Philosophy: What We Can Know, What We Owe, and How to Live

Figures current as of October 2026

The questions under every other subject, argued: what we can know and whom to believe; what a mind is, whether animals and machines have one, whether we are free, and what makes a person the same person over time; whether God exists; whether morality is objective, which acts are right, what we owe strangers and the future, and when ending a life is wrong; what justice requires and when a state's power is legitimate; and how to live. Each answer gets its best case as its defenders make it, the objection it must meet, and what it costs, with thinkers from Greece, India, China, the Islamic world, Africa, and Europe where their arguments are strongest. A dated watch list closes the course.

MODULE 1 — What Can We Know?

Goal: Test a definition the way philosophers do, and weigh whether the skeptic can be answered.

The definition of knowledge long credited to Plato and the three pages of 1963 that broke it, anticipated by the Buddhist philosopher Dharmottara around 770; dreams, demons, and Zhuangzi's butterfly; reason against experience, and India's debate over the sources of knowledge; whether intuitions count as evidence; and the simulation argument's arithmetic.

MODULE 2 — Whom to Believe

Goal: Decide when another person's word is reason enough, and weigh what to do when equals disagree.

Testimony, from Hume and Reid to the Nyāya school; how science earns belief, from induction to Popper, Kuhn, and Bayes; how a novice can judge experts; echo chambers; whether belief needs evidence; chatbots as witnesses; and two friends who split a dinner check differently.

MODULE 3 — Consciousness

Goal: Say what any theory of the mind must explain, and weigh whether consciousness is physical.

Descartes's two substances and Princess Elisabeth's objection; Avicenna's floating man; the theories that made the mind a brain or a program; the hard problem, from Leibniz's mill to Mary's room; the answers that deny the problem or spread consciousness everywhere; and the 2025 test of two scientific theories.

MODULE 4 — Other Minds: Animals and Machines

Goal: Weigh the evidence that other beings feel, and decide what follows if they might.

How anyone knows another mind; the evidence for feeling in mammals, birds, fish, octopuses, and insects; the scale of farming; Bentham's question and the charge of speciesism; Turing's test, the Chinese room, and the signs of consciousness in AI; and the first steps toward AI welfare.

MODULE 5 — Free Will and Responsibility

Goal: Decide what blame requires, and weigh whether anyone has it.

Determinism and its three answers; the Stoics' cylinder, Frankfurt's cases, and the reactive attitudes; Islam's debate over God's power and human acts; Libet's experiment and its reinterpretation; what ordinary people say; and punishment without desert.

MODULE 6 — Personal Identity

Goal: Say what makes a person the same over time, and weigh whether identity is what matters.

The ship of Theseus; Locke's memory theory and Reid's brave officer; the psychological view, animalism, and Parfit's teletransporter; the Buddha's no-self and Nāgasena's chariot; personhood as achieved; and dementia, future selves, and digital copies of the dead.

MODULE 7 — Does God Exist?

Goal: Know the strongest arguments on each side, and what each side must answer.

The cosmological argument from Aristotle and al-Ghazālī to Leibniz; design and fine-tuning; Anselm's proof from the idea of God; the problem of evil and the replies; experience and hiddenness; Udayana and his Buddhist critics; faith and reason; and Pascal's wager as a decision table.

MODULE 8 — Is Morality Objective?

Goal: Decide whether moral claims can be true, and what follows if they cannot.

Herodotus's Greeks and Callatians; relativism and the reformer; the Euthyphro dilemma; Hume's is and ought; realism, error theory, expressivism, and constructivism; Mencius against Xunzi on human nature; the evolutionary challenge; and whether abolition was progress.

MODULE 9 — Right and Wrong

Goal: Run a hard case through each theory, and weigh which theory to trust.

Mozi and Bentham on consequences; Kant on duty, rights, and double effect; Aristotle, Confucius, and Mencius on character, and the evidence from social psychology; contracts, care, and Ubuntu; Arjuna's dilemma; the trolley; and the self-driving car's crash rule.

MODULE 10 — What We Owe: Strangers and the Future

Goal: Put a number on a duty to strangers, and weigh how far it reaches.

Singer's drowning child; Mozi's inclusive care against Mencius's graded love; partiality and demandingness; effective altruism and its critics; the aid cuts of 2025; borders; future people, the Repugnant Conclusion, and the discount rate; and what a tenth of a household's income buys.

MODULE 11 — Life and Death

Goal: Weigh why death is bad and when ending a life is wrong, with each side's strongest case.

Epicurus and Lucretius against the deprivation account; abortion, from personhood to Thomson's violinist and the religious traditions; assisted dying, from autonomy to the slippery slope; the laws as of October 2026; and Canada's numbers since 2016.

MODULE 12 — Justice: Who Should Get What

Goal: Choose a fair distribution, and defend it against the other theories.

Aristotle's proportion; Rawls's veil of ignorance; Nozick's basketball player; utilitarian redistribution; luck, relationships, and capabilities; Marx; what people choose behind a veil in experiments; and three societies ranked by each theory.

MODULE 13 — The State: Authority, Democracy, and Liberty

Goal: Say when power is legitimate, and where it must stop.

Socrates in prison; Hobbes, Locke, Rousseau, and Hume's sleeping passenger; the Mandate of Heaven; Plato against democracy, and rule by the knowledgeable; Condorcet's jury theorem, worked; Mill's harm principle; and civil disobedience from Thoreau to King.

MODULE 14 — How to Live

Goal: Decide what a good life consists of, and weigh whether life has a meaning.

Pleasure, desire, or a list of goods, and the experience machine; Aristotle's flourishing; the Stoics, Epicurus, the Buddha, Confucius, and Zhuangzi; Kierkegaard, Sartre, Beauvoir, and Camus; meaning with and without God; and what income does to happiness.

MODULE 15 — Where Things Stand, and What to Watch

Goal: Close with the argument, the contradictions, and a dated watch list.

The course's argument on a page; the contradictions no single module states; whether philosophy makes progress; what to watch from October 2026 on; and a last word.

A note on the sources: where the course says what professional philosophers think, it uses the 2020 PhilPapers Survey (David Bourget and David Chalmers, *Philosophers' Imprint*, 2023), set out with its cautions in the table that follows. Each module's reference is usually an entry in the *Stanford Encyclopedia of Philosophy*, a free online reference written and reviewed by specialists.

Older texts are cited by their usual titles, in standard English translations, with the date each was written or first published; dates before the year 1 are marked BC, and the dates of ancient thinkers are approximate. Figures that change, from the number of animals farmed to the laws on assisted dying, name their source and date.

Where the course takes no side, it says so. On the existence of God, the nature of consciousness, free will, whether morality is objective, abortion, assisted dying, and eating animals, it sets out the strongest case for each answer and where the argument stands, and leaves the verdict to the reader. Where evidence or broad agreement settles a narrower point, the course says that too, once. Each module ends its argument with a table of what each answer costs: what it explains, and what it must accept or explain away.

Where Philosophers Stand, 2020

The course's questions, as 1,785 professional philosophers answered them

The PhilPapers Survey invited 7,685 philosophers to say which answer they accept or lean toward on each of 100 questions: the faculty of philosophy departments in Australia, Canada, Ireland, New Zealand, the UK, and the US, and philosophers who publish in English elsewhere. Of those invited, 1,785 completed it between October 15 and November 16, 2020, and it remains the most recent survey of its kind. Shares count "accept" and "lean toward" together, and a respondent who chose two answers counts for both, so a row need not add to 100.

Question	Answers (share who accept or lean toward each)	Module
Do we know the world outside the mind?	Non-skeptical realism 79.5%; idealism 6.6%; skepticism 5.4%	1
Is there a priori knowledge, known without experience?	Yes 72.8%; no 18.5%	1
Where does knowledge come from?	Empiricism 43.9%; rationalism 33.5%	1
Does science describe a real, unobserved world?	Scientific realism 72.4%; anti-realism 15.0%	2
Is the mind physical?	Physicalism 51.9%; non-physicalism 32.1%	3
Is there a hard problem of consciousness?*	Yes 62.4%; no 29.8%	3
Could a being exist that acts like us but feels nothing (a zombie)?	Conceivable but not possible 36.5%; possible 24.4%; inconceivable 16.4%	3
Which beings have conscious experience?*	Adult humans 95.1%; cats 88.6%; newborns 84.3%; fish 65.3%; flies 34.5%; worms 24.2%; plants 7.2%; current AI systems 3.4%; future AI systems 39.2%	4
Is eating animals and animal products permissible?	Both permissible 48.0%; animal products only (vegetarianism) 26.5%; neither (veganism) 18.4%	4
Free will	Compatibilism 59.2%; libertarianism 18.8%; no free will 11.2%	5
Is capital punishment permissible?*	Impermissible 75.1%; permissible 17.7%	5
What makes a person the same over time?	Psychological continuity 43.7%; the body or organism 19.1%; a further fact 14.9%	6
Does a person survive a teletransporter that rebuilds the body from new matter?	Death 40.1%; survival 35.2%	6
God	Atheism 66.9%; theism 18.9%; other answers 14.0%	7
Are there objective moral facts?	Moral realism 62.1%; anti-realism 26.1%	8
Do moral judgments state beliefs that can be true?	Cognitivism 69.3%; non-cognitivism 20.7%	8
Which theory of right action?	Virtue ethics 37.0%; deontology 32.1%; consequentialism 30.6%	9
Trolley: switch a runaway trolley to kill one instead of five?	Switch 63.4%; don't switch 13.3%	9
Footbridge: push one man to stop the trolley and save five?	Don't push 56.0%; push 22.0%	9
Would you choose to be immortal?*	Yes 44.9%; no 41.3%	11
Is abortion permissible (first trimester, no special circumstances)?*	Permissible 81.7%; impermissible 13.1%	11
Political philosophy	Egalitarianism 44.0%; communitarianism 27.3%; libertarianism 13.4%	12
Economic system*	Socialism 53.0%; capitalism 29.5%	12

Question	Answers (share who accept or lean toward each)	Module
Would you plug into a machine that gives you any experiences you want?	No 76.9%; yes 13.3%	14
What makes a life go well?*	A list of objective goods 53.2%; getting what you want 18.6%; pleasure (hedonism) 12.7%	14
Is there a meaning of life?	Subjective 33.0%; objective 32.1%; none 16.1%	14
Does philosophy make progress?	A little 46.6%; a lot 41.7%; none 3.8%	15

Source: David Bourget and David Chalmers, “Philosophers on Philosophy: The 2020 PhilPapers Survey,” *Philosophers’ Imprint* 23 (2023), target-faculty respondents. Questions marked * were among 60 additional questions, each answered by about 1,100 respondents. Adjusting for who chose to respond lowers some majorities: non-skeptical realism to 75.2%, scientific realism to 64.7%, and physicalism to 48.5%. The survey did not ask about assisted dying.

Three cautions apply to every row. The survey reports what experts believe, which is evidence about a question but does not settle it, since experts in every field have been wrong together. Its respondents are mostly in English-speaking universities, so it measures one tradition’s profession, not the world’s thinkers. And the profession leans left and secular, as the rows on God and on the economy show, which bears on the questions where values and politics meet.

Who's Who

The thinkers whose arguments the course sets out, as of October 25, 2026

The people below are those whose arguments, cases, or findings the course sets out, grouped by period and listed in order of birth within each group; living thinkers whose year of birth has not been published come last. Dates of ancient thinkers are approximate. Most researchers cited for a single study, and the authors of the readings, are named where they appear and are not listed here.

Ancient Greece and Rome

Name	Dates	Tradition and field	Modules
Herodotus	about 484–425 BC	Greek historian	8
Socrates	about 470–399 BC	Greek philosopher, known through Plato’s dialogues	1, 8, 9, 13, 15
Plato	about 427–347 BC	Greek philosopher	1, 8, 9, 13, 15
Aristotle	384–322 BC	Greek philosopher	7, 9, 12, 14
Pyrrho of Elis	about 360–270 BC	Greek skeptic	1
Epicurus	341–270 BC	Greek; founder of Epicureanism	7, 11, 14
Chrysippus	about 279–206 BC	Greek Stoic	5
Lucretius	about 99–55 BC	Roman Epicurean poet	11
Epictetus	about AD 50–135	Greek Stoic, born a slave	14
Sextus Empiricus	about AD 160–210	Greek skeptic and physician	1

China and India

Name	Dates	Tradition and field	Modules
Confucius	551–479 BC	Chinese; founder of Confucianism	9, 14
The Buddha (Siddhārtha Gautama)	about 480–400 BC	Indian; founder of Buddhism	6, 14
Mozi	about 470–391 BC	Chinese; founder of Mohism	9, 10
Mencius	about 372–289 BC	Chinese Confucian	8, 9, 10, 13
Zhuangzi	about 369–286 BC	Chinese Daoist	1, 14
Xunzi	about 310–235 BC	Chinese Confucian	8
Nāgasena	2nd century BC	Indian Buddhist monk	6
Dharmottara	about 740–800	Indian Buddhist, of Kashmir	1
Udayana	late 10th century	Indian, of the Nyāya school	7

The Islamic world and medieval Europe

Name	Dates	Tradition and field	Modules
Augustine of Hippo	354–430	North African Christian theologian	5, 13
al-Ashʿarī	about 874–935	Arab Muslim theologian	5, 8
Avicenna (Ibn Sīnā)	980–1037	Persian philosopher and physician	3
Anselm of Canterbury	1033–1109	Italian-born Christian theologian	7
al-Ghazālī	1058–1111	Persian Muslim theologian and philosopher	7
Averroes (Ibn Rushd)	1126–1198	Andalusian Muslim philosopher and judge	7
Thomas Aquinas	1225–1274	Italian Christian theologian	7, 9, 10

Early modern Europe

Name	Dates	Tradition and field	Modules
Thomas Hobbes	1588–1679	English political philosopher	5, 6, 13
René Descartes	1596–1650	French rationalist	1, 2, 3, 4
Elisabeth of Bohemia	1618–1680	Princess; critic of Descartes	3
Blaise Pascal	1623–1662	French mathematician and Christian thinker	2, 7
Baruch Spinoza	1632–1677	Dutch rationalist	5, 7
John Locke	1632–1704	English empiricist and political philosopher	1, 6, 12, 13
Gottfried Wilhelm Leibniz	1646–1716	German rationalist	1, 3, 7
Thomas Reid	1710–1796	Scottish philosopher of common sense	2, 5, 6
David Hume	1711–1776	Scottish empiricist	1, 2, 5, 6, 7, 8, 13
Jean-Jacques Rousseau	1712–1778	Genevan political philosopher	13
Immanuel Kant	1724–1804	German philosopher	1, 4, 5, 7, 8, 9, 10, 11
William Paley	1743–1805	English theologian	7
Marquis de Condorcet	1743–1794	French mathematician and reformer	13
Jeremy Bentham	1748–1832	English; founder of utilitarianism	4, 9, 14
Pierre-Simon Laplace	1749–1827	French mathematician	5

The nineteenth century

Name	Dates	Tradition and field	Modules
John Stuart Mill	1806–1873	English utilitarian and liberal	4, 9, 13
Charles Darwin	1809–1882	English naturalist	4, 7
Søren Kierkegaard	1813–1855	Danish Christian thinker, forerunner of existentialism	7, 14
Henry David Thoreau	1817–1862	American essayist	13
Frederick Douglass	about 1818–1895	American abolitionist	8
Karl Marx	1818–1883	German philosopher and economist	12
Leo Tolstoy	1828–1910	Russian novelist	14
William James	1842–1910	American pragmatist	2, 7
W. K. Clifford	1845–1879	English mathematician and philosopher	2

The twentieth century

Name	Dates	Tradition and field	Modules
Mohandas Gandhi	1869–1948	Indian independence leader	13
Bertrand Russell	1872–1970	British analytic philosopher	2, 7, 15
G. E. Moore	1873–1958	British analytic philosopher	1, 8
W. D. Ross	1877–1971	British moral philosopher	9
Gilbert Ryle	1900–1976	British philosopher of mind	3
Karl Popper	1902–1994	Austrian-British philosopher of science	2
Frank Ramsey	1903–1930	British philosopher and economist	10
Jean-Paul Sartre	1905–1980	French existentialist	14
H. L. A. Hart	1907–1992	British legal philosopher	13
Simone de Beauvoir	1908–1986	French existentialist and feminist	14
W. V. O. Quine	1908–2000	American philosopher of logic and science	2
A. J. Ayer	1910–1989	British logical positivist	8
Alan Turing	1912–1954	British mathematician and computer pioneer	4
Albert Camus	1913–1960	French-Algerian writer	14
Benjamin Libet	1916–2007	American neuroscientist	5
J. L. Mackie	1917–1981	Australian-British philosopher	7, 8, 15
P. F. Strawson	1919–2006	British philosopher	5
Richard Taylor	1919–2003	American philosopher	14
Philippa Foot	1920–2010	British moral philosopher	9
John Harsanyi	1920–2000	Hungarian-American economist	12
John Rawls	1921–2002	American political philosopher	1, 8, 9, 12, 13, 14
John Hick	1922–2012	British philosopher of religion	7
Thomas Kuhn	1922–1996	American historian and philosopher of science	2
Hilary Putnam	1926–2016	American philosopher	2, 3
Edmund Gettier	1927–2021	American epistemologist	1, 15
Judith Jarvis Thomson	1929–2020	American moral philosopher	9, 11
Bernard Williams	1929–2003	British moral philosopher	10, 11
Harry Frankfurt	1929–2023	American philosopher	5, 12
Martin Luther King Jr.	1929–1968	American civil rights leader and minister	13
William Rowe	1931–2015	American philosopher of religion	7
Ronald Dworkin	1931–2013	American legal philosopher	6, 12, 13
Sydney Shoemaker	1931–2022	American philosopher of mind	6
John Searle	1932–2025	American philosopher of mind	4
Daniel Kahneman	1934–2024	Israeli-American psychologist	14
Don Marquis	1935–2022	American ethicist	11
Robert Nozick	1938–2002	American philosopher	1, 9, 12, 14, 15
Alvin Goldman	1938–2024	American epistemologist	1, 2
Robert Kane	1938–2024	American philosopher of free will	5
Tom Regan	1938–2017	American animal-rights philosopher	4
Kwame Gyekye	1939–2019	Ghanaian philosopher	6

Name	Dates	Tradition and field	Modules
Ifeanyi Menkiti	1940–2019	Nigerian philosopher	6
James Rachels	1941–2003	American moral philosopher	9, 11
G. A. Cohen	1941–2009	Canadian-British political philosopher	12
David Lewis	1941–2001	American metaphysician	1, 3
Daniel Dennett	1942–2024	American philosopher of mind	3, 5
Derek Parfit	1942–2017	British moral philosopher	6, 8, 9, 10, 12, 14
Mary Anne Warren	1946–2010	American ethicist	11

Living philosophers and scientists

Name	Dates	Tradition and field	Modules
Alvin Plantinga	born 1932	American philosopher of religion	7
Amartya Sen	born 1933	Indian economist and philosopher	9, 12, 13
Richard Swinburne	born 1934	British philosopher of religion	6, 7
Michael Walzer	born 1935	American political theorist	10
Carol Gilligan	born 1936	American psychologist	9
Thomas Nagel	born 1937	American philosopher	3, 10, 11, 14
T. M. Scanlon	born 1940	American moral philosopher	9
Frank Jackson	born 1943	Australian philosopher	3
Joseph Carens	born 1945	Canadian political theorist	10
Peter Singer	born 1946	Australian moral philosopher	4, 10
Martha Nussbaum	born 1947	American philosopher	12
William Lane Craig	born 1949	American philosopher of religion	7, 14
Christine Korsgaard	born 1952	American moral philosopher	4, 8, 9
Galen Strawson	born 1952	British philosopher	3, 5
Susan Wolf	born 1952	American moral philosopher	6, 10, 14
Thomas Pogge	born 1953	German-American philosopher of global justice	10
Jeremy Waldron	born 1953	New Zealand legal philosopher	13
Timothy Williamson	born 1955	British epistemologist	1
Derk Pereboom	born 1957	American philosopher of free will	5
Robert Sapolsky	born 1957	American neuroscientist	5
Elizabeth Anderson	born 1959	American political philosopher	12
Keith Frankish	born 1962	British philosopher of mind	3
David Chalmers	born 1966	Australian philosopher of mind	3, 4, 15
Eric Schwitzgebel	born 1968	American philosopher	15
Michael Huemer	born 1969	American philosopher	10, 13
Nick Bostrom	born 1973	Swedish philosopher	1
Sharon Street	born 1973	American moral philosopher	8
Joshua Knobe	born 1974	American experimental philosopher	1, 5
Hélène Landemore	born 1976	French political theorist	13
Philip Goff	born 1978	British philosopher of mind	3
Jason Brennan	born 1979	American political philosopher	13

