averse than similar choices regarding possible gains.
**risk taking**: In economics, an investor is considered risk taking if he or she is not opposed to taking risks. Choices involving possible gains are often more risk taking than similar choices regarding possible losses.

**sampling issues**: Problems with the group from which statistics are derived that may lead to misrepresentations about the general characteristics of the group. Some sampling issues are size (the sample group is too small relative to the population to which it is being generalized) and bias (the sample group has been selected in a way that overrepresents certain subgroups and underrepresents other subgroups of the population).

**satisficing**: A heuristic, or rule of thumb, for decision making in which one picks a realistic goal or set of goals, then selects the first course of action that meets the goal or goals. This heuristic emphasizes that one is not seeking perfection, just something that is “good enough.”

**Scholastics**: Philosophers practicing the dominant Western Christian philosophy of the European Middle Ages.

**selective memory**: The phenomenon in which recall of past events is uneven, with particular classes of events overemphasized because of emotional tone, individual interest, and personal history. Gamblers have been shown to have a selective memory that favors past winnings over past losses.

**sentence**: A sentence is a series of words, spoken or written, that expresses a claim, statement, or assertion. See also *proposition*.

**set theory**: The branch of mathematics that studies the properties of sets and their interrelations—abstract properties of collections, regardless of what they contain.

**simplification**: Conceptual replacement of a complex concept, event, or phenomenon with a stripped-down version that is easier to understand or to manipulate mentally.

**sound argument**: An argument that is both valid—that is, the premises lead by strong logic to the conclusion—and in which the premises are true.
**square of opposition**: Aristotle’s visualization of the logical relations between categorical propositions.

**Stag Hunt**: In game theory, a two-person game in which the value of mutual cooperation is greater than the value of either defecting against a cooperator or mutual defection, as illustrated in the following matrix:

\[
\begin{array}{c|cc}
& \text{Cooperate} & \text{Defect} \\
\hline
\text{Cooperate} & 5, 5 & 0, 1 \\
\text{Defect} & 1, 0 & 1, 1 \\
\end{array}
\]

*Stag Hunt Matrix*

In the Stag Hunt, the real winner is mutual cooperation; for alternative payoff scenarios, see **Chicken** and **Prisoner’s Dilemma**.

**statistical significance**: A mathematical calculation that is used to determine whether the results of an observation or experiment might have occurred by chance alone.

**Stoics**: The Stoics held that the rational road to tranquility is to control one’s emotional reactions to such an extent that one becomes impervious to the slings and arrows of outrageous fortune.

**straw man fallacy**: The rhetorical strategy of exaggerating or misrepresenting the claims of the opposition in order to more easily appear to refute those claims.
**stress test**: A technique used to examine the strength or stability of an entity under operational conditions that are more extreme than what is expected normally. In analyzing argument flow, a technique for detecting dependency between reasons by eliminating them individually in order to see whether the argument still goes through.

**subcontraries**: The relationship between propositions at the bottom left (“Some S are P”) and right (“Some S are not P”) of Aristotle’s square of opposition. If two propositions are subcontraries, it is possible for both propositions to be true, but it is not possible for both propositions to be false. See also contraries, square of opposition.

**sufficient condition**: $X$ is a sufficient condition for $Y$ if $X$ is enough; once one has $X$, $Y$ is guaranteed. See also necessary condition.

**syllogism**: An argument using three categories (A, B, and C) that are linked in pairs to make categorical propositions (e.g., “All A’s are B’s,” or “Some B’s are not C’s”), which are then combined into a three-step argument. Some syllogisms are valid and some are not. See categorical proposition, validity.

**thermodynamics**: A branch of physics that is concerned with the relationships of heat, pressure, and energy.

**thumos**: Spirit or fortitude shown in battle; according to Plato, the third element of the soul.

**tit for tat**: In game theory, a strategy for playing an iterated game in which a player mirrors the play made by the opponent in the previous round; the player cooperates if the opponent cooperated in the previous round but defects if the opponent defected in the previous round.

**tripartite theory of the soul**: Plato’s theory that the soul includes three parts: *nous, epithumia*, and *thumos*.

**trivium**: The basis of a classical education throughout the Middle Ages; a three-part curriculum composed of grammar, logic, and rhetoric.
**tu quoque**: A fallacy in reasoning that tries to defuse an argument by claiming, “You did it, too.”

**Ultimatum Game**: In game theory, a game in which two participants are given the task of splitting a payoff, for example, a certain amount of money. One participant is randomly selected to be the Proposer of how to make the split, and the other, the Responder, must decide whether to accept the offer. If the Responder accepts, the payoff is split between them as proposed; if the Responder rejects the offer, no one gets a payoff.

**universal generalization**: A statement about “all” of a particular class: “All X’s are Y’s.”

**universal proposition**: In logic, a universal proposition refers to a claim either in the form “All S are P” (universal affirmative, e.g., all snakes are reptiles) or in the form “No S are P” (universal negative, e.g., no snakes are reptiles). The universal affirmative occupies the upper-left corner of Aristotle’s square of opposition; the universal negative occupies the upper right.

**valence**: A psychological term describing the positive or negative value (often emotional) that we ascribe to things.

**validity**: An argument is valid if the conclusion follows from the premises, if the premises offer sufficient logical or evidential support for the conclusion. An argument is deductively valid if it is impossible for all premises to be true and the conclusion to be false.

**value**: Desirability; psychological research indicates that relative valuations tend to be inconsistent over time and sensitive to context.

**Venn diagram**: A way of visualizing the relations between concepts and their extensions through the use of overlapping circles.

**verifiability**: The criterion for what distinguishes science from non-science, proposed by a group of 20th-century scientists and philosophers called the Vienna Circle. Science, they said, is that which can be confirmed by empirical or experiential data. See also **falsifiability**.
**visualization**: The process of using diagrams or imagery as an aid to thought in representing a problem, calculation, or set of possibilities.

**warrant**: A general underlying principle that licenses an inference from data to a conclusion. In a probability warrant, the strength of the link between premise and conclusion is expressed in terms of probabilistic connection (e.g., 90 percent of the time, premise A is linked to conclusion B). In a definitional warrant, the premises are linked to conclusion as a matter of definition (e.g., whales are mammals by definition because they breathe air and give live birth). A legal warrant relies on a point of law as the link between the premise and conclusion (e.g., a contract requires a signature; thus, this unsigned document is unenforceable). An ethical warrant relies on an underlying ethical belief (e.g., if there is a shared belief that one should not deceive, then the conclusion that a deliberately deceitful act was wrong is warranted).

**Wason selection task**: A psychological task used to examine the psychology of reasoning. If shown four cards, one displaying the letter E, one the letter K, one the number 4, and one the number 7, using a minimum number of cards, which cards do you turn over to test the theory “If a card has a vowel on one side, it has an even number on the other”? It has been demonstrated that people tend to seek confirmation when working with unfamiliar or abstract concepts, such as “vowels” and “even numbers,” but will seek falsifiability in similar tasks using familiar situations, such as underage drinking.

**wisdom-of-crowds heuristic**: A rule of thumb in which one bases one’s decision on the most popular answer selected by a random group of people. In order for this heuristic to be effective, certain conditions of randomness must be met. The reason it works can be explained statistically by the fact that in an appropriately random group of people, most will be ignorant about the topic, but a few will actually know the answer. The answers of the “ignorant” will distribute randomly and cancel each other out, allowing the answers of those who really know to tip the scale.

**word**: A linguistic representation of an idea; as distinct from concept.


Best, Joel. Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists. Berkeley: University of California Press, 2001. The source of “the worst statistic ever”: the claim that the number of children killed with guns has doubled each year since 1950.

Bishop, Bill. *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*. Boston: Houghton Mifflin Harcourt, 2009. Bishop makes the case that polarization is the result of our increased tendency to associate only with others like ourselves.


Crease, Robert P. *The Prism and the Pendulum: The Ten Most Beautiful Experiments in Science*. New York: Random House, 2004. What are the most beautiful experiments? Crease is a columnist for *Physics World* and
asked that question of his readers. Here, he offers a philosophical discussion of some of the foremost candidates.


Feynman, Richard P., Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics*. Reading, MA: Addison-Wesley, 1963. Physics that can be read for the entertainment value alone. The best and most accessible introduction to aspects of physics that anyone has produced or is likely to.


their book in terms of practical hints on polarized negotiation, but their suggestions apply to polarized discussion generally.


Grim, Patrick, ed. *Philosophy of Science and the Occult*. 1st and 2nd eds. Albany, NY: SUNY Press, 1982, 1991. The core of the book is devoted to the problem of demarcation, with other sections that focus on particular topics, such as astrology and parapsychology. UFOlogy has a section in the first edition, replaced with one on quantum mysticism in the second.

MIT Press, 1998. Logic and game theory as seen with the techniques of computer modeling.


Hume, David. “Of the Passions.” In An Enquiry Concerning Human Understanding, 1748. A contemporary version is edited by Tom L. Beauchamp in the Oxford Philosophical Texts series (New York: Oxford University Press, 1999). This is the classic source for “emotion, not reason, ought to rule.”


Kelley, David. The Art of Reasoning. 3rd ed. W. W. Norton, 1998. There is no current informal logic book that I can recommend without reservation, but Kelley’s has two good chapters on Venn diagrams for categorical propositions and syllogisms.


Schopenhauer, Arthur. The Art of Controversy. New York: Cosimo Classics, 2007. Also freely available online. Schopenhauer is often infuriating; a valuable read for precisely that reason.


Simons, Daniel J. Surprising Studies of Visual Awareness. Champaign, IL: VisCog Productions, 2008. This DVD includes full video of the “invisible gorilla” and a number of astounding change-blindness experiments.