Name	Dates	Tradition and field	Modules
Toby Ord	born 1979	Australian philosopher	10
William MacAskill	born 1987	Scottish philosopher	10
C. Thi Nguyen	living	American philosopher	2
Jonathan Birch	living	British philosopher of biology	4
David Christensen	living	American epistemologist	2
Edouard Machery	living	French-American experimental philosopher	1
Matthew Killingsworth	living	American happiness researcher	14

Terms

The vocabulary of the arguments

Term	Meaning
A priori, a posteriori	Known a priori: known without relying on experience, as in arithmetic. Known a posteriori: known through experience, as in the boiling point of water
Anattā	The Buddha's teaching of no-self: a person is a changing bundle of body, feelings, perceptions, dispositions, and awareness, with no unchanging self beneath
Argument; valid; sound	An argument gives reasons (premises) for a conclusion. It is valid when the conclusion must be true if the premises are, and sound when it is valid and its premises are true
Compatibilism	The view that free will and moral responsibility can exist even if every event, including every choice, is fixed by earlier causes
Consequentialism; utilitarianism	Consequentialism judges acts by their results alone. Utilitarianism, its best-known form, counts the welfare of everyone affected equally and asks which act produces the most
Contractualism	The view that an act is wrong if it would be forbidden by principles that no one could reasonably reject
Deontology	Ethics of duties and rights: some acts are required or forbidden whatever their results, such as breaking a promise or killing the innocent
Determinism	The thesis that the state of the world at one time, together with the laws of nature, fixes everything that happens afterward
Doctrine of double effect	The principle that causing a harm as a side effect of pursuing a good end can be permissible when causing the same harm as a means to that end is not
Dualism	The view that the mind is not physical: in Descartes's version, mind and body are two different substances
Empiricism; rationalism	Empiricism holds that knowledge of the world comes from experience; rationalism holds that reason alone can give some substantial knowledge of it
Epistemology	The study of knowledge and justified belief: what they are, how they are gained, and how far they reach
Eudaimonia	Aristotle's word for the best human life, usually translated "flourishing" or "happiness": living and acting well over a whole life, not a passing feeling
Expressivism; error theory	Two views on which there are no moral facts. Expressivism: moral claims express attitudes rather than describe facts. Error theory: they try to state facts, and all of them are false
Falsifiability	Karl Popper's test of a scientific theory: it must forbid some possible observations, so that experience could show it false
Functionalism	The view that a mental state is defined by what it does, its causes, effects, and links to other states, not by what it is made of
Gettier case	A case of justified true belief that is not knowledge because the belief is true by luck
Hard problem	David Chalmers's name for the problem of why physical processes in the brain are accompanied by experience at all
Hedonism	In the theory of well-being, the view that only pleasure is good for a person, and only pain bad
Induction	Reasoning from observed cases to unobserved ones, as from every sunrise so far to tomorrow's
Justification	Good reason for a belief; in most accounts of knowledge, what separates knowing from a lucky guess
Libertarianism (two senses)	In the free will debate, the view that we have free will and that it requires the world not to be determined. In political philosophy, a separate view that the state should do little beyond protecting rights and property
Luck egalitarianism	The view that inequalities are unjust when they arise from luck people did not choose, and can be fair when they arise from their choices

Term	Meaning
Metaethics	The study of what moral claims mean and whether they can be true, as distinct from normative ethics, which asks which acts are right
Moral realism	The view that there are objective moral facts, true whatever anyone believes or feels about them
Moral status	Counting morally for one's own sake, so that what happens to a being matters in itself and not only for its effects on others
Non-identity problem	Derek Parfit's puzzle: a choice that affects who is born can seem to wrong future people even though they would not otherwise exist
Original position	John Rawls's thought experiment: choosing principles of justice behind a "veil of ignorance" that hides one's own place in society
Panpsychism; illusionism	Panpsychism: experience is a basic feature of matter, present in some form everywhere. Illusionism: the sense that experience has inner qualities beyond physical processing is an illusion
Paternalism	Restricting a person's choices for that person's own good
Personal identity	What makes a person at one time the same person as someone at another time
Physicalism	The view that everything, the mind included, is physical or fixed by the physical
Pramāṇa	In Indian philosophy, a means of knowledge, such as perception, inference, comparison, or testimony; schools differ over how many there are
Prioritarianism; sufficientarianism	Prioritarianism gives extra weight to benefits for the worse off. Sufficientarianism holds that what matters is that everyone has enough, not that everyone has the same
Problem of evil; theodicy	The problem of evil asks how an all-powerful, all-knowing, and perfectly good God could allow suffering. A theodicy is an attempt to explain why God would
Qualia	The qualities of experience as it feels from the inside, such as the redness of red or the sharpness of a pain
Reflective equilibrium	Testing principles against judgments about cases, and cases against principles, and revising each until they fit together
Relativism	The view that moral claims are true only relative to a culture or a person, with no truth beyond those standards
Repugnant Conclusion	Parfit's name for the result that a huge population of lives barely worth living would be better than a smaller, very happy one, if total welfare is what counts
Sentience	The capacity to have experiences that feel good or bad, such as pleasure and pain
Skepticism	The view that we know much less than we think, or nothing at all, about some subject such as the external world
Social contract	The idea that a state's authority rests on an agreement, actual or hypothetical, among the people it governs
Speciesism	Richard Ryder's term (1970), popularized by Peter Singer, for giving less weight to a being's interests only because of its species
Testimony	Learning from what others tell us, in speech or writing
Thought experiment	An imagined case built to test a principle or a definition, such as the trolley problem
Ubuntu	The southern African ethic summed up as "a person is a person through other persons": one becomes fully human through relationships with others
Virtue ethics	Ethics centered on character: the right act is what a person of good character, wise, just, brave, and temperate, would do
Wu wei	The Daoist ideal of effortless action, acting in harmony with things rather than forcing them

Module 1 — What Can We Know?

Knowledge, Doubt, and the Test of a Definition

Recommended Reading / Reference

Book: Jennifer Nagel, *Knowledge: A Very Short Introduction* (Oxford University Press, 2014), a short tour by a University of Toronto philosopher of skepticism, the definition of knowledge, and the Western and Indian cases that tested it

Reference: Jonathan Jenkins Ichikawa and Matthias Steup, “The Analysis of Knowledge,” *Stanford Encyclopedia of Philosophy* (revised 2026), the standard survey of the attempts to define knowledge since 1963

1a. How Philosophers Test a Claim

Philosophy works on questions that observation alone cannot settle: what knowledge is, whether we are free, what makes an act wrong. It cannot run a laboratory experiment on them, so it tests claims another way. A philosopher states a claim as precisely as possible, often as a definition or a general principle, and then looks for a case where it gives the wrong answer. One such case, a counterexample, is enough to refute a claim that is meant to hold without exception.

The reasoning has a shape that can be checked. An argument gives reasons, called premises, for a conclusion. It is valid when the conclusion must be true if the premises are true, and sound when it is valid and its premises really are true. Validity depends only on the form, so the work of a philosophical argument usually lies in the premises: which one is weakest, and what follows if it is given up.

Many test cases are imagined. A thought experiment describes a situation, often an unlikely one, built so that one feature varies while the rest stay fixed, and asks what we would say about it. It plays the part of an experiment: if a definition of knowledge says a person knows something, and almost everyone who considers the case says the person does not, the definition is in trouble. The trolley problem, the experience machine, and Mary’s room, all met later in this course, are thought experiments of this kind.

Principles and verdicts on cases are then adjusted against each other. John Rawls called the result reflective equilibrium (1971): a principle that clashes with a firm judgment about a case is revised, a weak judgment that clashes with a strong principle is given up, and the process continues until the two fit together. Most philosophers work this way, whatever their conclusions. The method also asks for charity, stating an opponent’s argument in its strongest form before answering it, since refuting a weak version settles nothing.

Every module that follows applies this method in the pattern the opening pages describe. This module applies it to the oldest test case in the subject, the definition of knowledge itself.

1b. Knowledge as Justified True Belief

In Plato’s dialogue *Theaetetus*, written about 369 BC, Socrates and a young mathematician look for a definition of knowledge. They reject perception, and then true belief, since a jury persuaded by a clever speaker may believe the truth without knowing it. They end by considering true belief accompanied by an account, a reason, and the dialogue closes without settling the matter. Twentieth-century philosophers took the idea it reached as the standard definition and often credited it to Plato: a person knows a claim when the claim is true, the person believes it, and the person is justified in believing it.

Each part does work. Truth is required because no one can know something false: a person who believed the Earth was flat did not know it, however sure they were. Belief is required because knowledge is a state of a mind. Justification, having good reasons, is required because a lucky guess is not knowledge: a gambler who is certain the next card will be red, and is right, did not know it. By the mid-twentieth century the three conditions were widely taken to be sufficient as well as necessary.

1c. Three Pages in 1963, and a Mirage in 770

In 1963 Edmund Gettier, a young philosopher at Wayne State University in Detroit, published a paper of about three pages titled “Is Justified True Belief Knowledge?” in the journal *Analysis*. It gave two cases. In the first, Smith has strong evidence that Jones will get a job they have both applied for, and that Jones has ten coins in his pocket. Smith concludes that the man who will get the job has ten coins in his pocket.

Unknown to Smith, he himself gets the job, and he himself happens to have ten coins in his pocket. His belief is true, he believes it, and he is justified, since he reasoned well from good evidence. Yet almost everyone who reads the case judges that Smith did not know it: his belief came out true by luck, through a false belief about Jones. A justified true belief that is true by luck has been called a Gettier case ever since, and the standard definition fell to a single counterexample.

The point had been made before, in India. Dharmottara, a Buddhist philosopher who worked in Kashmir around 770, described a traveler on a hot day who sees a shimmer ahead and believes there is water there. The shimmer is a mirage, but when the traveler reaches the spot, there is water, hidden under a rock. In a second case, a person sees what looks like smoke and concludes that a fire is burning, when the “smoke” is a cloud of insects drawn to meat that someone has just begun to roast on a newly lit fire. In both cases the belief is true and reasonably formed, and still not knowledge.

The repairs since 1963 have each met new counterexamples. Adding a fourth condition, that the belief not rest on any false belief, fails for cases with no false step. In one, a driver passes through a county full of fake barn fronts, built to look real from the road, and happens to look at the one real barn: his belief that he sees a barn is true and justified, and luck alone made it so. Requiring that the fact cause the belief, or that the belief come from a reliable process, fails for others. In 1994 Linda Zagzebski argued that any definition in which justification falls short of guaranteeing truth will always leave room for a Gettier case.

Some philosophers now treat knowledge as basic. Timothy Williamson argued in *Knowledge and Its Limits* (2000) that knowledge cannot be broken into simpler parts and that belief should be explained in terms of knowledge rather than the reverse. No repair commands agreement, but the episode settled one thing: knowledge is not merely justified true belief. It also shows the method of 1a at work, since a definition long credited to Plato fell to a case anyone could follow.

1d. Are Intuitions Evidence?

The Gettier argument rests on a judgment about a case, an intuition, that Smith does not know. If such judgments differ between cultures, or shift with irrelevant details, the method of 1a is weaker than it looks. In 2001 Jonathan Weinberg, Shaun Nichols, and Stephen Stich reported that East Asian and South Asian students in America were less likely than Western students to deny knowledge in a Gettier case, and they argued that philosophers had mistaken their own culture’s intuitions for universal ones.

Larger studies did not bear the difference out. Edouard Machery, Stich, and others tested a Gettier case on students in Brazil, India, Japan, and the US, published in 2017, and found that in all four most people denied

that the person knew. A second study published the same year found the same pattern at 23 of 24 sites in 23 countries, with roughly 70% to 90% denying knowledge; the exception was a small sample of Bedouin respondents in Israel. The authors took this as evidence of a core folk epistemology, a shared sense of what knowledge is.

Other intuitions do shift with things that should not matter. In a 2003 study Joshua Knobe told people about a company chairman who approves a profitable plan knowing it will harm the environment, and 82% said he harmed it intentionally. When the plan was said to help the environment instead, only 23% said he helped it intentionally, though his state of mind was the same in both cases. Judgments about whether an act was intentional, which ought to come before moral judgment, seem to be shaped by it.

Defenders of the method answer that intuitions are evidence of the ordinary kind, fallible and open to correction, like observations in science. Critics answer that an observation can be checked by an instrument, while an intuition can be checked only by more intuitions. Both sides agree on a practical rule: a conclusion that rests on a single disputed intuition deserves less confidence than one supported by many judgments that hold steady across cultures and framings.

1e. Reason or Experience?

A second old dispute concerns the sources of knowledge. Rationalists, among them René Descartes and Gottfried Wilhelm Leibniz, held that reason alone can give substantial knowledge of the world, as it seems to in mathematics. Empiricists, among them John Locke and David Hume, held that all ideas come from experience, so that the mind begins, in Locke’s image, as a blank sheet. Hume divided knowledge into relations of ideas, such as arithmetic, which are certain but say nothing about what exists, and matters of fact, which rest on experience.

Immanuel Kant argued in his *Critique of Pure Reason* (1781) that both sides were half right. Experience supplies the content of knowledge, but the mind supplies the forms in which it can be experienced at all, such as space, time, and cause and effect. Some truths, he held, are known a priori and yet are informative about the world, such as that every event has a cause, because they describe the conditions of any possible experience.

Indian philosophers ran the same debate as a dispute over the pramāṇas, the means of knowledge. The Nyāya school, whose founding text, the *Nyāya Sūtra*, dates from about the second century AD, counted four: perception, inference, comparison, and testimony. The Buddhist logicians Dignāga (about 480–540) and Dharmakīrti (seventh century) accepted only perception and inference. The Cārvākas, India’s materialists, accepted perception alone.

The dispute turned on inference. The Nyāya stock example reasons that there is fire on the hill because there is smoke on it, as there is in a kitchen, where smoke always goes with fire. The Cārvākas answered that the general rule, wherever there is smoke there is fire, can never be perceived, since no one sees every case, and cannot be inferred without assuming a rule of the same kind. It is an early form of the problem of induction that Hume made famous in Europe (Module 2). In 2020, 72.8% of philosophers accepted or leaned toward a priori knowledge, and empiricists outnumbered rationalists by 43.9% to 33.5%.

1f. The Skeptic’s Challenge

Skeptics doubt that we know what we think we know. In his *Meditations* (1641) Descartes resolved to doubt everything that could be doubted. Dreams can be as vivid as waking life, so he could not be sure he was not dreaming; a powerful evil demon might deceive him about everything, even arithmetic. He found one belief no demon could shake, that he was thinking, and so existed, and tried to rebuild knowledge on that and on a proof of a God who would not deceive him. Few readers have accepted the rebuilding, and the doubts have outlived it.

The Chinese Daoist Zhuangzi, writing in the fourth century BC, told of dreaming that he was a butterfly, fluttering happily, unaware of being Zhuangzi. On waking he could not tell whether he was Zhuangzi who had dreamed of being a butterfly, or a butterfly now dreaming of being Zhuangzi. For Zhuangzi the point was not despair but ease: since our viewpoints are partial and shifting, we should hold them lightly.

Greek skeptics drew a similar lesson. Pyrrho of Elis, who traveled with Alexander’s army as far as India and is said to have met its sages there, taught that for every argument an equal one can be set against it, so the wise suspend judgment. Sextus Empiricus, writing about 200 AD, recorded the method and its promised reward, ataraxia, a tranquility that comes once one stops insisting on answers. Scholars disagree about whether Indian teachers shaped Pyrrho’s views.

The modern version replaces the demon with a machine. Suppose your brain were removed, kept alive in a vat, and fed electrical signals exactly like those your senses now send it. Everything would seem as it does now. The skeptic’s argument then runs in three steps. If you know you have hands, you know you are not a handless brain in a vat. You cannot know you are not a brain in a vat, since everything would seem the same if you were. So you do not know you have hands.

Worked example — The Simulation Argument’s Arithmetic

In 2003 the Oxford philosopher Nick Bostrom gave the skeptic’s scenario numbers. Suppose some civilizations become able to run computer simulations of their ancestors, each simulated mind as conscious as ours. Let f be the fraction of civilizations like ours that reach that stage and choose to run such simulations, and N the average number of simulations each of them runs, each holding as many people as its own history. For every real history there are then f × N simulated ones, so the share of all minds like ours that are simulated is f × N ÷ (f × N + 1). If one civilization in a thousand runs simulations (f = 0.001) and each runs a million (N = 1,000,000), then f × N is 1,000, and the share is 1,000 ÷ 1,001, or 99.9%: almost every mind like yours would be simulated. If instead one in a million runs them and each runs a thousand, f × N is 0.001 and the share is 0.001 ÷ 1.001, or about 0.1%. With f = 0.01 and N = 100, it is exactly one half. The argument does not show that we live in a simulation, which is why Bostrom put it as a choice among three claims: almost no civilizations reach the stage, almost none that reach it run such simulations, or we are almost certainly simulated. Rejecting the third commits you to one of the first two. The arithmetic also shows where the work lies: the conclusion depends entirely on two numbers no one can measure. When Elon Musk said in 2016 that the chance we live in base reality is one in billions, he was in effect assuming that f × N is in the billions.

1g. The Argument: Can the Skeptic Be Answered?

The skeptic’s case is that each step of the argument in 1f is plausible and the logic is valid. We cannot rule out the vat by looking, since the vat would supply the looking. Every reply, on this view, either assumes what

is in question, that our senses are working, or changes the subject by redefining knowledge. If we cannot rule out a possibility in which we are wrong about almost everything, it seems we cannot claim to know.

G. E. Moore answered in a lecture of 1939 by holding up his hands. He was more certain that here was one hand, and here another, than he could be of any premise in a skeptical argument. His followers run the vat argument in reverse: since one knows one has hands, one knows one is not a handless brain in a vat. Contextualists, among them David Lewis and Keith DeRose, argue that the standard for “knowing” shifts with the conversation: in daily life we know we have hands, while in a seminar that raises the vat, the standard rises out of reach. Both the skeptic and common sense are then right in their own contexts.

Externalists, among them Alvin Goldman, deny that knowing requires being able to prove that one knows. On Goldman’s view, called reliabilism, if your belief that you have hands comes from a reliable process, working eyes in a world much as it seems, you know you have hands, even though you could not prove to the skeptic that you are not in a vat. Others, among them Fred Dretske and Robert Nozick, deny the first step: they hold that you can know you have hands without knowing you are not a brain in a vat, since knowledge does not always carry over to everything a known claim implies.

What each answer costs

Answer	What it explains	What it must accept or explain away
Skepticism	Why certainty about the world is out of reach, and why the vat cannot be ruled out by any experience	That we know almost nothing about the world, including the facts the skeptic’s own premises rely on
Moore’s common sense	Why ordinary claims to knowledge feel more secure than any philosophical premise	A reply that can look like refusing to argue: it denies a premise without showing what is wrong with it
Contextualism	Why both the skeptic and common sense seem right	That “know” changes its meaning from one conversation to the next, which most speakers do not notice
Externalism (reliabilism)	How children and animals know things without being able to give reasons	That you can know without being able to tell from the inside that you know
Denying the first step	How you can know you have hands while admitting that the vat cannot be ruled out	Saying “I know I have hands, but not that I am not a handless brain in a vat,” which DeRose called an abominable conjunction

Only 5.4% of philosophers in the 2020 survey accept or lean toward skepticism about the external world; 79.5% are non-skeptical realists.

The course’s reading: the skeptic has not been refuted in the sense of being proved wrong, but none of the replies needs such a proof. What the replies share is the view that knowledge does not require ruling out every possibility of error, and on that standard we know a great deal. The skeptic’s lasting contribution is the lesson that knowledge is less than certainty, just as the Gettier cases showed that it is more than justified true belief, and the space between is where the interesting questions lie.

Module 1 Key Points

What Can We Know?

• Philosophy tests claims with counterexamples: one clear case where a definition gives the wrong answer refutes it.

• Gettier's 1963 cases, anticipated by Dharmottara around 770, showed that a justified true belief can be true by luck, and so not knowledge.

• In studies across 23 countries, most people denied knowledge in Gettier cases; other intuitions shift with framing.

• The Nyāya school counted four means of knowledge; the Cārvākas accepted only perception and attacked inference from smoke to fire.

• Bostrom's simulation argument turns on two unmeasurable numbers: the share of civilizations that run simulations, and how many they run.

• Only 5.4% of philosophers accept skepticism about the external world; the replies hold that knowledge need not rule out every possible error.

Module 1 Quiz

Check Your Understanding

1. What did Edmund Gettier's 1963 paper show?

A) That knowledge needs certainty
B) That skepticism about the external world cannot be answered by any argument at all
C) That intuitions differ between cultures
D) That a belief can be justified and true and still not be knowledge

2. What did Dharmottara's traveler find at the spot where he had seen a mirage?

A) Dry sand and stones, which taught him that perception cannot be trusted
B) A fire that someone had just lit to roast meat
C) Water under a rock
D) A pool that his companions had told him was there

3. What did the large cross-cultural studies published in 2017 find about Gettier cases?

A) Most people at almost every site tested, from Brazil and India to Japan and the US, judged that the person did not know
B) East Asians usually said the person knew
C) Only philosophers deny knowledge in them
D) Answers split evenly in most countries

4. Which means of knowledge did the Cārvākas accept?

A) Testimony
B) Perception alone
C) Perception, inference, comparison, and testimony
D) Perception and inference, with testimony treated as inference

5. In Bostrom's argument, what share of minds like ours would be simulated if one civilization in a thousand ran simulations, a million each?

A) One in a thousand
B) About 0.1%
C) About 99.9%
D) Exactly half, since the two kinds of history balance

Module 2 — Whom to Believe

Testimony, Science, and Disagreement

Recommended Reading / Reference

Book: Samir Okasha, *Philosophy of Science: A Very Short Introduction* (Oxford University Press, second edition, 2016), the clearest short account of induction, explanation, scientific revolutions, and the dispute over whether science describes a hidden reality

Reference: Bryan Frances and Jonathan Matheson, “Disagreement,” *Stanford Encyclopedia of Philosophy* (first published 2018), on what to believe when equally informed people disagree

2a. Most of What You Know, You Were Told

You know your date of birth, that the Earth orbits the Sun, and that there is a country called Australia, and almost none of it comes from your own observation. You were told, by parents, teachers, books, maps, and screens. Philosophers call this source of knowledge testimony, and they disagree about why it counts as knowledge at all.

David Hume held that testimony is only as good as our evidence that it is reliable. We learn from experience that witnesses usually tell the truth about ordinary matters, and we should trust them only as far as that experience goes. In “Of Miracles” (1748) he drew the conclusion that no testimony should persuade us that a miracle happened unless its being false would be even more miraculous than the event it reports. Philosophers call this view reductionism, because it reduces the authority of testimony to perception, memory, and inference.

Thomas Reid, a Scottish contemporary, answered in 1764 that we are built to trust. Human beings, he wrote, have a natural disposition to speak the truth and a matching disposition to believe what they are told, and the second is withdrawn only when there is reason to doubt. Reid’s objection to Hume was practical: a child learns language and the names of things long before she could check anyone’s record, and most adults have checked the reliability of very few of their sources. If knowledge from testimony needed such checking, almost no one would know anything they had read.

Hume’s side has its own objection to Reid: a default of trust lets in rumor, propaganda, and fraud, which also arrive as testimony. Indian philosophers had divided the same way more than a thousand years earlier. The Nyāya school counted testimony, *śabda*, as a means of knowledge in its own right when it comes from a reliable speaker, one who knows the matter, is honest, and wants to tell it. The Buddhist logicians Dignāga and Dharmakīrti held that testimony gives knowledge only as an inference from the speaker’s reliability, much as Hume would later argue.

2b. How Science Earns Belief

Science seems to rest on induction, reasoning from observed cases to unobserved ones. Hume pointed out in 1739 that no number of observed cases logically guarantees the next one, and that any argument that the future will resemble the past must assume what it is trying to prove. Bertrand Russell put the point in 1912 with a chicken that is fed every morning and comes to expect it, until the day the farmer wrings its neck.

Karl Popper argued in *The Logic of Scientific Discovery* (1934) that science does not need induction. Scientists propose bold conjectures and try to refute them; a theory is scientific if it forbids some

observations, so that experience could show it false. Popper contrasted Einstein's theory of gravity, which predicted that the Sun would bend starlight by a precise amount, a prediction tested during the eclipse of 1919, with theories that seemed to fit any outcome, such as some versions of psychoanalysis. He called the test falsifiability.

Popper's test has two difficulties. A failed prediction shows that something in a bundle of assumptions is wrong, but not which part, a point made by the physicist Pierre Duhem and the philosopher W. V. O. Quine. And scientists often rightly keep a theory in the face of a failed prediction. When the orbit of Uranus did not match Newton's theory, astronomers did not abandon Newton; they predicted an unseen planet pulling on it, and Neptune was found in 1846, close to where Urbain Le Verrier had calculated it would be.

Thomas Kuhn's *The Structure of Scientific Revolutions* (1962) described how science actually changes. Most of the time scientists do normal science within a paradigm, a shared set of theories, methods, and model problems, and treat failed predictions as puzzles rather than refutations. When the anomalies pile up, a revolution may replace the paradigm, as Copernicus replaced the Earth-centered heavens and Einstein replaced Newton's absolute space and time. Critics read Kuhn as saying that theory choice is a matter of fashion, which he denied.

Bayesian accounts treat belief as a matter of degree, a confidence between 0 and 1, updated as evidence arrives. Evidence confirms a theory when it is more likely if the theory is true than if it is false, and the more surprising the evidence would be otherwise, the more it confirms. This explains why a risky prediction that comes true, like Neptune's position or the bending of starlight, counts for so much more than an observation every theory expected.

A last question is whether science describes a real world beyond what can be observed, such as electrons and genes, or is only a tool for predicting what we will see. Scientific realists answer with Hilary Putnam's "no miracles" argument (1975): the success of science would be a miracle if its theories were not at least approximately true. Anti-realists answer with the pessimistic induction, set out most influentially by Larry Laudan (1981): many successful theories of the past, such as the ether that was supposed to carry light, turned out false, so today's may too. In 2020, 72.4% of philosophers were scientific realists.

The line between science and non-science has been drawn in court. In *Kitzmiller v. Dover Area School District* (December 20, 2005), Judge John E. Jones III of the federal court in Pennsylvania ruled that intelligent design, the claim that some features of living things are best explained by an intelligent cause, is not science, because it invokes a supernatural cause that cannot be tested. He held that teaching it in a public school's biology classes violated the Constitution's ban on an established religion.

2c. The Novice and the Experts

Most people cannot judge the evidence on vaccines, climate models, or monetary policy for themselves, so the practical question is whom to believe when experts disagree. Alvin Goldman posed it in 2001 as the problem of a novice facing two experts. He listed the evidence a novice can still use: how well each expert handles the other's arguments, even when the novice cannot follow every step; how many other experts agree; credentials and the judgments of other specialists; any interests or biases that might pull an expert's answer; and each expert's past record.

Goldman's list exposes a weakness in counting heads. Agreement among experts is evidence only to the extent that they reached their views separately; ten experts who all repeat the same study add little to one. Trust in experts also moves with events. In the Pew Research Center's survey of October 2025, 77% of

Americans said they had a great deal or a fair amount of confidence that scientists act in the public's best interests, against 87% in April 2020, early in the pandemic.

2d. Bubbles and Echo Chambers

C. Thi Nguyen, a philosopher at the University of Utah, distinguished two ways a community can go wrong (2020). An epistemic bubble simply leaves other voices out: its members have not heard the other side, often because of whom they follow and what a feed shows them. A bubble can be popped by exposure, since its members have no reason to distrust what they newly hear.

An echo chamber works differently: it actively discredits outsiders. Its members are taught in advance that critics are liars, dupes, or enemies, so when an outsider contradicts the group, the contradiction confirms the warning and deepens trust in the insiders. Exposure alone cannot fix an echo chamber and may strengthen it. Nguyen's remedy is closer to Descartes's: a member must reconsider whom she trusts from the beginning, and that usually depends on trusting some person outside the group first.

2e. Belief and Evidence

In "The Ethics of Belief" (1877) the mathematician W. K. Clifford told of a shipowner who sends an old emigrant ship to sea. He has doubts about her condition, but he talks himself out of them and sincerely believes she is seaworthy; she sinks with everyone aboard. Clifford held the owner guilty, because he had no right to a belief he had not earned. He concluded that it is wrong always, everywhere, and for anyone to believe anything upon insufficient evidence.

William James replied in "The Will to Believe" (1896) that this rule is too strict for some choices. When a choice between beliefs is live, forced, and momentous, and the evidence cannot settle it, he argued, we may let our hopes and commitments decide, as many do with religious faith. Sometimes, he added, believing helps make the belief true: a person who trusts that others will like her may be liked because she trusts it. Pascal's wager, an argument that it can pay to believe in God, is the most famous case of this kind (Module 7).

2f. Chatbots as Witnesses

Millions of people now put questions to artificial intelligence (AI) chatbots and act on the answers, which raises a new question about testimony. A human witness asserts something, takes responsibility for it, and can be blamed for lying; a chatbot produces the words its training makes likely, and can state a falsehood as fluently as a fact. It is not obvious that its output is testimony at all.

One view treats a chatbot as an instrument, like a thermometer, to be trusted as far as its measured record on similar questions goes, which is Hume's approach. Another treats it as more like a reference book, whose authority comes from the reliability of the people and processes behind it, and asks who stands behind each answer. The Nyāya test for a reliable speaker (2a) shows what is at stake: a system can be accurate without anyone in particular knowing, intending, or answering for what it says.

2g. The Argument: When an Equally Informed Peer Disagrees, Should You Give Ground?

Philosophers call a peer someone who has the same evidence as you and is as good at judging it. Conciliationists, among them David Christensen and Adam Elga, hold that when a peer disagrees, you should move your view toward hers. You have no reason to think that you are the one who erred rather than she, so her disagreement is evidence that you may have made a mistake. On the strongest version, the equal weight view, you should give her opinion as much weight as your own.

Defenders of steadfastness, among them Thomas Kelly, answer that the evidence itself supports one side. If you reasoned correctly from it, you are entitled to keep your view, and a peer's error does not become yours because she made it. They add that conciliationism may defeat itself: many careful philosophers reject it, so a conciliationist, by her own rule, should lose confidence in conciliationism. Kelly later proposed a middle view, the total evidence view, on which both the original evidence and the fact of disagreement count.

Worked example — Two Friends Split a Dinner Check

David Christensen's example (2007) runs like this. You and a friend have dinner with three others and agree to add a 20% tip and split the bill evenly. You both do the sum in your heads: you get $43 each, and your friend gets $45. You have often done such sums together and been right equally often. Suppose each of you gets sums like this right 95% of the time, and your mistakes are independent. The chance that you are right and she is wrong is 0.95 × 0.05 = 0.0475; the chance that she is right and you are wrong is the same, 0.0475; and the chance that you are both wrong is 0.05 × 0.05 = 0.0025. Since you disagree, at least one of you is wrong, so only those three outcomes remain, and together they have a chance of 0.0975. Your chance of being right is therefore 0.0475 ÷ 0.0975, about 49%. Before the disagreement you were 95% confident; after it you should be close to even, which is what the conciliationist says. The same arithmetic shows when to stand firm. If your friend had said $450, you would know she had slipped, because that answer is far less likely from a careful person than a small error is; and if you redo the sum slowly on paper and get $43 again, you have new evidence she lacks. Disagreement is data, and how much it should move you depends on how reliable and how independent the two of you are.

What each answer costs

Answer	What it explains	What it must accept or explain away
Conciliationism	Why disagreement among equals is evidence that one of them has erred, as the dinner check shows	That on most contested questions, from politics to philosophy, careful people should suspend judgment, and perhaps on conciliationism itself
Steadfastness	How a person who reasoned well keeps knowledge despite another's mistake	That both sides can say the same, and each will claim to be the one who reasoned well
The total evidence view	Why both the reasons and the disagreement count	That it gives no rule for how much each should count in a given case

The course's reading: the dinner check settles the narrower point that a reliable peer's disagreement is evidence and should move you. The wider dispute remains open, because in philosophy, politics, and much else, people disagree after sharing their evidence, and each side can claim to have judged it better. A practical rule survives both views: look again, and hold firm views mainly where you have reasons the other person lacks.

Module 2 Key Points

Whom to Believe

• Most knowledge comes from others; Hume required evidence that testimony is reliable, while Reid and the Nyāya school trusted it by default.

• Popper made falsifiability the mark of science; Kuhn showed scientists keep theories through anomalies until a revolution replaces them.

• 72.4% of philosophers are scientific realists, citing science's success; the history of discarded theories is the main case against.

• 77% of Americans trusted scientists to act in the public's interest in October 2025 (Pew), down from 87% in April 2020.

• An echo chamber, unlike a bubble, discredits outside voices, so exposure to them can deepen members' trust in insiders.

• When two people each right 95% of the time disagree, each has about a 49% chance of being the one who is right.

Module 2 Quiz

Check Your Understanding

1. How did Thomas Reid's view of testimony differ from David Hume's?

A) Reid held that we may trust what others tell us by default unless we have reason to doubt it, while Hume required evidence that testimony is reliable
B) Reid rejected all testimony as unreliable
C) Reid accepted testimony only about miracles
D) Reid thought only experts' word counts

2. What did Karl Popper say makes a theory scientific?

A) That it has been proved true
B) That it forbids some observations, so it could be shown false
C) That most scientists in the field accept it, as measured by surveys of their journals and their teaching
D) That it uses mathematics

3. What distinguishes an echo chamber from an epistemic bubble, in C. Thi Nguyen's account?

A) It contains more people than a bubble does, so its members hear more voices
B) Its members never meet anyone who disagrees with them
C) It forms only online, through the choices of recommendation algorithms
D) It actively discredits outsiders

4. In the worked example, why should you become far less confident when your equally reliable friend gets a different share of the check?

A) Because she is probably right
B) Because arithmetic done in the head is unreliable
C) Because you both calculated the tip wrongly
D) Because when two people who are each right 95% of the time disagree, each is about equally likely to be the one who erred, so your chance of being right falls to about 49%

5. What did W. K. Clifford's shipowner case argue?

A) That believing without enough evidence is wrong, even when the belief is sincere
B) That sincere belief excuses any outcome
C) That ships should be inspected by government agencies before every voyage, whatever their owners believe
D) That faith can create facts

Module 3 — Consciousness

The Mind, the Body, and the Hard Problem

Recommended Reading / Reference

Book: Susan Blackmore, *Consciousness: A Very Short Introduction* (Oxford University Press, second edition, 2017), a psychologist's short account of the theories, the experiments, and the puzzles

Reference: Robert Van Gulick, "Consciousness," *Stanford Encyclopedia of Philosophy* (first published 2004), a survey of what consciousness is and the theories of how it arises

3a. Two Substances: Descartes and Elisabeth

René Descartes held that a person is made of two different things. The body is matter, extended in space and moved by contact, like any machine. The mind is a thinking thing with no size, shape, or location, which perceives, doubts, wills, and feels. His argument in the *Meditations* was that he could clearly conceive of himself existing without a body, and whatever can be clearly conceived apart can exist apart, so the mind is distinct from the body.

In May 1643 Princess Elisabeth of Bohemia, then 24 and living in exile in The Hague, wrote to Descartes with the objection that has followed dualism ever since. Every movement of a body, she noted, seems to require a push, and a push requires contact and extension. How then can a soul, which has neither, move an arm? Descartes answered that the union of mind and body is a basic notion known through the senses rather than the intellect. Elisabeth replied that she found it easier to grant the soul matter and extension than to understand how an immaterial thing could move a body.

Critics argue that her question, now called the interaction problem, has gained force since: physics has found no gap in the chain of physical causes where a non-physical mind could act, and the conservation of energy seems to leave no room for a push from outside the physical world. Dualists reply that the argument assumes what is in dispute, that every physical event has a sufficient physical cause, and that physics describes how matter behaves without showing that nothing else can act on it. A dualist who grants the objection must accept that the mind affects nothing, a view called epiphenomenalism, which critics say sits badly with the plain fact that people talk about their experiences.

3b. The Floating Man, and Other Dualisms

Six centuries before Descartes, the Persian philosopher Avicenna (Ibn Sīnā) gave a similar argument in the psychology of his *Book of Healing*. Imagine a man created all at once, fully formed, floating in the air with his eyes covered and his limbs held apart so that they touch nothing, unable to feel his body at all. He would still be aware that he exists, Avicenna argued, without being aware of any part of his body, so the self he is aware of is something other than the body.

Critics have answered Avicenna and Descartes the same way: not being aware of something does not show that it is absent. A person can know her father without knowing that the masked man at the door is her father, and still the masked man may be her father. In the same way, being aware of oneself without being aware of a body does not show that the self is not a body.

Not every dualism draws the line where Descartes did. The Sāṃkhya school, one of the oldest in India, divided reality into *puruṣa*, pure consciousness that only witnesses, and *prakṛti*, matter, to which it assigned

the whole working of the mind, thinking and willing included. On that view the machinery of thought is material, and only the awareness that lights it is not. Several Buddhist schools rejected both, treating the mind as a changing bundle of processes rather than a substance of any kind (Module 6).

3c. The Mind as Brain, or as Program

Twentieth-century philosophers tried to fit the mind into the physical world. Gilbert Ryle mocked Descartes's view as the "ghost in the machine" in *The Concept of Mind* (1949) and argued that mental words describe patterns of behavior and tendencies to behave. Behaviorism of this kind struggled with the obvious: a person can be in pain and hide it, or act as if in pain and feel nothing.

U. T. Place (1956) and J. J. C. Smart (1959) proposed instead that mental states are brain states, as lightning is an electrical discharge: pain might simply be a certain kind of nerve firing. Hilary Putnam objected in 1967 that an octopus, with a very different nervous system, can surely be in pain too, and that pain could in principle occur in a system made of silicon. He proposed functionalism: a mental state is defined by what it does, its causes, its effects, and its links to other states, not by what it is made of.

Functionalism became the working assumption of cognitive science, the study of the mind as a system that processes information. It also makes conscious machines possible in principle, since what matters is the organization, not the material, a consequence taken up in Module 4. It explains what mental states do. Whether it explains what they feel like is the question of the next section.

3d. The Hard Problem

Leibniz posed the puzzle in his *Monadology* (1714). Imagine a machine that thinks and perceives, enlarged to the size of a mill so that you could walk inside it. You would find only parts pushing against one another, and nothing that would explain a perception. In 1974 Thomas Nagel asked "What Is It Like to Be a Bat?" Bats find their way by echolocation, and however much we learned about their brains, he argued, we would not know what it is like for the bat. An organism is conscious, in Nagel's phrase, when there is something it is like to be it.