Sorensen, Roy A. *Thought Experiments*. New York: Oxford University Press, 1992. Sorensen argues that thought experiments are literally that: experiments. He includes a range of wonderful examples in the discussion.

Sterrett, Susan G. *Wittgenstein Flies a Kite: A Story of Models of Wings and Models of the World*. New York: Pi Press, 2006. This is a profound book, both theoretical and historical, on the obstacles conceptual modeling faces in scaling up to the real world. Well worth reading, despite some technically difficult later chapters.


University of Illinois Advertising Exhibit. http://www.library.illinois.edu/adexhibit/technology.htm. Sometimes it’s easier to see through advertising ploys in older ads. Here, you will find a great selection of old ads by topic (celebrities, tobacco, diet).


Lecture 1, Exercise

Lines 1, 7, and 12 come from John Keats, “Ode to a Nightingale.”

Lines 2, 5, and 10 come from Robert Burns, “A Red, Red Rose.”

Lines 3, 8, and 11 come from Walt Whitman, “Out of the Cradle Endlessly Rocking.”

Lines 4, 6, and 9 come from T. S. Eliot, “The Love Song of J. Alfred Prufrock.”

Lecture 3, Exercise

In a sketch, the initial information lays out like this:

<table>
<thead>
<tr>
<th>Phil</th>
<th>Bill</th>
<th>John</th>
<th>2 x Bill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$50</td>
<td></td>
<td></td>
<td>$100</td>
</tr>
</tbody>
</table>

If John makes $900 a week, twice Bill’s salary is $1,000, and half of that is $500. Phil makes $50 less; thus, he must make $450 a week.
Lecture 6, Exercise

Here, we’ve rearranged Aristotle’s square of opposition. Your sketch should look something like this:

![Venn Diagram](image)

Lecture 7, Exercises

Here are progressive diagrams for premises and conclusions in the exercises:

![Venn Diagrams](image)
Step 1: “No M are P” means that nothing is in the intersection of M and P—we blank that out.

Step 2: “All Q are P” means that nothing in Q is outside of the P area. We add that information in the second diagram.

Step 3: The conclusion, “No M are Q,” would mean that nothing is in the overlap of M and Q. But that information is already in the second diagram. Given the premises, the conclusion must be true already; the argument is valid.

Step 1: “All M are P” means that nothing is in M outside of the P area.

Step 2: We add “All Q are P”: There are no Q’s outside of the P area.

Step 3: The conclusion would tell us that there is definitely something in the M area outside of Q. That would be new information—something that doesn’t follow from the premises alone. The argument is invalid.
Step 1: We sketch in “All M are P.”

Step 2: “Some P are Q” tells us that there is at least one thing in the overlap between P and Q. But it doesn’t tell us where in that overlap, so we leave it “indefinite” on the line. It might be on either side.

Step 3: The conclusion tells us that some M are Q: that there is an X on that particular side of the line. That goes beyond the information we already have and, thus, doesn’t follow from the premises. The argument is invalid.

Lecture 8, Questions to Consider

(a) Collect 1985 pennies and you will have $19.85.

(b) No, because he would be dead, and it’s not legal for a dead person to marry.

(c) Meat.

(d) Those letters can be rearranged to form: one word.

(e) They are two of a set of triplets.

Lecture 10, Exercise

(a) Rio de Janeiro

(b) Bill Gates, Larry Ellison, Christy Walton

(c) Sudan

(d) Ferrell, Drysdale, Johnson

On which questions did the recognition heuristic work for you? On which questions did it fail? Can you tell why?
Lecture 14, Questions to Consider

*Post hoc ergo propter hoc.* Because the magician “disappears” one coin at point A and then produces a second coin at point B, we are led to believe that the second event happened because the first one did: that the disappearance “caused” the appearance and that a single coin has “traveled” from point A to B.

Lecture 17, Exercise

There are lots of ways to construct the quiz scores, but here’s an example. Let the children’s scores be:

1 1 1 1 2 3 8 9 10 10 10

(a) The mode is 1. Create a “most people” category first and aim for something different than the “middle,” which you can expect to be more like the mean or the median.

(b) The median is 3, with as many above as below.

(c) The mean is 5.09. Having kept the median low, you can make the mean significantly higher by adding high numbers at the end.

Lecture 19, Exercise

Here’s how to start thinking about the exercise:

Option 1 has a 20 percent chance of gaining $1 million over investment and an 80 percent chance of gaining nothing. Probability multiplied by desirability is \(0.2 \times \$1\text{ million} = \$200,000\) expected utility.

Option 2 has a 40 percent chance of gaining $500,000 over investment. Probability multiplied by desirability is \(0.4 \times \$500,000 = \$200,000\) expected utility.
Option 3 has a 90 percent chance of gaining $150,000 over investment. Probability multiplied by desirability is $0.9 \times $150,000 = $135,000 expected utility.

But that may be only the start of your deliberations. Are there other considerations you would take into account?