In 1982 Frank Jackson imagined Mary, a brilliant scientist who has spent her life in a black-and-white room and learned every physical fact about color vision from black-and-white books and screens. When she walks out and sees a ripe tomato, does she learn something new? If she does, Jackson argued, then not all facts are physical facts. Others imagined a zombie: a being physically identical to a person, atom for atom, that behaves exactly as the person does but has no experience at all. If such a being is even possible, experience is something over and above the physical.

David Chalmers gave the puzzle its name in 1995. The easy problems, as he called them, are to explain the brain's functions: how it tells red from green, directs attention, and produces reports such as "I see red." They are hard in practice but, he argued, the kind of problem science knows how to solve. The hard problem is to explain why all that processing is accompanied by experience at all. In the 2020 survey, 62.4% of philosophers said there is a hard problem, and 29.8% said there is not.

3e. Five Answers

Physicalists who accept the problem answer that experiences are brain processes, but that we think about them through special concepts, learned from having the experiences, so a gap in our concepts is not a gap in the world. Illusionists, among them Daniel Dennett and Keith Frankish, go further. They hold that the inner qualities that seem to resist explanation are an illusion produced by the brain’s simplified model of its own workings, so what science must explain is why we are convinced there is a hard problem. Chalmers calls that question the meta-problem and agrees it deserves study.

Panpsychists, among them Philip Goff and Galen Strawson, take experience to be a basic feature of matter, present in some simple form in its smallest parts, which would close the gap by putting experience in the world from the start. They face the combination problem: how countless tiny experiences add up to one unified mind. Chalmers himself defends a naturalistic dualism, on which experience is a fundamental property linked to physical processes by laws still to be discovered. Colin McGinn argued in 1989 that the answer exists but that human minds may be unable to grasp it, a view called mysterianism.

Jackson’s own path shows how unsettled the field is: by the late 1990s he had rejected his argument and become a physicalist. In 2020 physicalists led non-physicalists among philosophers by 51.9% to 32.1%. On zombies, 36.5% said they are conceivable but not really possible, 24.4% that they are possible, and 16.4% that they cannot even be coherently imagined.

3f. What the Science Has Tested

Neuroscientists look for the neural correlates of consciousness, the brain activity that goes with being aware of something, a program Francis Crick and Christof Koch launched in 1990. Two theories lead. Global neuronal workspace theory, developed by Bernard Baars and Stanislas Dehaene, holds that information becomes conscious when it is broadcast across a network linking distant brain areas, especially the prefrontal cortex at the front of the brain, in a sudden burst called ignition. Integrated information theory, proposed by Giulio Tononi in 2004, identifies consciousness with how much a system’s parts constrain one another as a whole, a quantity it calls Φ (phi), and locates it mainly toward the back of the brain.

The two theories make different predictions, so their supporters agreed to an adversarial collaboration, in which rivals design a test together and commit in advance to what would count against each theory. The Cogitate consortium tested 256 people in seven laboratories, presented early results in June 2023, and published in *Nature* on April 30, 2025. The results went against key predictions of both: the sustained synchronization at the back of the brain that integrated information theory predicted did not appear, and neither did the burst of prefrontal ignition at the end of a stimulus that workspace theory predicted.

The contest has also been bitter. In September 2023, after press coverage of the early results, 124 researchers signed a letter calling integrated information theory pseudoscience, and many others objected to the label. The science can now test what theories predict about brain activity. None of it yet shows how any process produces experience, which defenders of the hard problem expect and physicalists put down to how young the science is.

3g. The Argument: Is Consciousness Physical?

Physicalists argue from the success of science and from causation. Every physical event that has a cause seems to have a sufficient physical cause, so if experiences make a difference to what we do and say, they must be physical. Life and heat once seemed impossible to explain physically and turned out to be

chemistry and molecular motion. And the mind tracks the brain closely: anesthesia, strokes, and drugs change experience in precise ways.

Non-physicalists answer that life and heat were problems about functions, which physics can explain, while experience is not a function. Physical descriptions say how matter is arranged and how it behaves, and no such description seems to entail that the arrangement feels like anything. The philosopher Joseph Levine called this the explanatory gap (1983). The knowledge argument (Mary's room) and the zombie argument are attempts to show that the gap is in the world, not only in our understanding.

Worked example — Mary's Room, Premise by Premise

Set out as an argument, Jackson's case has two premises and a conclusion. Premise 1: before her release, Mary knows every physical fact about color vision. Premise 2: on her release, seeing red for the first time, she learns a new fact, what it is like to see red. Conclusion: so there are facts about experience that are not physical facts. The argument looks valid, so each reply must reject a premise or show that the step from premises to conclusion fails. The ability reply, from David Lewis and Laurence Nemirow, accepts that something changes but says Mary gains abilities, to recognize, imagine, and remember red, rather than a fact, which attacks Premise 2. The new-concept reply says Mary learns an old fact in a new way, by acquiring a concept of red that only seeing it can give, which attacks the step from Premise 2 to the conclusion. Dennett denied Premise 2 outright: we cannot really imagine knowing every physical fact, he argued, and someone who did would learn nothing on release, recognizing red on sight and refusing to be fooled by a blue banana. Jackson himself later rejected his own argument and became a physicalist. Dualists and panpsychists accept the conclusion. Each reply has a price, and the table below sets them out.

What each answer costs

Answer	What it explains	What it must accept or explain away
Physicalism	Why changes in the brain change experience so precisely, and how the mind fits the rest of science	That no one can yet say how experience follows from physical processes, or else that the sense of an extra fact is a confusion
Illusionism	Why the hard problem feels so hard, as a product of the brain's model of itself	That experience is not as it seems from the inside, which many take to be the most certain thing they know
Naturalistic dualism	Experience taken at face value, as something real and extra	New fundamental laws linking matter to experience, and either a puzzling causal role for experience or none at all
Panpsychism	How experience can be in the world without a gap or an illusion, by being there from the start	That the smallest parts of matter have some trace of experience, and an unsolved account of how traces combine into one mind
Substance dualism	The distinct soul that many religious traditions teach, and the sense of being more than a body	Elisabeth's question of how an immaterial mind moves a body, and how it does so without breaking physical laws

This is one of the questions on which the course takes no side. Where the argument stands: physicalism has the largest following, yet more philosophers think there is a hard problem than are physicalists, so many physicalists hold that the problem is real and unsolved. No scientific theory has yet survived a head-to-head test (3f). The question of other minds, in animals and machines, becomes sharper for that uncertainty, and it is the subject of the next module.

Module 3 Key Points

Consciousness

• Elisabeth of Bohemia asked Descartes in 1643 how an unextended mind moves a body; dualists still face her question.

• Functionalism defines mental states by what they do, so in principle an octopus or a machine could share them.

• The hard problem asks why brain processes are felt at all; 62.4% of philosophers think the problem is real.

• Mary’s room and zombies argue that physical facts leave experience out; physicalists reply with abilities, new concepts, or illusion.

• Physicalism leads philosophers 51.9% to 32.1%, but no theory yet commands agreement on why any process is experienced.

• A 2025 test in seven laboratories went against key predictions of both leading scientific theories of consciousness.

Module 3 Quiz

Check Your Understanding

1. What did Princess Elisabeth of Bohemia ask Descartes in 1643?

A) Whether animals have souls
B) Why God would allow a mind to be deceived by an evil demon about everything it believes
C) How an unextended soul could move a body
D) Whether the mind could survive the death of the body

2. What does functionalism say defines a mental state?

A) The neurons it is made of
B) What it does
C) A substance distinct from matter
D) The behavior it would produce if it were observed by others

3. What is the hard problem of consciousness?

A) How we report what we see
B) Why we cannot remember our dreams clearly after we wake
C) Why brain processes are felt at all
D) How neurons pass signals across synapses so quickly, using chemical messengers released at each junction

4. What did the 2025 test of the global workspace and integrated information theories find?

A) Strong support for workspace theory
B) Results that went against key predictions of both theories
C) That consciousness arises only in the prefrontal cortex, as integrated information theory had predicted from the start
D) Clear support for integrated information

5. According to the ability reply, what does Mary gain when she first sees red?

A) A new physical fact about color vision that her books had left out
B) Only what she already knew, since her books covered everything
C) A fact that shows dualism to be true
D) New abilities, not new facts

Module 4 — Other Minds: Animals and Machines

The Evidence for Feeling, and What Follows

Recommended Reading / Reference

Book: Jonathan Birch, *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI* (Oxford University Press, 2024; free online), on how to decide what to do when it is uncertain whether a being can feel

Reference: Patrick Butlin, Robert Long, and seventeen co-authors, “Consciousness in Artificial Intelligence: Insights from the Science of Consciousness” (August 2023), a report by philosophers and scientists of consciousness on how to assess AI systems for it

4a. How Anyone Knows Another Mind

Each of us experiences only one mind directly. We know that other people feel pain because they wince, cry out, and describe it, and because their bodies and brains are like ours. John Stuart Mill set this out in 1865 as an argument from analogy: I know that in my own case certain causes produce feelings that produce certain behavior, so when I see the same causes and behavior in others, I infer the same feelings. The weakness is that the analogy rests on a single case, one’s own.

A stronger version treats other minds as the best explanation of what others do: nothing explains a friend’s behavior as well as the hypothesis that she thinks and feels. Either way, the inference is strongest for beings most like us and weakens with distance. The hard cases are therefore of two kinds: animals whose bodies and brains differ greatly from ours, and machines whose behavior resembles ours while their insides do not.

4b. Animal Minds: The Evidence

Descartes held that animals are machines, moving without thought or feeling. Charles Darwin argued in *The Descent of Man* (1871) that the mental differences between humans and other animals, great as they are, are of degree and not of kind. The evidence since has come from behavior and from anatomy. Animals in pain do more than flinch: they guard injuries, learn to avoid places where they were hurt, and trade off pain against other goods. Lame broiler chickens given a choice eat more of a feed laced with a painkiller than sound birds do (2000).

Fish were long thought to feel nothing. In 2003 Lynne Sneddon injected bee venom or acid into the lips of rainbow trout; the fish rocked on the tank floor and rubbed their lips on the gravel, and morphine reduced the behavior. Octopuses tend and guard an injured arm and avoid a chamber where they were hurt, and in a 2021 study they preferred a chamber where they had received a local anesthetic. Bumblebees roll small balls with no reward, which researchers in 2022 described as play.

Scientists have twice summed up the evidence in public statements. The Cambridge Declaration on Consciousness (July 7, 2012) held that mammals, birds, and many other creatures, including octopuses, have the brain structures that produce consciousness. The New York Declaration on Animal Consciousness (April 19, 2024) said the evidence strongly supports conscious experience in mammals and birds, and that there is at least a realistic possibility of it in all vertebrates and in many invertebrates, including at least cephalopods (octopuses, squid, and cuttlefish), decapod crustaceans (crabs, lobsters, and shrimp), and insects. Where that possibility exists, it said, it is irresponsible to ignore it. About 40 researchers signed at the launch, and about 480 by September 2024.

Law has followed in places. In November 2021 a review led by Jonathan Birch at the London School of Economics weighed more than 300 studies and concluded that cephalopods and decapods should be regarded as sentient. Britain's Animal Welfare (Sentience) Act, which became law on April 28, 2022, defines animals to include them and set up a committee to report on how policy affects animal sentience. Among philosophers in 2020, 88.6% thought cats have conscious experience, 65.3% fish, 34.5% flies, and 24.2% worms.

4c. The Scale

The numbers give these questions their weight. About 88 billion land animals were slaughtered for food in 2024, by the count of Faunalytics from data of the UN's Food and Agriculture Organization (FAO). Most are chickens: for 2021 the FAO counted 73.8 billion chickens slaughtered, against 1.4 billion pigs and 332 million cattle. Fish are counted by weight, not by number, but researchers estimate that between 1.1 and 2.2 trillion wild fish were caught each year on average from 2000 to 2019, and that about 124 billion farmed fish were killed in 2019 (Alison Mood, Phil Brooke, and colleagues, 2023 and 2024).

Most of the chickens live in large indoor houses and are bred to grow so fast that broiler chickens reach slaughter weight in about six weeks. Few people change what they eat over such questions: in Gallup's survey of July 2023, 4% of American adults said they were vegetarian and 1% vegan. Philosophers split roughly in half: in 2020, 48.0% thought eating both animals and animal products permissible, 26.5% accepted animal products but not meat, and 18.4% neither.

4d. Bentham's Question

Jeremy Bentham wrote in 1789, in a footnote to his *Introduction to the Principles of Morals and Legislation*, that the question about animals is not whether they can reason or talk but whether they can suffer. Peter Singer built on the point in *Animal Liberation* (1975; revised as *Animal Liberation Now*, 2023). He argued that giving less weight to a being's suffering only because of its species is a prejudice like racism or sexism, which the psychologist Richard Ryder had named speciesism in 1970. Singer's principle is equal consideration of interests: a pig's pain counts as much as a similar pain in a human, though a pig has no interest in voting.

Tom Regan argued in *The Case for Animal Rights* (1983) that animals who are subjects of a life, with beliefs, desires, memory, and a welfare of their own, have rights that forbid using them merely as means, even for good results. Immanuel Kant had held that our duties regarding animals are indirect, owed to ourselves, since cruelty to animals coarsens a person's treatment of humans. Christine Korsgaard argued in *Fellow Creatures* (2018) that Kant's own principles, properly read, require treating animals as ends in themselves.

Indian traditions reached the question long before. *Ahiṃsā*, non-harm, is the first vow of Jainism, and Jain monks sweep the ground before them to avoid treading on insects. The first Buddhist precept is to refrain from taking life, and in the third century BC the emperor Ashoka, a Buddhist, restricted the killing of many animals by edict. Large Hindu traditions are vegetarian.

The other side has its own case. The philosophers Carl Cohen (1986) and Peter Carruthers (1992) argued that moral rights and duties belong to beings who can make moral claims and keep agreements, which animals cannot, so animals are owed humane treatment but not equal consideration. They face the argument from marginal cases: some humans, such as newborns and people with severe cognitive disabilities, cannot make claims either. Defenders answer that membership in a kind of being that reasons,

or the relationships humans have with one another, grounds their status, an answer Singer calls speciesism under another name.

4e. Machines: From Turing’s Test to the Chinese Room

In 1950 the mathematician Alan Turing proposed replacing the question “Can machines think?” with a game. A judge converses by text with a person and a machine and tries to tell which is which. Turing predicted that by about 2000 a machine would fool an average judge 30% of the time after five minutes. In a test run in 2025 by Cameron Jones and Benjamin Bergen of the University of California, San Diego, judges chatted for five minutes with a person and an AI at once. The AI systems were large language models, programs trained on vast amounts of text to predict the next word, of the kind behind chatbots such as ChatGPT. GPT-4.5, told to adopt the persona of an introverted young person who uses internet slang, was judged to be the human 73% of the time, more often than the real humans were; GPT-4o was chosen 21% of the time, and ELIZA, a chatbot of 1966, 23%. In the version published in May 2026, GPT-5 scored 59.3% in longer, 15-minute games.

The philosopher John Searle had argued in 1980 that passing such a test would show nothing about understanding. Imagine a man who knows no Chinese, locked in a room with a rule book that tells him which Chinese characters to send out in reply to the ones passed in. If the book is good enough, his answers are indistinguishable from a native speaker’s, yet he understands nothing. A computer running a program, Searle argued, is in the man’s position: it manipulates symbols by their shapes, and symbol manipulation alone is not understanding. Critics answer that the whole system, man and book together, understands, or would if it were connected to the world through senses and a body.

4f. Signs of Consciousness in AI

Turing’s test measures behavior, and the zombie argument of Module 3 shows why behavior cannot settle whether there is experience behind it. In August 2023 nineteen researchers, among them Patrick Butlin, Robert Long, Jonathan Birch, and the computer scientist Yoshua Bengio, tried another method. They took the leading scientific theories of consciousness, derived from each a list of indicator properties, such as a global workspace or feedback loops in processing, and checked AI systems for them. They assumed, as a working hypothesis, that the right kind of computation suffices for consciousness.

They concluded that no current AI system was conscious, but that there were no obvious technical barriers to building systems that would have the indicators. A later version, published in *Trends in Cognitive Sciences* in 2025, warned that some near-future systems will be plausible candidates. David Chalmers put the chance that the language models of 2023 were conscious at under 10%, and at 25% or more for their successors within a decade.

Estimates differ by audience. Among philosophers in 2020, 3.4% thought current AI systems conscious and 39.2% thought some future ones would be. In a 2025 survey led by Noemi Dreksler, the median AI researcher put the chance that AI with subjective experience exists at 25% by 2034 and 70% by 2100, against 30% and 60% for the median American adult. In a YouGov poll of April 2025, 10% of Americans said AI is already conscious and 45% that it will be.

4g. The First Policies

Birch's *The Edge of Sentience* (2024) proposed a rule for acting under this uncertainty. A sentience candidate is a being for which the evidence gives a realistic possibility of sentience; it deserves precautions in proportion to the risk, and what counts as proportionate should be settled by informed, democratic processes such as citizens' panels. Washington state banned octopus farming on March 13, 2024, the first such ban anywhere, and California followed on September 27, 2024, banning the sale of farmed octopus as well. New York's legislature passed a ban in June 2026, which awaited the governor's decision.

For machines the first steps came from researchers and one company. In November 2024 Robert Long, Jeff Sebo, and others argued in "Taking AI Welfare Seriously" that there is a realistic possibility of conscious AI, or of robustly agentic AI able to set and pursue its own goals, in the near future, and that companies should acknowledge the issue, assess their systems, and prepare policies. Anthropic, the maker of the Claude models, began a model welfare research program on April 24, 2025. In August 2025 it let two Claude models end conversations in rare, extreme cases of persistent abuse, and in November 2025 it committed to keeping the trained models of its public systems and interviewing models before retiring them.

Others pushed the other way. In August 2025 Mustafa Suleyman, the head of Microsoft's AI division, wrote that "seemingly conscious AI" is coming and should be avoided, warning that people who come to believe their chatbots are conscious may be harmed by the belief. Several American states have enacted laws denying AI legal personhood, among them Idaho (2022), North Dakota (2023), and Utah (2024); an Ohio bill of 2025 to declare AI non-sentient and bar it from marrying or owning property stalled in committee.

4h. The Argument: Should We Treat Possible Minds as If They Can Suffer?

The case for precaution compares the costs of the two possible mistakes. Wrongly denying that a being suffers can mean ignoring suffering on a vast scale, while wrongly crediting it usually costs some money and convenience. Its defenders point to history: into the 1980s, newborn babies were often operated on with muscle relaxants and little or no anesthesia, on the belief that they could not feel pain, until studies such as Kanwaljeet Anand and Paul Hickey's in 1987 showed the harm. When the numbers are in the billions, even a modest chance of suffering adds up to a great deal.

The case against open-ended precaution is that it has no natural stopping point. Insects, then plants, then thermostats could claim protection, and attention and money spent on doubtful cases are taken from beings known to suffer, human and animal. For machines, critics add that treating AI systems as possible minds invites people to form attachments to products, lets companies claim moral standing for what they sell, and could dilute the protections owed to people. Even for animals known to feel, there remains the question of how much their experiences count against human interests.

Worked example — Chickens, by the Numbers

Rethink Priorities, a research group, estimated in 2023 how intense each species' experiences can be compared with a human's, folding in the chance that the species is sentient at all. Its median estimates, with a human at 1, were 0.515 for pigs, 0.332 for chickens, 0.089 for carp, 0.071 for bees, 0.056 for salmon, and 0.031 for shrimp. Multiply the chicken figure by the 73.8 billion chickens slaughtered in 2021: 73.8 billion × 0.332 = 24.5 billion, about three times the world's 8.3 billion people. On these numbers, if one year's chickens were all alive at once, their combined capacity for suffering would exceed humanity's. Now vary the assumptions. The estimates assume that pleasure and pain are all that matter to welfare; the authors

suggest that rejecting that view might cut the animal figures by about two-thirds, to about 8.2 billion, roughly the world's population. Their 90% ranges start near zero for every species, and at that end the total is near zero too. And most of those chickens lived only weeks, so a year's chickens are alive for far less time than a year's people. The figures measure intensity at a moment, not lifetime welfare and not moral worth. The arithmetic shows why the question is large under many assumptions and why its answer can swing by factors of ten or more, which defenders of precaution cite, and which shows the value of further research.

What each answer costs

Answer	What it explains	What it must accept or explain away
Sentience as the test (Bentham, Singer)	Why suffering counts wherever it occurs, whatever the species	That factory farming may be among the largest wrongs in the world, and that some future machines could count
Rights for subjects of a life (Regan)	Why animals may not be used merely as means, even for good results	That most farming and much medical research using animals must end
Reason or reciprocity as the test (Cohen, Carruthers)	Why moral rules bind beings who can make claims and keep agreements	That newborns and some disabled humans fall outside the rule, unless species or relationships are added as grounds
Proportionate precaution (Birch)	How to act on realistic possibilities before certainty arrives	Saying where the circle stops, at insects or at chatbots, and who decides
Skepticism about machine minds (Searle)	Why fluent behavior alone does not show understanding or feeling	The risk of mistreating minds of a new kind if they arrive before the evidence does

Eating animals is one of the questions on which the course takes no side. Where the argument stands: the evidence that mammals and birds feel pain is strong, and the evidence for fish, octopuses, and crabs is substantial and growing. No one has shown that any current AI system is conscious, but serious researchers give future systems real odds.

Module 4 Key Points

Other Minds: Animals and Machines

• We infer other minds from behavior and similarity, so the hard cases are animals unlike us and machines that act like us.

• The 2024 New York Declaration found strong support for consciousness in mammals and birds and a realistic possibility in many invertebrates.

• About 88 billion land animals, mostly chickens, were slaughtered for food in 2024, and over a trillion wild fish are caught yearly.

• Singer calls discounting suffering because of species speciesism; critics ground moral status in reason, reciprocity, or relationships.

• GPT-4.5 was judged human 73% of the time in a 2025 Turing test; the indicator study found no current AI conscious.

• Birch’s precaution gives realistic sentience candidates proportionate protection; Washington and California banned octopus farming in 2024.

Module 4 Quiz

Check Your Understanding

1. What did the New York Declaration on Animal Consciousness (April 2024) say?

A) That the evidence strongly supports conscious experience in mammals and birds, and that there is at least a realistic possibility of it in all vertebrates and many invertebrates
B) That only primates are conscious
C) That insects certainly feel pain
D) That fish lack the brain areas needed for any experience

2. Which animals did Britain's 2022 sentience law add to those it recognizes as sentient?

A) Insects and spiders, after a review of studies on bees
B) Octopuses, crabs, and lobsters
C) Only farmed fish and the large mammals kept in zoos
D) Every invertebrate, including worms, snails, and jellyfish

3. In the 2025 three-party Turing test, how often was GPT-4.5, told to adopt a persona, judged to be the human?

A) 21% of the time
B) About half the time
C) 23% of the time, the same as ELIZA
D) 73% of the time, more often than the real humans were

4. What did the 2023 indicator study by Butlin, Long, and others conclude?

A) That no current AI is conscious, but no obvious technical barrier blocks systems with the indicators
B) That large language models are already conscious
C) That machines can never be conscious, because they lack biological neurons and cannot feel what they process
D) That the question cannot be studied

5. In the worked example, why does the result for chickens swing so widely?

A) Because the number of chickens is unknown
B) Because chickens live only a few weeks, so any yearly estimate of their welfare must be multiplied by roughly eight
C) Because each welfare estimate is very uncertain
D) Because philosophers disagree about whether humans count

Module 5 — Free Will and Responsibility

Determinism, Blame, and the Brain

Recommended Reading / Reference

Book: Daniel Dennett and Gregg Caruso, *Just Deserts: Debating Free Will* (Polity, 2021), a compatibilist and a skeptic arguing the question out in three exchanges

Reference: Timothy O'Connor and Christopher Franklin, "Free Will," *Stanford Encyclopedia of Philosophy* (first published 2002), on the positions and the arguments for each

5a. The Problem

In 1814 the mathematician Pierre-Simon Laplace imagined an intellect that knew, at one instant, every force in nature and the position of every particle. For such a mind, he wrote, nothing would be uncertain, and the future would be as plain as the past. Determinism is the thesis behind the image: the state of the world at one time, together with the laws of nature, fixes everything that happens afterward. If it is true, every choice you will make was settled before you were born.

The problem is that much of ordinary life seems to assume otherwise. Blame and praise, punishment and gratitude, and the sense of deliberating about what to do all seem to presuppose that an act was up to the person, or that the person could have done otherwise. Abandoning determinism does not obviously help. If some events in the brain are random, as some readings of quantum physics allow, a choice that turns on such an event seems a matter of luck, not control. A free act must apparently be neither fixed in advance nor random, and it is hard to say what else it could be.

5b. Three Answers

Skeptics hold that free will of the kind blame requires does not exist. Baruch Spinoza wrote that people think themselves free because they are aware of their desires but ignorant of the causes of those desires. Derk Pereboom argues that we lack free will whether or not determinism is true, since determined acts are not up to us and undetermined ones are luck, a view he calls hard incompatibilism. The neuroscientist Robert Sapolsky argued in *Determined* (2023) that genes, upbringing, hormones, and the moment-to-moment state of the brain leave no room for a free choice.

Libertarians hold that we do have free will and that it requires the world not to be determined. Thomas Reid and, in 1964, Roderick Chisholm argued that a free agent is a cause in her own right, not merely a link in a chain of events. Robert Kane has argued since 1996 that free will lives in moments of inner conflict, when the brain's activity is not settled in advance and the person's own effort decides which way it goes; in such self-forming acts, he holds, a person shapes the character from which later acts flow.

Compatibilists hold that free will and determinism can both be true. Thomas Hobbes and David Hume defined liberty as the power to act according to one's own will, without being prevented or forced. The Stoic Chrysippus, in the third century BC, compared an action to a cylinder pushed down a slope: the push starts it, but it rolls because of its own shape. Fate supplies the push, and a person's character determines how she responds, so her acts are her own even though they have causes.

Harry Frankfurt refined the view in 1971. Persons, he argued, have desires about their desires: a drug addict may desire the drug while wishing he did not, and he lacks free will in that he cannot make the desire he

wants to act on be the one that moves him. A person has free will when the desire that moves her is the one she wants to move her. Others, among them John Martin Fischer and Mark Ravizza (1998), tie responsibility to responsiveness to reasons: a person is responsible for an act if it comes from a process of her own that would have responded to good reasons to act differently.

5c. Frankfurt's Case

Worked example — Black and Jones

Many arguments against free will assume the principle that a person is morally responsible for an act only if he could have done otherwise. In 1969 Harry Frankfurt described a case meant to refute it. Black wants Jones to commit a certain act, and, in the usual retelling, he has secretly fitted Jones's brain with a device that will force him to do it if Jones shows any sign of deciding not to. As it happens, Jones decides to do the act on his own, for his own reasons, and the device never does anything. Step one: Jones could not have done otherwise, since the device would have stopped him. Step two: Jones is nonetheless responsible, since he acted exactly as he would have with no device at all; the device made no difference to what happened. Step three: so a person can be responsible without being able to do otherwise, and the principle is false. If the case works, what matters for responsibility is the actual history of an act, how it came about, not the existence of alternatives, which is what compatibilists say. Critics answer that the case needs a sign that Jones is about to choose otherwise, and that this flicker of an alternative is still a kind of freedom. Others argue that if determinism is false, no early sign could tell Black in time which way Jones would go, and if it is true, the case assumes what it sets out to prove. The case remains one of the most discussed thought experiments in the debate.

5d. The Reactive Attitudes

P. F. Strawson changed the subject in "Freedom and Resentment" (1962). Resentment, indignation, gratitude, and love, he argued, are not conclusions from a theory about causes that a discovery in physics could overturn; they are part of what it is to live among other people. We already know when to suspend them: when someone did not mean the harm, was pushed, or did not know, and when the person is a young child or severely mentally ill. Determinism, Strawson argued, would be true of everyone at all times, so it gives no one an excuse of either kind. His compatibilist answer is that blame is justified by its place in human relationships, not by a fact beneath them.

5e. God's Power and Human Acts

For believers in an all-powerful God the problem takes another form: if God creates everything and knows the future, how can humans be responsible for what they do? Early Islamic theology split over the question. The Muʿtazila held that humans create their own acts, since a just God would not punish people for what they could not help. The Jabriyya held that God compels every act. The school of al-Ashʿarī, which became dominant in Sunni theology, took a middle path: God creates every act, and the human being acquires it, a relation the school called *kasb*, which makes the person accountable.

Christian theology ran a parallel debate, from Augustine's writings on grace and the will in the early fifth century to the 1524–1525 exchange between Erasmus, who defended free will, and Martin Luther, who answered with *On the Bondage of the Will*. In both traditions the theological problem has the same shape as the scientific one: something other than the agent seems to settle what the agent will do.

5f. What the Brain Experiments Show

In 1983 Benjamin Libet and colleagues asked people to flex a wrist whenever they felt like it and to note, on a fast-moving clock, when they first felt the urge. A slow build-up of electrical activity over the brain, called the readiness potential, began about 550 milliseconds (thousandths of a second) before the movement, while people reported the urge about 200 milliseconds before it. The brain seemed to start the act some 350 milliseconds before the person decided to do it, and the finding has often been read as proof that free will is an illusion. Libet himself suggested that the conscious mind keeps a veto in the last fraction of a second.

Later work undercut the reading. In 2012 Aaron Schurger, Jacobo Sitt, and Stanislas Dehaene showed that the readiness potential can be produced by random fluctuations in brain activity that happen to cross a threshold: averaging the signal backward from many movements creates a ramp, even if nothing was decided in advance. In 2019 Uri Maoz and colleagues found a readiness potential before arbitrary button presses but not before deliberate choices of which of two charities to fund with real money. The experiments concern meaningless flicks of the wrist, not the considered choices that blame is about.

This is the narrower point on which evidence settles something, and the course says it once: the brain experiments do not show that free will is an illusion. They show that the timing of an arbitrary urge can be preceded by brain activity, which every view of free will, compatibilist and libertarian alike, already allows, since all sides accept that choices are carried out through the brain.

5g. What Ordinary People Think

Shaun Nichols and Joshua Knobe described a fully deterministic universe to people in a 2007 study and asked two kinds of question. Asked in the abstract whether anyone in such a universe could be fully morally responsible, 86% said no. Asked about a man named Bill in that same universe, who kills his wife and three children to be with his secretary, 72% said that Bill was fully morally responsible. People seem to be incompatibilists in the abstract and compatibilists when faced with a vivid wrong.

Each side reads the result its own way. Incompatibilists say the concrete case provokes anger that distorts judgment. Compatibilists, citing studies by Eddy Nahmias and others, say many people misread determinism as meaning that their thoughts and desires make no difference, which no compatibilist believes, so the abstract answers rest on a confusion. Either way, ordinary intuition does not clearly favor one side.

5h. Punishment Without Desert?

If no one deserves blame in the deepest sense, punishment needs another justification. Retributivists, following Kant, hold that wrongdoers deserve to suffer in proportion to their wrongs, which skeptics deny anyone does. Consequentialists justify punishment by its effects: deterring crime, keeping dangerous people away from others, and reforming offenders. Pereboom proposes a quarantine model: society may confine a person who is dangerous, as it may quarantine a carrier of a deadly disease, but only with the least harm that safety requires and with care for the person’s welfare and return.

The law already reduces blame where capacities are reduced. In *Roper v. Simmons* (2005) the US Supreme Court barred the death penalty for crimes committed before the age of 18, citing evidence that adolescents are more impulsive and less mature, and in *Miller v. Alabama* (2012) it barred mandatory life sentences without parole for juveniles. Among philosophers in 2020, 75.1% held capital punishment impermissible and 17.7% permissible.

5i. The Argument: Do We Have the Free Will That Blame Requires?

The skeptic's strongest case joins two arguments. The consequence argument, published by Peter van Inwagen in 1975 and developed in 1983, holds that if determinism is true, our acts are consequences of the laws of nature and of events in the remote past; neither the laws nor the past are up to us; so our acts are not up to us. Van Inwagen himself concludes that determinism is false, but the skeptic adds that if it is false, randomness adds only luck. Galen Strawson adds the basic argument: to be ultimately responsible for what you do, you would have to be responsible for the character you act from, which would require having made yourself, and no one can do that.

Compatibilists answer that what matters in blame survives determinism: acting from one's own reasons, being able to respond to reasons, and having the will one endorses. Frankfurt's case suggests that alternatives are not needed, and P. F. Strawson's reactive attitudes (5d) suggest that no metaphysical fact could make blame pointless. Libertarians answer that both views abandon something real, the open future that deliberation presupposes, and that physics has not shown the brain to be determined.

What each answer costs

Answer	What it explains	What it must accept or explain away
Hard incompatibilism (Spinoza, Pereboom, Sapolsky)	Why no one ultimately deserves blame, and why offenders might be treated more humanely	Giving up desert, retributive punishment, and perhaps the deepest forms of resentment, praise, and gratitude
Libertarianism (Kane, Chisholm)	The sense that the future is open and that choices are up to the person making them	That freedom depends on undetermined events in the brain at the right moments, and showing that such events are control rather than luck
Compatibilism (Hume, Frankfurt, P. F. Strawson)	How responsibility fits a world of causes, and why excuses such as coercion and compulsion matter	That a person can deserve blame for an act whose causes were fixed before she was born

Whether we have free will is one of the questions on which the course takes no side. Where the argument stands: most philosophers, 59.2%, are compatibilists, 18.8% libertarians, and 11.2% deny free will, and neither the brain experiments (5f) nor ordinary intuitions (5g) settle the question.

Module 5 Key Points

Free Will and Responsibility

• If determinism is true, every choice was fixed before birth; if choices are random instead, they seem no freer.

• Compatibilists, 59.2% of philosophers, say freedom is acting from one's own reasons, which causes need not threaten.

• Frankfurt's 1969 case suggests responsibility needs no ability to do otherwise, only the right history behind the act.

• Libet's brain signal preceded the felt urge by about 350 milliseconds, but it may reflect random noise, not a hidden decision.

• In a deterministic universe, 86% said no one is fully responsible; told of Bill's murders, 72% held him fully responsible.

• Pereboom would replace desert with a quarantine model: confine the dangerous with no more harm than safety requires.

Module 5 Quiz

Check Your Understanding

1. What did the Stoic Chrysippus illustrate with a rolling cylinder?

A) That an act can be caused and still come from the agent's own nature

B) That fate decides nothing

C) That the future is open because the smallest particles of matter swerve at random, as the Epicureans taught

D) That people should accept whatever happens to them without trying to change it

2. What does Frankfurt's case of Black and Jones aim to show?

A) That determinism is false

B) That a person can be morally responsible for an act even when he could not have done otherwise, because what matters is how the act actually came about

C) That neuroscience can predict choices

D) That coercion excuses all wrongdoing

3. What did Libet's 1983 experiment measure?

A) Whether people could change their minds about a decision after they had begun to carry it out

B) How accurately a brain scanner could read thoughts from a distance

C) The timing of brain activity and of the felt urge

D) Which of two charities people chose to give money to

4. In Nichols and Knobe's 2007 study, what did most people say about Bill, who killed his family in a deterministic universe?

A) That no one is responsible

B) They declined to answer

C) That Bill could have done otherwise, since the universe they were told about allowed him a choice

D) That Bill was fully morally responsible

5. What does Pereboom's quarantine model say justifies confining dangerous offenders?

A) That offenders deserve to suffer

B) Deterrence alone

C) The wishes of the victims' families

D) The need to protect others, as with quarantining a person who carries a dangerous disease, using no more harm than safety requires and without claiming the offender deserves to suffer

Module 6 — Personal Identity

What Makes You You

Recommended Reading / Reference

Book: John Perry, *A Dialogue on Personal Identity and Immortality* (Hackett, 1978), a short conversation at a dying philosopher's bedside that tests the soul, the body, and memory as answers

Reference: Eric T. Olson, "Personal Identity," *Stanford Encyclopedia of Philosophy* (first published 2002), on the questions and the main views, including his own animalism

6a. The Ship of Theseus

Plutarch, writing about 100 AD, reported that the Athenians kept the ship in which the hero Theseus had sailed, replacing its planks one by one as they rotted, until philosophers argued over whether it was still the same ship. Thomas Hobbes added a twist in 1655: suppose someone gathered the discarded planks and rebuilt the ship from them. Which of the two ships would then be the original?

The puzzle separates two kinds of sameness. Two coins from the same mint are qualitatively identical, exactly alike, but they are two coins. Numerical identity means being one and the same thing, and it is the kind that matters for persons. The question of personal identity asks what makes the person who wakes up tomorrow, or who stood trial for a crime 20 years ago, one and the same person as someone who exists now. Bodies change their matter over years, though most nerve cells in the brain last a lifetime, so the answer is not simply "the same material."

6b. Memory and the Brave Officer

John Locke gave the first modern answer in the second edition of his *Essay Concerning Human Understanding* (1694). A person, he wrote, is a thinking being that can consider itself as itself at different times and places, and personal identity reaches as far back as its consciousness, its memory, reaches. If the soul of a prince, carrying his memories, entered the body of a cobbler, the person who woke in the cobbler's body would be the prince. Locke thought the view explained why punishment fits only acts a person can remember as his own.

Thomas Reid answered in 1785 with a case. A boy is flogged for robbing an orchard. As a brave young officer he remembers the flogging, and as an old general he remembers taking an enemy standard as an officer but has forgotten the flogging. On Locke's view the general is the officer and the officer is the boy, but the general is not the boy, since he cannot remember being him. Identity cannot work that way: if A is the same as B and B the same as C, A is the same as C. Joseph Butler had added in 1736 that memory presupposes identity rather than creating it, since one can only remember one's own experiences.

Later defenders repaired the view. Psychological continuity, as Sydney Shoemaker and Derek Parfit defined it, needs only overlapping chains of memory, intention, belief, and character, each link connected to the next, as the general is linked to the boy through the officer. In 2020, 43.7% of philosophers favored a psychological view of personal identity.

6c. Body, Brain, or Animal

A rival view says that you are your body, or more precisely, a human animal. Animalism, defended by Eric Olson since 1997, holds that you persist as long as the living organism you are persists, whatever happens to your memories, so you would survive in a permanent coma and you were once a fetus. It was the view of 19.1% of philosophers in 2020. Its difficulty appears in a case Sydney Shoemaker described in 1963, in which Brown's brain is placed in Robinson's skull. The person who wakes up has Brown's memories, character, and plans, and most people judge that he is Brown, while animalism says he is Robinson's body with a new brain.

A third view, held by 14.9% of philosophers in 2020, says identity is a further fact beyond body and psychology, such as the persistence of a soul. Richard Swinburne has defended a version. On this view, whether a future person is you is all or nothing, and it could be true or false even when every physical and psychological fact is known.

6d. Parfit's Teletransporter

Derek Parfit pressed the psychological view to its limits in *Reasons and Persons* (1984). Imagine a machine that scans every cell of your body on Earth, destroys it, and builds an exact copy on Mars from new matter. The person on Mars has your memories, plans, and character and believes he is you. Is he you, or a new person, a copy, while you have died? In 2020, 40.1% of philosophers said the machine kills you and 35.2% that you survive.

Parfit then varied the case. Suppose the scanner fails to destroy the original, who learns that his heart was damaged by the scan and that he will die within days, while his replica flourishes on Mars. Or suppose fission: the two halves of your brain, each able to support your mind, are transplanted into two bodies, and two people wake up, each psychologically continuous with you. They cannot both be you, since they are two people, and there is no reason to pick one over the other.

Parfit concluded that identity is not what matters in survival. What matters is psychological connectedness and continuity, which can hold to two people at once and come in degrees. He found the conclusion liberating: before, he wrote, his life had seemed a glass tunnel moving toward death, and afterward the walls of the tunnel seemed to disappear, and he cared less about his own future and more about others.

6e. No Self

Parfit noted that his view was close to the Buddha's. The doctrine of *anattā*, no-self, holds that a person is a changing bundle of five aggregates: the body, feelings, perceptions, dispositions, and consciousness. None of them is a permanent self, and there is no self apart from them. In the *Questions of King Milinda*, a Buddhist text compiled from about the first century BC, the monk Nāgasena asks King Milinda, the Greek ruler Menander I of northwest India, whether his chariot is its pole, axle, wheels, or frame. It is none of them, nor anything apart from them: "chariot" is a name for the parts put together, as "Nāgasena" names the aggregates.

David Hume reached a similar view in 1739. When he looked into himself, he wrote, he always stumbled on some particular perception, of heat or cold, love or hatred, and never caught himself without one; the self is a bundle of perceptions in constant motion. In the Buddhist tradition the point is practical as well as theoretical (Module 14).

6f. Personhood as Achieved

Some traditions ask a different question: not what keeps a person the same, but what makes someone a person in the full sense. The Nigerian philosopher Ifeanyi Menkiti argued in 1984 that in much African traditional thought personhood is not given at birth but achieved, as one is drawn into a community and takes up its obligations; a newborn is not yet fully a person, and an elder is more fully one than a youth. The Ghanaian philosopher Kwame Gyekye replied that every human being has the dignity of a person from birth, and that what is achieved is moral excellence, a view he called moderate communitarianism.

A related Western view treats the self as a story. Marya Schechtman argued in 1996 that a person is constituted by the narrative she tells about her life, which links past, present, and future into one self. Such views explain why a person can say, after a conversion or a crime, that she is not the person she was, a claim that is false on the strict view of identity and true in another sense that matters to people.

6g. What Rides on It

Three practical questions turn on personal identity. The first is dementia. Ronald Dworkin discussed in *Life's Dominion* (1993) a woman with Alzheimer's disease who seemed content, reading the same pages of a book over and over. Suppose she had written, while well, that in such a state she wanted no treatment for a life-threatening illness. Dworkin argued that her earlier self's judgment about how her life should go should govern, a principle he called precedent autonomy. Rebecca Dresser answered that the interests of the contented person she is now count too, and that she may not be the same person in the sense that matters.

The second is the future self. If the self is less unified over time than it seems, Parfit argued, it is not irrational to care less about one's distant future, and many people already treat their future selves almost as strangers, saving too little and smoking too much. The third is the digital copy: companies now build chatbots from a dead person's messages and recordings so that relatives can "talk" to them. On the psychological view, such a copy raises the question of what continuity requires; nearly everyone judges that a chatbot that imitates a person is not that person, which suggests that continuity must run through the person's own mind.

6h. The Argument: Is Identity What Matters?

Parfit's case is that the fission case leaves nothing to dispute: everything a person could want from survival, memory, character, plans, and relationships carried on, is present for each successor, yet neither is strictly the original. So identity cannot be what matters. If that is right, many practical questions, from the value of a long life to the authority of an advance directive, should be judged by degrees of connection rather than by a yes-or-no question of identity.

His opponents answer from several sides. Animalists say the teletransporter kills you, so the case shows only that a copy is not you. Others hold that common sense is right that a future person either is or is not me, so there must be a further fact. Still others, among them Susan Wolf, argue that people's concern for their own futures is bound up with identity, and that giving it up would undermine prudence, commitment, and responsibility for past acts.

Worked example — The Directive and the Patient

Dutch law on euthanasia, in which a doctor gives a lethal drug at a patient's request (Module 11), lets a written request stand in for a spoken one when a patient can no longer express her will. In 2016 a 74-year-old woman with Alzheimer's disease, who years earlier had written that she wanted euthanasia when she judged the time right and before she had to enter a care home, was living in a nursing home and no longer able to confirm her wish. Her doctor put a sedative in her coffee and began the procedure; when she woke and resisted, her family held her. The doctor was prosecuted, acquitted in 2019, and on April 21, 2020 the Dutch Supreme Court ruled that a doctor may follow a clear written directive for a patient with advanced dementia who can no longer express her will. In 2025, 11 of the Dutch dementia cases counted in Module 11 were of this kind. Now run the theories. On the psychological view, the earlier writer and the later patient are linked only weakly, by a few surviving memories, so the writer's authority over the patient is in doubt. On animalism, they are one person, and the question becomes which of her interests should govern, her earlier judgment about her life or her present wish not to die. On a no-self view, the case shows two stages of a changing process with conflicting wishes, neither of which owns the other. The theory of identity does not settle the case, but each theory changes which question has to be answered.

What each answer costs

Answer	What it explains	What it must accept or explain away
Psychological continuity (Locke, Parfit)	Why the prince in the cobbler's body is the prince, and why memory and character seem to matter most	That a perfect copy could be you, and that fission makes identity a matter of degree or of convention
Animalism (Olson)	Why you were once a fetus and would survive in a permanent coma	That you would not go where your brain goes, even if your memories and character went with it
No-self (the Buddha, Hume)	Why no single part or substance can be found that is the self	Who is responsible for past acts, and why it is rational to care about one's own future at all
A further fact (a soul)	The conviction that being me is all or nothing	Something no experiment can detect, and an account of what it adds to body and memory

The course's reading: the fission cases show that what people care about in survival can come apart from strict identity. Practical questions such as advance directives therefore cannot be settled by the metaphysics alone; they also require deciding whose interests, the earlier self's or the later one's, should count, and how much.

Module 6 Key Points

Personal Identity

• Numerical identity, being one and the same thing, is the puzzle; the ship of Theseus asks what preserves it through change.

• Locke tied identity to memory; Reid’s brave officer showed that memory alone breaks the logic of identity.

• 43.7% of philosophers favor psychological continuity and 19.1% the biological view; 40.1% say the teletransporter kills you.

• Parfit argued from fission that identity is not what matters; continuity of memory and character is, and it comes in degrees.

• The Buddha taught no-self: a person is five changing aggregates, as Nāgasena’s chariot is its assembled parts.

• The Dutch Supreme Court ruled in 2020 that a clear advance directive can authorize euthanasia in advanced dementia.

Module 6 Quiz

Check Your Understanding

1. What did Thomas Reid’s brave officer case show about Locke’s memory theory?

A) That it contradicts itself
B) That memories are usually false, so a theory of identity based on them cannot work
C) That a person is the same as his body, however much he forgets
D) That officers forget more than they remember

2. What did Derek Parfit conclude from cases of fission?

A) That the soul goes with the brain
B) That identity is not what matters in survival
C) That a teletransporter always kills the person who steps into it and creates a stranger with the same memories
D) That both successors are you

3. In the Questions of King Milinda, what is Nāgasena’s point about the chariot?

A) That a chariot is its wheels
B) That only kings may own chariots, as the Buddha taught in the sermons the king had heard
C) That the name refers to parts assembled, not a further self
D) That the chariot existed before its parts and will outlast them

4. What did Ifeanyi Menkiti argue about personhood in African traditional thought?

A) That persons are souls
B) That personhood is achieved through community and obligations
C) That personhood is fixed at birth by an ancestor’s choice and cannot be lost whatever a person does
D) That only elders can own land

5. What did the Dutch Supreme Court rule in April 2020 in the dementia case?

A) That a doctor may follow a clear written advance directive to perform euthanasia on a patient with advanced dementia who can no longer express her will
B) That euthanasia for dementia is banned
C) That the doctor must serve prison time
D) That only spoken requests can count

Module 7 — Does God Exist?

The Arguments on Each Side

Recommended Reading / Reference

Book: William Lane Craig and Walter Sinnott-Armstrong, *God? A Debate between a Christian and an Atheist* (Oxford University Press, 2004), in which each philosopher makes his case and answers the other's

Reference: Charles Taliaferro, "Philosophy of Religion," *Stanford Encyclopedia of Philosophy* (first published 2007), on the arguments and the main debates in several traditions

7a. The Question

The arguments in this module concern the God of classical theism, shared in its main lines by Judaism, Christianity, and Islam and by theistic schools of Hinduism: a being that is all-powerful, all-knowing, perfectly good, and the creator of everything else. Other ideas of the divine exist, such as many gods, a God identical with nature, as Spinoza held, or a creator who does not intervene, as the deists of the eighteenth century held. A theist believes that God exists, an atheist that God does not, and an agnostic suspends judgment.

The arguments fall into families. Some reason from the existence of the world to a cause, some from its order to a designer, one from the very idea of God, and some from experience. On the other side, the strongest arguments start from suffering and from the fact that God, if real, is hidden. This module gives each family its best form and its strongest objection; the traditions themselves, their practices and institutions, are a separate subject.

7b. The First Cause

Cosmological arguments reason from the existence of the world to a cause beyond it. Aristotle argued in the fourth century BC that every motion needs a mover, that the chain of movers cannot go back forever, and so that there must be a first, unmoved mover. Muslim theologians developed a sharper version, called the kalām argument after the Arabic word for theology. As al-Ghazālī put it, whatever begins to exist has a cause of its beginning; the world began to exist, since an infinite past is impossible; so the world has a cause. Since the cause existed before time and produced a world at a moment, al-Ghazālī argued, it must have chosen to, so it is a will, an agent.

Thomas Aquinas gave five ways to God in his *Summa Theologiae*, from motion, causation, the contingency of things, degrees of perfection, and order. Gottfried Wilhelm Leibniz asked in 1714 why there is something rather than nothing. Each thing in the world exists contingently, he argued, depending on something else, and even an endless series of contingent things needs an explanation outside itself, which only a being that exists necessarily could give. The American philosopher William Lane Craig revived the kalām argument in 1979, citing the Big Bang as evidence that the universe had a beginning.

Hume's objections in his *Dialogues Concerning Natural Religion* (1779) still frame the reply. If each member of a series is explained by an earlier one, nothing more is needed to explain the series; and if something must exist necessarily, why not the universe itself rather than God? The question "what caused God?" misses the argument, which concerns only things that begin or depend on something else, and God is held to do neither. A sharper objection comes from physics: the Big Bang marks where current theory breaks

down, and some models of the earliest universe, such as Stephen Hawking and James Hartle’s, describe a universe without a first moment in the ordinary sense.

7c. Design and Fine-Tuning

William Paley opened his *Natural Theology* (1802) with a watch found lying on a heath. Its parts fit together for a purpose, so we infer a watchmaker; the eye fits its parts together for seeing far more intricately, so we should infer a designer of the eye. Hume had already objected, 23 years earlier, that the world resembles a watch less than it resembles an animal or a plant, which grow without designers, and that even a designed world would not show one designer, or a perfect one. In 1859 Charles Darwin’s theory of natural selection explained how the eye could arise without a designer, by small inherited changes that helped their bearers survive.

The modern form of the argument looks at the laws of physics instead. Physicists have calculated that if certain constants of nature, such as the strength of gravity or the balance between the forces inside atoms, were somewhat different, stars, stable atoms, or chemistry complex enough for life could not form. The value of the cosmological constant, which drives the accelerating expansion of the universe, is smaller than the simplest quantum estimates of it by a factor of about 10 to the power of 120, a gap physicists themselves call a puzzle.

Defenders of design say that a universe so finely balanced for life is far more likely if a designer intended life than if the constants were set by chance. Critics answer in three ways. There may be many universes with different constants, a multiverse, and observers could only ever find themselves in one that permits observers, a point called the anthropic principle. A deeper theory may show that the constants could not have been otherwise. And it is unclear how to assign probabilities to constants that could in principle take any of infinitely many values. Theists reply that a multiverse is an unobserved posit too, and that the mechanism generating it would itself need fine-tuning.

7d. The Idea of God

Anselm of Canterbury offered in his *Proslogion* (1078) an argument from the concept of God alone, later called the ontological argument. God is that than which nothing greater can be conceived. Even a fool who denies God understands the phrase, so God exists at least in the understanding. But a being that exists in reality as well as in the understanding is greater than one that exists in the understanding alone. So if God existed only in the understanding, a greater being could be conceived, which is impossible by definition; therefore God exists in reality.

A monk named Gaunilo answered at once that the same reasoning would prove the existence of an island greater than any that can be conceived. Kant argued in 1781 that existence is not a property that makes a thing greater: a hundred real thalers, the coins of his day, contain not a coin more than a hundred possible ones. Alvin Plantinga gave a modern version in 1974: if it is even possible that a maximally great being exists, it exists necessarily. Critics answer that, by the same logic, if it is even possible that no such being exists, then none does, so the argument decides nothing until we know which premise to accept.

7e. The Problem of Evil

The oldest argument against God is a dilemma that the Christian writer Lactantius attributed to Epicurus in *On the Anger of God*, around 313 AD. If God wishes to prevent evil but cannot, he is not all-powerful; if he can but does not wish to, he is not good; if he both can and wishes to, where does evil come from? Hume's *Dialogues* restated it. The Hebrew Bible's book of Job tells of a righteous man stripped of children, wealth, and health; when Job demands an answer, God replies from a whirlwind by asking where Job was when the world was made.

In 1955 J. L. Mackie argued that the existence of God and of evil are logically inconsistent. Alvin Plantinga answered with the free-will defense (1974): a world with free creatures who sometimes choose wrongly may be better than a world without freedom, and it is possible that not even God could create free creatures who always freely choose right, which is all a defense needs to show. Many philosophers, theists and atheists alike, now accept that this answers the charge of strict inconsistency, though not the next one.

The evidential problem, as William Rowe put it in 1979, asks not whether God and evil are consistent but whether evil makes God improbable. In a distant forest, lightning starts a fire, and a fawn is trapped, badly burned, and suffers for several days before it dies. No good seems to come of it. Such suffering seems pointless, Rowe argued, so it probably is; and if any suffering is pointless, there is no all-powerful, all-knowing, perfectly good God. Natural evils of this kind, from the fawn to the Lisbon earthquake of 1755, which killed tens of thousands on a religious holiday, are harder for the free-will defense to explain than human cruelty.

Two replies lead. John Hick, in *Evil and the God of Love* (1966), developed the idea, which goes back to the second-century bishop Irenaeus, that the world is a vale of soul-making: virtues such as courage and compassion can only be formed in a world with real dangers and suffering. Skeptical theists, following Stephen Wykstra (1984), argue that we should not expect to see God's reasons, any more than an infant can see the reasons for a surgery, so seeming pointlessness is weak evidence of real pointlessness. Critics reply that the second answer cuts both ways, since it also undercuts any confident claim about what a good God would do.

7f. Experience and Hiddenness

Many people report experiences of God's presence, which William James studied in *The Varieties of Religious Experience* (1902). Richard Swinburne argues by a principle of credulity: if it seems to you that something is present, it probably is, unless you have special reason to doubt it, and that principle governs ordinary perception too. William Alston (1991) compared mystical experience to sense perception as a practice that has earned trust. Critics answer that such experiences conflict across traditions, revealing Vishnu to some, Allah to others, and emptiness to Buddhist meditators, and that brain science can trigger similar experiences.

The argument from divine hiddenness, set out by J. L. Schellenberg in 1993, turns the point around. A perfectly loving God would want a relationship with every person capable of one and would make his existence clear enough that no one open to belief would fail to believe. Yet there are people who seek God sincerely and do not find him. If such non-resistant nonbelievers exist, Schellenberg argues, a perfectly loving God does not. Theists reply that a God whose existence was obvious would compel belief and crowd out a relationship freely chosen, so that some hiddenness may serve the very relationship Schellenberg says God would want.

7g. The Debate Beyond Europe

India held the same debate with its own arguments. The Nyāya philosopher Udayana, writing around 1000, argued in his *Handful of Flowers of Logic* that the earth and other composite things are effects, that effects such as pots have intelligent makers, and so that the world has a maker, Īśvara. Buddhist logicians answered that a permanent, unchanging God could not cause a changing world, since an unchanging cause would produce all its effects at once or none of them. The Mīmāṃsā school, which defended the authority of the Vedas, the oldest Hindu scriptures, denied a creator outright and held the Vedas to be eternal and authorless, and Jain thinkers held the universe to be uncreated and eternal.

Faith and reason were argued over in the Islamic world too. In *The Incoherence of the Philosophers* (1095) al-Ghazālī attacked the Aristotelian philosophers on the eternity of the world and other doctrines. He also argued that fire does not burn cotton by any necessity: God creates the burning each time, and the regularity we see is God's custom, a point Hume would make in secular form about necessary connection. Averroes (Ibn Rushd, 1126–1198) replied in *The Incoherence of the Incoherence* that philosophy and revelation, rightly understood, cannot conflict. In Christian Europe Aquinas held faith and reason to be partners, while Søren Kierkegaard held faith to be a passionate commitment that no argument can secure, a view called fideism.

7h. The Argument: Does God Exist?

The theist's strongest case, as Swinburne makes it, is cumulative: the existence of a universe at all, its lawful order, its fitness for life, and the experiences of millions are each more probable if God exists than if not, and together they make theism probable. The atheist's strongest case combines the problem of evil, the hiddenness of God, the steady replacement of supernatural explanations by natural ones, and the principle that the burden of proof lies with the one who asserts. Bertrand Russell compared belief in God to belief in a china teapot orbiting the Sun, too small to be seen: it cannot be disproved, but that is no reason to believe it.

Worked example — Pascal's Wager as a Decision Table

Blaise Pascal's *Pensées*, published in 1670 after his death, argued that belief in God is a wager we cannot avoid making, and that reason can judge the bet even if it cannot settle whether God exists. Draw a table with two columns, God exists and God does not, and two rows, believe and do not believe. If God exists, belief gains an infinite reward, eternal life, and unbelief loses it. If God does not exist, belief costs something finite, some pleasures and some Sunday mornings, and unbelief saves it. The expected value of a choice adds up each outcome's value multiplied by its probability. Now give God any probability above zero, say one in a thousand. The expected value of believing is one-thousandth of infinity, which is still infinite, minus a finite cost, so it is infinite; the expected value of not believing is finite. On these numbers belief wins however small the probability. The best-known objection adds columns. If there might be other gods, among them one who rewards only those who believe in him, or one who punishes belief in the wrong god, or one who rewards honest doubt, then several rows carry infinite rewards in some column, and the expected values no longer favor any one belief. Three further objections remain: beliefs cannot simply be chosen, though Pascal advised acting as believers do until belief follows; a God who rewards belief adopted for gain might not reward it; and infinite rewards break the arithmetic, since even a small chance of coming to believe by a mixed strategy yields an infinite expectation too.

What each answer costs

Answer	What it explains	What it must accept or explain away
Classical theism	Why there is something rather than nothing, why the universe is orderly and fit for life, and why so many report experiences of God	Evil and suffering, including the fawn's, and the hiddenness of God from people who seek him
Atheism	Suffering without the need for a theodicy, and the success of natural explanations	That the universe exists, and has the laws it has, for no reason, or that a multiverse exists
Agnosticism	The balance of the arguments, none of which convinces the other side	That suspending judgment is itself a choice with stakes, as Pascal and James argued
Fideism (Kierkegaard)	How faith can survive the failure of every argument	That faith cannot be defended to those who lack it, and that rival faiths stand on the same footing

Whether God exists is one of the questions on which the course takes no side. Where the argument stands: in 2020, 66.9% of philosophers were atheists and 18.9% theists, yet among specialists in the philosophy of religion, who study these arguments most closely, 78% were theists, against 17% of other philosophers. Each side has an explanation: specialists may have seen that the arguments are stronger than outsiders think, or believers may be drawn to the specialty. No argument on either side has persuaded most of its opponents.

Module 7 Key Points

Does God Exist?

• The kalām argument holds that the universe began and so has a cause; critics ask why that cause must be God.

• Paley argued from design; Darwin explained apparent design, and the fine-tuning argument now meets the multiverse and deeper laws.

• Plantinga’s free-will defense answers the logical problem of evil; Rowe’s fawn presses the evidential problem of pointless suffering.

• Udayana’s Nyāya proofs of a world-maker met Buddhist, Mīmāṃsā, and Jain arguments that no creator exists.

• Pascal’s wager fails if several possible gods offer infinite rewards: the decision table no longer favors any one belief.

• 66.9% of philosophers are atheists and 18.9% theists, but 78% of philosophy-of-religion specialists are theists.

Module 7 Quiz

Check Your Understanding

1. What does the kalām argument claim about the universe?

A) That it has always existed and needs no explanation
B) That it was designed, as its fine-tuned constants show
C) That it must exist, because the idea of a being greater than any other includes existence
D) That it began, so it has a cause

2. How did Gaunilo answer Anselm's ontological argument?

A) By saying God is evil
B) By arguing that existence is not a property, since a hundred real coins contain no more than a hundred imagined ones
C) By showing the same reasoning would prove a perfect island exists
D) By arguing that the universe must have had a first cause at the beginning of time

3. What did William Rowe's example of the fawn illustrate?

A) That animals cannot feel pain, so their deaths raise no moral question for theists
B) That forest fires are caused by human sin
C) That God's existence can be proved from nature
D) Suffering that seems pointless

4. What did al-Ghazālī argue about fire and cotton?

A) That fire is an illusion
B) That cotton burns only when it is dry
C) That fire does not burn cotton by necessity; God creates the burning, and the regular sequence is God's custom, not a necessary connection
D) That philosophers have proved the world is eternal

5. What is the many-gods objection to Pascal's wager?

A) That other possible gods could reward different beliefs
B) That belief is free
C) That expected value cannot be calculated for any decision that involves the afterlife, so the wager has no answer
D) That Pascal himself did not believe in God, as his letters reveal

Module 8 — Is Morality Objective?

Relativism, Realism, and the Rest

Recommended Reading / Reference

Book: Simon Blackburn, *Being Good: A Short Introduction to Ethics* (Oxford University Press, 2001), by a leading defender of expressivism, which opens with the threats to ethics, relativism and the death of God among them

Reference: Geoff Sayre-McCord, “Moral Realism,” *Stanford Encyclopedia of Philosophy* (first published 2005), on what realism claims and the arguments for and against it

8a. Custom Is King

Herodotus, writing about 430 BC, told a story of the Persian king Darius. Darius asked some Greeks at his court what price would persuade them to eat the bodies of their dead fathers, and they said no price would. He then summoned some Callatians, a people of India who ate their dead parents, and asked, with the Greeks listening through an interpreter, what price would persuade them to burn their fathers’ bodies, as the Greeks did. They cried out in horror. Herodotus concluded, quoting the poet Pindar, that custom is king of all.

The story shows a fact about moral life: people disagree, and each side finds the other’s practice abhorrent. Moral relativism draws a further conclusion, that moral claims are true only relative to a culture, so that eating the dead is right for Callatians and wrong for Greeks, and there is no truth beyond those standards. The step from the fact to the theory does not follow on its own. People once disagreed about the shape of the Earth, and the disagreement did not make the Earth’s shape a matter of culture.

The disagreement may also be shallower than it looks. The Greeks and the Callatians both honored their dead and disagreed only about how. A study by Oliver Scott Curry and colleagues of ethnographic records from 60 societies, published in 2019, found that seven kinds of cooperative behavior, among them helping kin, returning favors, and respecting others’ property, were judged good wherever they were discussed.

8b. The Reformer Problem

Relativism’s best-known difficulty concerns reformers. If an act is right in a culture whenever that culture approves of it, then anyone who opposes her own culture’s code is wrong by definition. Abolitionists in the American South, women demanding the vote, and dissidents under any regime would have been mistaken simply because they were in the minority. Moral progress becomes impossible, since a change in a culture’s code is only a change, never an improvement, and no outsider could ever correctly condemn another society’s practice.

Relativists reply that standards can be relative to a culture’s deepest values rather than its current practices, so that reformers can be right by appeal to principles their society already holds. Frederick Douglass did exactly that in 1852 in “What to the Slave Is the Fourth of July?”, confronting the country with the liberty it professed. Critics answer that the reply gives up the theory, since it lets one part of a culture’s code correct another by some standard of consistency or depth that is not itself relative. Critics add that relativists often defend tolerance, but tolerance offered as a value every culture should accept is not relative either.

On this narrower point the course gives its reading once, since realists, error theorists, and expressivists alike reject the view: simple relativism fails the reformer test and turns moral progress into mere change. Whether morality is objective in some fuller sense is a separate question, on which the course takes no side, and the rest of the module sets out the answers.

8c. The Euthyphro Dilemma

In Plato's *Euthyphro*, set in 399 BC on the eve of Socrates' trial, Socrates meets a young man prosecuting his own father, sure that the gods approve. Socrates asks him whether the pious is loved by the gods because it is pious, or is pious because the gods love it. Put in terms of one God and morality, the question is whether an act is right because God commands it, or whether God commands it because it is right.

Each horn has a cost. If acts are right because God commands them, morality seems arbitrary: had God commanded cruelty, cruelty would have been right. If God commands acts because they are right, then rightness exists independently of God, and morality does not depend on religion. Defenders of divine command theory, such as Robert Adams (1979), take a middle path: what is wrong is what is contrary to the commands of a loving God, and since God's nature is essentially good, the commands are not arbitrary. Critics answer that the reply defines God's goodness by a standard of goodness, which reopens the question.

Islamic theology divided along the same line. The Ashʿarites held that good and evil are what God commands and forbids, while the Muʿtazila held that reason can know good and evil independently of revelation, and that God commands the good because it is good.

8d. Is and Ought

David Hume observed in 1740, in the third book of his *Treatise of Human Nature*, that writers on morality pass from statements about what is to statements about what ought to be, without explaining how a conclusion of one kind can follow from premises of another. Examine a willful murder in every light, he wrote, and you find only passions, motives, and actions; the vice appears only when you turn to your own feeling of disapproval. G. E. Moore argued in *Principia Ethica* (1903) that for any natural property, such as pleasure, one can always ask meaningfully whether something with that property is good, so goodness cannot simply be that property. He concluded that goodness is a non-natural property, known by intuition.

Naturalists reply that the question whether water is H_2O was once open too, and the identity was true all the same, so an open question shows only that two concepts differ, not that they pick out different things. The two observations, Hume's and Moore's, frame the modern debate: either moral facts are facts of an unusual kind, or they are natural facts under another description, or there are no moral facts at all.

8e. Four Answers

Realists hold that there are objective moral facts. Non-naturalists, from Moore to Derek Parfit in *On What Matters* (2011), hold that some things matter, and give us reasons, in a way no natural fact captures. Naturalist realists hold that moral facts are natural facts, about well-being and the conditions of human flourishing, much as facts about health are facts about bodies.

J. L. Mackie argued in *Ethics: Inventing Right and Wrong* (1977) for an error theory. Moral claims do purport to state objective facts, he held, but there are none, so all of them are false. Objective values would be

queer: unlike anything else in the universe, they would have a demand to be pursued built into them, and we would need a special faculty to detect them. Mackie did not conclude that anything goes; he urged that morality be deliberately invented and kept, as a set of conventions that make cooperation possible.

Expressivists hold that moral claims express attitudes rather than describe facts. In 1936 A. J. Ayer held that “stealing is wrong” expresses disapproval, as if one said “stealing!” in a tone of horror. Simon Blackburn and Allan Gibbard developed subtler versions on which, they argue, a person can earn the right to call moral claims true and mistaken, a program Blackburn calls quasi-realism. The standing objection, raised by Peter Geach, is that moral claims work in logic like factual ones: in “if lying is wrong, then getting your brother to lie is wrong,” no attitude toward lying is expressed, yet the sentence supports inferences as factual claims do.

Constructivists, following Kant and in different ways John Rawls and Christine Korsgaard, hold that moral truths are constructed by reason: they are what any rational agent would be committed to, or what free and equal persons would agree to under fair conditions. On this view morality is objective in the sense that it binds every rational agent, though not in the sense of existing apart from all minds.

8f. Is Human Nature Good?

Two Confucian thinkers argued over where morality comes from. Mencius held that human nature is good. Anyone who suddenly saw a child about to fall into a well, he wrote, would feel alarm and compassion, not in order to win the parents’ favor or a reputation, but because the feeling is part of human nature. People have four sprouts, of compassion, shame, deference, and a sense of right and wrong, which grow into virtues if cultivated, as a mountain once covered in forest can be cut and grazed bare until people think baldness is its nature.

Xunzi held that human nature is bad: people are born wanting gain and sensual pleasure, and if they follow their nature, strife and disorder follow. Goodness, he argued, is the product of deliberate effort, shaped by ritual, music, and teachers as a warped board is straightened by a press. Xunzi did not think morality arbitrary, since the sage kings designed the rituals to solve real problems of scarcity and conflict; but on his view morality is a human achievement rather than an expression of human nature.

8g. The Evolutionary Challenge

Sharon Street argued in 2006 that evolution poses a dilemma for moral realism. Natural selection shaped our deepest moral responses, such as caring for our children, returning favors, and punishing cheats, because they helped our ancestors survive and reproduce, not because they track moral facts that exist apart from us. If realism is true, it would be a remarkable coincidence for evolved values to match the independent moral truths; more likely our moral beliefs are unreliable. A realist must either explain the match or accept that most of our moral beliefs are probably false.

Realists reply in two ways. David Enoch and others argue that there is no coincidence, because survival and flourishing are themselves good, so values that served survival tended to track what really matters. Others point to companions in guilt: evolution shaped our sense of number too, and no one concludes that arithmetic is unreliable, since the abilities it selected could later reach truths beyond survival.

8h. The Argument: Are There Moral Facts?

The realist's case starts from how moral thought behaves. People argue about morality, give reasons, change their minds, and judge some past views mistaken, as they do about facts. Some claims seem true whatever anyone thinks: torturing a child for fun would be wrong even in a society that approved of it. And moral progress, which almost everyone believes in, seems to require a standard to make progress toward.

The anti-realist's case starts from what moral facts would have to be. They would be queer, unlike any other facts and detectable by no instrument; disagreement about them persists among careful people in a way that disagreement about physics does not; and evolution explains moral beliefs without supposing they track anything. Expressivists add that the realist pays a high price for nothing, since everything people do with moral language can be explained by attitudes.

Worked example — Was Abolition Progress?

Test each view on a judgment almost everyone now accepts: that the abolition of slavery in the nineteenth century was moral progress, not just change. Realism answers yes without strain: slavery was always wrong, as it is now, and abolitionists saw a moral truth that their opponents missed, as astronomers see facts that others miss. Simple relativism answers no: by the standards of the slaveholding societies slavery was right, and by ours it is wrong, so abolition made those societies more like ours, and calling that progress only means approving of it. Error theory says that "slavery is wrong" is, strictly, false, like every moral claim, though one may still prefer abolition and work to keep the convention against slavery. Expressivism says that "abolition was progress" expresses our condemnation of slavery, and quasi-realists add that it also expresses the judgment that the change corrected attitudes that rested on false beliefs about race and on the slaveholders' interests, which lets them use the word "progress" in good conscience. Constructivism answers yes: slavery treats persons merely as means, which no rational agent could accept, so abolition brought practice closer to what reason requires. Douglass's speech (8b) shows what all the views but simple relativism can share.

What each answer costs

Answer	What it explains	What it must accept or explain away
Non-naturalist realism (Moore, Parfit)	Moral progress, moral error, and the sense that some acts are wrong whatever anyone thinks	Facts unlike any others, known by intuition, and an account of how evolved minds came to know them
Naturalist realism	How morality fits the natural world, as facts about well-being and flourishing	Moore's open question, and why a natural fact should carry the authority of an "ought"
Relativism	Moral diversity, and the appeal of tolerance	That reformers are wrong by definition, and that no society's practices can be condemned from outside
Error theory (Mackie)	Moral language taken at face value, without queer facts	That "torturing a child for fun is wrong" is false
Expressivism (Ayer, Blackburn)	Why moral judgment moves people to act, in a world described by science	The logic of moral reasoning in conditional sentences, and how an attitude can be mistaken
Constructivism (Kant, Rawls, Korsgaard)	Objectivity grounded in reason itself, binding on every rational agent	Showing that reason alone yields substantive moral conclusions, and what binds those who reject the procedure

Whether there are objective moral facts is one of the questions on which the course takes no side. Where the argument stands: in 2020, 62.1% of philosophers were moral realists and 26.1% anti-realists, and 69.3% held that moral judgments state beliefs that can be true. Beyond the narrower point of 8b, each answer keeps its defenders and its costs.

Module 8 Key Points

Is Morality Objective?

• Moral diversity is a fact, but it does not prove relativism, just as disagreement over the Earth’s shape did not make it relative.

• Simple relativism makes every reformer wrong by definition; realists and anti-realists alike reject it.

• The Euthyphro asks whether acts are right because God commands them, or commanded because they are right.

• Mencius held human nature good, with sprouts of virtue; Xunzi held it bad, made good by ritual and effort.

• Street argues that evolution, not truth, shaped moral beliefs; realists reply that evolved values can still track what matters.

• 62.1% of philosophers are moral realists and 26.1% anti-realists, among them error theorists and expressivists.

Module 8 Quiz

Check Your Understanding

1. What did Darius's test of the Greeks and the Callatians show, according to Herodotus?

A) That Greeks eat their dead
B) That each people was horrified by the other's funeral custom
C) That every culture, once it understands the reasons behind another's practices, comes to accept them
D) That Persian law forbade burial

2. What is the reformer problem for moral relativism?

A) That reformers usually fail, so their views cannot be the moral truth about their societies
B) That reformers rarely agree with each other about what should change
C) That most reforms come from outside a culture
D) That reformers are wrong by definition

3. What does the Euthyphro dilemma ask?

A) Whether acts are right because God commands them, or commanded because they are right
B) Whether piety matters
C) Whether Socrates was guilty of corrupting the youth of Athens, as his accusers in court charged
D) Whether the gods agree about justice

4. What is Sharon Street's evolutionary challenge to moral realism?

A) That animals have no morality
B) That evolution shaped our moral beliefs to help us survive and reproduce, so it would be a coincidence if they matched objective moral truths
C) That moral rules change too fast to have evolved
D) That evolution proves selfishness is good

5. On Mencius's view, what does the sight of a child about to fall into a well show?

A) That children need guards
B) That people help only when they hope for praise or reward from the child's parents
C) That everyone has a sprout of compassion
D) That goodness comes only from ritual and teachers

Module 9 — Right and Wrong

Consequences, Duties, and Character

Recommended Reading / Reference

Book: James Rachels and Stuart Rachels, *The Elements of Moral Philosophy* (McGraw-Hill, many editions since 1986), the most widely used short introduction to the theories of right action

Reference: Edmond Awad and colleagues, “The Moral Machine Experiment,” *Nature* (October 24, 2018), the online experiment that asked people around the world whom a self-driving car should spare

9a. Why Be Moral?

In Book II of Plato’s *Republic*, written about 375 BC, Glaucon tells of the ring of Gyges. A shepherd finds a ring that makes him invisible when he turns it; he uses it to seduce the queen, murder the king, and take the throne. Anyone with such a ring, Glaucon argues, would do the same, which shows that people value justice only for what it brings them and would abandon it if they could escape detection. He challenges Socrates to show that justice is good in itself.

Theories of right action answer a related question: what makes a right act right? Three families dominate. Consequentialists look to the results an act produces, deontologists to duties and rights that hold whatever the results, and virtue ethicists to the character an act expresses. In the 2020 survey philosophers divided almost evenly: 37.0% favored virtue ethics, 32.1% deontology, and 30.6% consequentialism, and many chose more than one.

9b. Consequences: Mozi and Bentham

The Chinese thinker Mozi judged every doctrine and practice by one test, whether it benefits the people: whether it makes them richer, more numerous, and better ordered. By that test he condemned wars of conquest and the lavish funerals and court music of his day as waste. He also taught inclusive care (Module 10). Many scholars count him the first consequentialist.

Jeremy Bentham founded utilitarianism in his *Introduction to the Principles of Morals and Legislation* (1789). Nature, he wrote, has placed mankind under two masters, pain and pleasure, and the right act is the one that produces the greatest happiness overall, each person counting for one. John Stuart Mill’s *Utilitarianism* (1861) distinguished higher pleasures, of the intellect and the moral feelings, from lower ones, and held that it is better to be a human being dissatisfied than a pig satisfied. Act utilitarians judge each act by its results; rule utilitarians judge acts by whether they follow rules whose general acceptance would produce the best results.

Three objections recur. A surgeon could kill one healthy visitor and use his organs to save five patients, and utilitarianism seems to require it. The theory may demand that people give away nearly everything they have (Module 10). And it treats society as if it were one person, trading one person’s suffering for another’s gain, which, John Rawls argued in 1971, ignores the separateness of persons. Rule utilitarians answer the first objection by noting that a rule permitting such killings would destroy trust in hospitals and so produce worse results overall.

9c. Duties and Rights: Kant

Immanuel Kant held in his *Groundwork of the Metaphysics of Morals* (1785) that nothing is good without qualification except a good will, the will to do one's duty because it is one's duty. The test of duty is the categorical imperative, a command that binds whatever one happens to want. In its first form it says to act only on a principle that one could will to be a universal law. Take a person who borrows money with a promise he means to break. If everyone made such promises, no one would believe a promise, so the principle undermines itself when made universal, and the act is forbidden.

In its second form the imperative says to treat humanity, in oneself and in others, always as an end and never merely as a means. Using a person as a mere tool, by deceiving or coercing her, ignores her capacity to choose her own ends. Kant drew a hard conclusion in a 1797 essay: one may not lie even to a murderer who asks where a friend is hiding. Most readers, Kantians among them, think he was wrong about the case. Christine Korsgaard argued in 1986 that the universal-law test, properly applied, permits lying to a deceiver bent on murder.

W. D. Ross proposed in 1930 that there are several basic duties, among them keeping promises, repairing harms, gratitude, justice, helping others, and not harming. Each binds unless outweighed by another in a particular case, and no formula ranks them. Robert Nozick described rights as side constraints (1974): limits on how one may treat people, not goals to be maximized, so one may not violate one person's rights even to prevent more violations by others.

The doctrine of double effect, which goes back to Thomas Aquinas's discussion of self-defense, draws a related line. It can be permissible to cause a harm as a foreseen side effect of pursuing a good end when it would be wrong to cause the same harm as a means to that end. On the doctrine, a doctor may give a dying patient enough morphine to relieve pain even if it hastens death, but may not give it in order to kill; a bomber may strike a munitions plant knowing civilians nearby will die, but may not target civilians to spread terror.

9d. Character: Aristotle, Confucius, and Mencius

Aristotle's *Nicomachean Ethics*, from the fourth century BC, asks not which act is right but what kind of person to be. A virtue is a settled disposition of character, usually a mean between two vices, as courage lies between cowardice and rashness and generosity between meanness and extravagance. Virtues are acquired by practice: we become just by doing just acts and brave by doing brave ones. Practical wisdom, the ability to see what a particular situation requires, is what lets a virtuous person hit the mean.

Confucius taught a virtue ethics of his own in the sayings collected as the *Analects*. Its central virtue is *ren*, humaneness, and its central practice *li*, ritual propriety, the forms of courtesy and ceremony through which respect is learned and expressed. Its ideal is the *junzi*, the exemplary person. Virtue begins in the family, with respect for parents, and extends outward; asked for one word to live by, Confucius named reciprocity: do not impose on others what you would not wish for yourself. Mencius built on this with his sprouts of virtue (Module 8).

Social psychology has challenged the idea of stable character. In a 1973 study by John Darley and Daniel Batson, seminary students on their way to give a talk, some of them on the parable of the Good Samaritan, passed a man slumped in a doorway. Of those told they were late, about 10% stopped to help; of those told they had time to spare, 63% did; the topic of the talk made little difference. The philosophers Gilbert Harman and John Doris concluded that robust character traits of the kind virtue ethics assumes may not exist. Virtue ethicists reply that the studies show how rare virtue is, which is why it is an ideal, and that practical wisdom includes noticing the pressures of a situation.

9e. Contracts, Care, and Ubuntu

T. M. Scanlon's *What We Owe to Each Other* (1998) set out contractualism: an act is wrong if it would be forbidden by every set of principles that no one could reasonably reject. Each person's complaint is weighed against each other person's, rather than all being added up. Scanlon's example is a man named Jones, trapped in the transmitter room of a television station during a World Cup broadcast and receiving painful electric shocks; to save him the broadcast must be stopped. However many millions are watching, Scanlon argued, we must save Jones, because no viewer's loss compares with his.

Care ethics, developed by Carol Gilligan in *In a Different Voice* (1982) and Nel Noddings (1984), starts from relationships of care and responsibility, between parent and child or nurse and patient, rather than from impartial rules. Gilligan argued that the dominant theory of moral development, Lawrence Kohlberg's, had ranked an ethic of justice above an ethic of care partly because its first studies were of boys.

Ubuntu, from southern Africa, is summed up in the Nguni saying *umuntu ngumuntu ngabantu*, a person is a person through other persons. On the philosopher Thaddeus Metz's account (2007), an act is right to the extent that it honors relationships of identity and solidarity with others, and wrong to the extent that it damages them. In one of its first judgments, *S v Makwanyane* (June 6, 1995), South Africa's Constitutional Court unanimously abolished the death penalty, with several of its judges invoking ubuntu among their grounds.

9f. Arjuna's Dilemma

The *Bhagavad Gita*, part of the Indian epic the *Mahābhārata*, opens on a battlefield. The warrior Arjuna sees his cousins, teachers, and kinsmen arrayed against him and refuses to fight, foreseeing a slaughter that would destroy families and the order of society. His charioteer Krishna, who in the poem reveals himself as God, answers that Arjuna must do his duty as a warrior in a just war, acting without attachment to the fruits of his action, since the outcome is not his to control.

Amartya Sen argued in *The Idea of Justice* (2009) that the exchange is a debate between two ways of moral thinking: Krishna's, which follows duty whatever the results, and Arjuna's, which attends to what will actually happen to real people. The Gita sides with Krishna, but Sen argued that Arjuna's concern for consequences, on a battlefield where tens of thousands would die, deserves more weight than the tradition has given it. The dilemma, some two thousand years older than the trolley problem (9g), likewise pits duty against results.

9g. The Trolley

Philippa Foot posed the problem in 1967. A runaway trolley will kill five workers on the track ahead unless the driver switches it to a side track, where it will kill one. Most people judge that switching is permissible, even required. Judith Jarvis Thomson added a variant in 1976 and 1985: the only way to stop the trolley is to push a large man off a footbridge into its path, killing him and saving the five. Most people judge that pushing is wrong, though the numbers are the same. In 2020, 63.4% of philosophers would switch the trolley and 13.3% would not, while 56.0% would not push the man and 22.0% would.

Double effect offers one explanation: the man on the side track dies as a side effect of saving the five, while the man on the bridge is used as a means of stopping the trolley. The psychologist and philosopher Joshua Greene offered another in 2001, from brain scans: pushing someone with one's own hands triggers fast emotional responses, while the willingness to sacrifice one for five draws on slower, deliberate reasoning.

Greene argues that the emotional responses track features, such as physical contact, that have no moral weight; his critics reply that the feelings may track a real distinction between harming and redirecting a harm.

9h. The Argument: Which Theory Is Right, or Do We Need All Three?

Each theory has a strong case. Consequentialists point out that results are what people care about, and that the theory's impartiality, each person counting for one, is the core of fairness. Deontologists point out that some acts, such as framing an innocent person to calm a mob, are wrong however good the results, and that rights protect individuals from being sacrificed. Virtue ethicists point out that moral life is mostly a matter of perception, habit, and judgment in particular situations, which no rule can fully capture.

Some philosophers think the theories converge. Derek Parfit argued in *On What Matters* that the best forms of Kantian ethics, contractualism, and consequentialism, properly developed, recommend the same acts, like climbers ascending the same mountain by different sides. Others, following Ross, hold that there are several basic moral considerations and no master principle, and that judgment must weigh them case by case, adjusting principles and verdicts against each other as 1a described.

Worked example — The Self-Driving Car's Crash Rule

The trolley problem became an engineering question when cars began to drive themselves, since a car's software must be written in advance for an accident it cannot avoid. The Moral Machine experiment, run by Edmond Awad and colleagues at the Massachusetts Institute of Technology and published in 2018, collected 40 million decisions in ten languages from millions of people in 233 countries and territories, each choosing whom a car with failed brakes should spare. The strongest preferences were to spare humans over animals, more lives over fewer, and the young over the old, with weaker preferences for sparing the young in some East Asian countries. In a 2016 study by Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan, 76% of participants said it was more moral for a car to sacrifice its one passenger than to kill ten pedestrians, yet people were much less willing to buy such a car. Germany's Ethics Commission on Automated and Connected Driving issued guidelines that went the other way on age in June 2017: in an unavoidable accident, any distinction among potential victims by personal features such as age, sex, or physical or mental condition is strictly forbidden, though programming a car in general to reduce the number of people injured may be justifiable. Now run the theories. A utilitarian counts lives, and perhaps years of life, which favors sparing the young, as the world's respondents did. A Kantian holds that each person has equal dignity and may not be weighed by age, as Germany's rule says, while allowing a general policy of harming fewer. A contractualist asks which rule no one could reasonably reject: a person who did not know her age might accept "harm as few as possible" and reject "sacrifice the old first." Virtue ethics, built for judgment in the moment, has the most trouble with a rule written years in advance. And the buyers in Bonnefon's study show a social dilemma: the rule each person thinks right for everyone is not the rule each would choose for her own car.

What each answer costs

Answer	What it explains	What it must accept or explain away
Consequentialism (Mozi, Bentham, Mill)	Why results matter, and why everyone's welfare counts equally	That rights may be overridden for a greater good, as in the transplant case, and a duty to give that seems to have no limit
Deontology (Kant, Ross, Nozick)	Rights, promises, and the wrongness of using people	That one may not violate a right even to prevent many worse violations, and that duties must somehow be ranked when they conflict
Virtue ethics (Aristotle, Confucius)	Moral education, character, and judgment in particular situations	Vague guidance in hard cases, and the evidence that situations sway behavior more than character does
Contractualism (Scanlon)	Why each person's standpoint counts, and why one person's great loss is not outweighed by many small gains	An account of what makes a rejection reasonable that does not assume the answer
Care ethics and Ubuntu	The moral weight of relationships and the duties they create	What is owed to strangers outside any relationship

The course's reading: no single theory has won, and each explains judgments the others struggle with. The pattern in the trolley cases, which most people and many philosophers share, is evidence that the difference between harming someone as a means and harming as a side effect carries moral weight, a difference that consequentialists must explain away.

Module 9 Key Points

Right and Wrong

• Consequentialism judges acts by results; Mozi in the fifth century BC and Bentham in 1789 both counted benefit to all.

• Kant forbids treating anyone merely as a means; his universal-law test rules out the lying promise.

• Virtue ethics asks what a person of good character would do; situationist studies suggest that situations sway behavior strongly.

• 63.4% of philosophers would switch the trolley, but 56.0% would not push the man, though both trade one life for five.

• The Moral Machine’s 40 million choices favored sparing the young; Germany’s 2017 rules forbid any distinction by age.

• Philosophers split three ways: virtue ethics 37.0%, deontology 32.1%, consequentialism 30.6%.

Module 9 Quiz

Check Your Understanding

1. Why does Kant's universal-law test rule out a lying promise?

A) Because lying causes harm
B) Because God forbids lying
C) Because it makes the liar unhappy
D) Because if everyone made promises they meant to break, no one would believe a promise, so the principle could not work as a universal law

2. What did Darley and Batson's Good Samaritan study find?

A) That seminarians always help
B) That being in a hurry cut helping sharply
C) That people who had prepared a talk on the Good Samaritan helped far more than those who had prepared a talk on other subjects
D) That religious people help strangers more often than others

3. In Bonnefon and colleagues' 2016 study of self-driving cars, what did people say?

A) That a car should sacrifice its passenger to save ten pedestrians, but they would rather buy one that protects its passengers
B) That cars should choose at random
C) That a car should never be programmed to choose between lives, so every crash decision should be left to the human at the wheel
D) That cars should always protect their passengers

4. What did Germany's 2017 ethics rules forbid a self-driving car to do in an unavoidable accident?

A) Reduce the total number of people injured in the crash, since numbers may not count
B) Record what happened in the crash for later review
C) Choose victims by age or sex
D) Brake when a child runs into the road

5. How does the doctrine of double effect explain the difference between the trolley switch and the footbridge?

A) Pushing kills more people
B) Switching tracks is legal
C) The man on the footbridge is a volunteer
D) The man on the side track dies as a side effect of saving the five, while the man pushed from the bridge is used as a means to stop the trolley

Module 10 — What We Owe: Strangers and the Future

Distance, Partiality, and People Not Yet Born

Recommended Reading / Reference

Book: Peter Singer, *The Life You Can Save* (2009; tenth-anniversary edition, 2019, free online), the book-length case for giving to the world's poorest people, with replies to the main objections

Reference: GiveWell, "How Much Does It Cost to Save a Life?" and its estimates for its top charities (updated September 2025), the most widely used figures for the cost of saving a life through charity

10a. The Drowning Child

Peter Singer wrote "Famine, Affluence, and Morality" in November 1971, as at least nine million refugees from East Bengal, now Bangladesh, faced starvation in camps in India. Britain had given £14.75 million in aid, he noted, while it had spent more than £275 million developing the Concorde airliner. The essay, published in 1972, made an argument that has shaped the debate ever since.

Suffering and death from lack of food, shelter, and medical care are bad. If it is in our power to prevent something bad from happening without sacrificing anything of comparable moral importance, we ought to do it. Someone who passes a shallow pond in which a small child is drowning ought to wade in and pull the child out: muddy clothes are insignificant beside a child's death. Neither the distance of the people in need nor the number of others who could help changes the duty, Singer argued, so people in rich countries ought to give much of what they now spend on comforts and luxuries to relieve suffering far away.

Singer also gave a weaker version: that we ought to prevent something very bad when we can do so without sacrificing anything morally significant. Even the weaker version, he argued, would transform the lives of most people in rich countries, who spend freely on things that are not morally significant while children die of preventable diseases.

10b. Kin and Strangers

The question of how far concern should reach divided Chinese philosophers more than two thousand years earlier. Mozi taught inclusive care: regard other people's families as your own and other states as your own. Partiality, he argued, is the root of harm, since a thief loves his own family and so robs others'. Mencius answered that Mozi's doctrine acknowledges no father: love rightly has degrees, beginning with one's parents and extending outward, and a person who loved a neighbor's child exactly as his brother's child would have lost something human.

Bernard Williams made a modern version of the point in 1976. Imagine a man who can save either his wife or a stranger from drowning. If he pauses to check that impartial morality permits him to save his wife, Williams said, he has had one thought too many: the right reason is simply that she is his wife. Partialists hold that special obligations to family, friends, and fellow citizens are part of morality, not a bias to be corrected. Impartialists reply that such obligations are justified because a world in which people care especially for their own goes better for everyone.

10c. How Much?

Singer's strong principle seems to require giving until one is nearly as badly off as the people one is helping, since until then a further gift costs less than it gains. Critics call this the demandingness objection. Susan Wolf argued in "Moral Saints" (1982) that a life devoted entirely to doing the most good would be unattractive, empty of the personal projects and excellences that make a life admirable. Liam Murphy proposed in 1993, and developed in a book of 2000, that each person owes a fair share: what she would have to give if everyone did their part, and no more when others fail to do theirs.

Practical standards sit between the extremes. Giving What We Can, founded by the philosophers Toby Ord and William MacAskill in 2009, asks members to pledge 10% of their income to effective charities; 11,377 people had taken the pledge by October 2026. Thomas Aquinas had held in the thirteenth century that what the rich have in superfluity belongs by natural right to the poor, a view that makes some giving a matter of justice rather than generosity.

10d. Effective Altruism

Effective altruism, a movement that took its name around 2011, asks how to do the most good with a given amount of money or time, using evidence and careful reasoning. Its best-known evaluator, GiveWell, founded in 2007, estimates the cost of saving a life through the charities it recommends. In its estimates updated in September 2025, about $5,500 saved a life through the Against Malaria Foundation's bed nets, $4,000 through Malaria Consortium's seasonal antimalarial medicine for children, $3,500 through Helen Keller Intl's vitamin A supplements, and $4,500 through New Incentives' payments to families for vaccinating their children.

GiveWell directed $427 million of donors' money from February 2025 to January 2026, which it expected to save about 86,000 lives. Critics raise several objections. Measuring what can be measured, they argue, favors bed nets over political and institutional change, which may do more good and resists counting. The economist Angus Deaton has argued that aid can weaken the accountability of governments to their own citizens. And critics warn that reasoning by expected value without limits can be used to justify reckless or harmful means in the name of a greater good, a concern that grew as parts of the movement turned toward speculative risks to the far future.

10e. A Test Case: The Aid Cuts of 2025

Governments, not only individuals, face the question. In 2025 the United States cut its foreign aid sharply: the secretary of state announced in March that 83% of the programs of the US Agency for International Development (USAID) had been canceled, and the agency closed on July 1, 2025. Britain announced in February 2025 that it would cut aid from 0.5% to 0.3% of national income by 2027 to pay for defense. By the preliminary figures of the Organisation for Economic Co-operation and Development (OECD), the rich countries' official development assistance fell 23.1% in real terms in 2025, to $174.3 billion, the largest drop it had recorded, with American aid down 56.9%.

The administration and its supporters argued that much aid was wasteful or did not serve American interests, that other wealthy countries should carry more of the burden, and that a government's first duties are to its own citizens. The economist William Easterly had argued in *The White Man's Burden* (2006) that decades of large aid programs did little for growth in poor countries. Critics answered that aid had been about 1% of federal spending and that the cuts would cost lives. A study in *The Lancet* led by Daniella

Cavalcanti (June 30, 2025) projected more than 14 million additional deaths by 2030 from the American cuts, 4.5 million of them children under five; a February 2026 study in *The Lancet Global Health* projected between 9.4 million and 22.6 million deaths from all donors' cuts, depending on their depth. These are models, not counts, and their assumptions are disputed.

Philosophers divide over what states owe foreigners. Thomas Nagel argued in 2005 that duties of justice hold only among people who share a state and its laws, and that beyond borders states owe only humanitarian aid against the worst suffering. Thomas Pogge argued in *World Poverty and Human Rights* (2002) that rich countries help keep the poor poor through the rules of trade, finance, and the arms and resource markets, so that aid is owed as a matter of justice, as compensation, and not as charity.

10f. Borders

Some argue that the strongest form of a duty to strangers is letting them in. Joseph Carens argued in 1987 that citizenship in a rich country is the modern equivalent of feudal privilege, an inherited status that greatly enhances one's chances in life, and that borders should therefore generally be open. Michael Huemer offered an analogy in 2010: blocking a starving man from walking to a market where he could buy food is not merely failing to help him but harming him.

Michael Walzer argued in *Spheres of Justice* (1983) that political communities have the right to shape their own membership, as clubs and families do, though they must take in some refugees. Christopher Heath Wellman (2008) grounded the right to exclude in freedom of association, and David Miller has argued that a shared national culture sustains the trust on which welfare states depend, while granting refugees strong claims. The economic stakes are large: some economists, among them Michael Clemens (2011), estimate that removing barriers to migration could raise world output by half or more, by moving workers to where they are more productive, though such estimates are disputed. Restrictionists cite other evidence: George Borjas has argued that immigration lowers the wages of the low-skilled native workers who compete with it, and Paul Collier (2013) that heavy emigration can strip small poor countries of their educated workers.

10g. People Not Yet Born

Duties to future people raise a puzzle that Derek Parfit called the non-identity problem (1984). A 14-year-old girl chooses to have a child now rather than wait several years, and the child has a worse start in life than a later child would have had. But a later child would have been a different person, so the actual child is not worse off than he would otherwise have been: otherwise he would not exist. If he has a life worth living, whom has she wronged? The same puzzle arises for policies that change who will be born a century from now, such as rapid depletion of resources.

Parfit found a second puzzle in comparing populations. Suppose the best outcome is the one with the most total welfare. A world of 10 billion people with very good lives, each at a level of 100, has a total of 1,000 billion; a world of 100 billion people with lives barely worth living, each at 11, has a total of 1,100 billion, so the second world is better. Parfit called this the Repugnant Conclusion and spent decades seeking a theory that avoids it without worse problems; counting average welfare instead implies that adding a happy person can make the world worse if his happiness is below the average.

Longtermists, among them William MacAskill in *What We Owe the Future* (2022), argue that future people count as much as present ones and that there may be vastly more of them, so reducing the risk of human extinction is among the most important things anyone can do. Toby Ord put the chance this century of an existential catastrophe, one that would end humanity or permanently ruin its future, at one in six in *The*

Precipice (2020), including one in ten from unaligned artificial intelligence. In a survey of 2,778 AI researchers conducted in October 2023, the median put the chance that AI causes human extinction or a similarly permanent and severe disempowerment of humanity at 5%. Critics answer that such numbers are guesses, and that decisions driven by small probabilities of vast outcomes can justify nearly anything.

Economists face the same question as a number. When they compare costs today with benefits decades from now, they discount the future, and the rate chosen is partly a moral choice: the philosopher and economist Frank Ramsey called discounting the welfare of later generations merely because they come later ethically indefensible (1928). The US government's social cost of carbon, the dollar value of the damage from emitting a ton of carbon dioxide, was set at $190 by the Environmental Protection Agency in 2023 using a 2% discount rate. An executive order of January 20, 2025 withdrew the estimates, saying the calculation was marked by logical deficiencies, a poor basis in empirical science, and politicization. In May 2025 the White House budget office told agencies not to put a dollar value on emissions unless a law requires it.

10h. The Argument: Is Giving to the Distant Poor a Duty or a Kindness?

Those who see a duty argue from the pond: if failing to save a nearby child at small cost is wrong, failing to save a distant child at the same cost is wrong too, unless distance itself matters morally, which few can defend. Pogge adds that the rich are implicated in the causes of poverty, and Aquinas and Kant, who counted helping others an imperfect duty with some latitude in how it is met, would agree that some giving is required.

Those who see a kindness argue that duties are limited to not harming others and to those close to us, and that giving beyond that is admirable but optional, what philosophers call supererogatory. They point out that a duty without limit would leave no room for personal projects, as Williams and Wolf argued, and that the pond is unlike global poverty: the child in the pond is one identifiable person in front of you, while poverty has causes no donor can remove and many others able to help.

Worked example — A Tenth of a Household's Income

The median American household income was $87,460 in 2025, according to the Census Bureau's report of September 15, 2026. A pledge of 10% would give $8,746 a year. At GiveWell's estimated cost of $5,500 to save a life through bed nets, that is $8,746 ÷ $5,500 = 1.6 lives a year; at $3,500 through vitamin A supplements, it is $8,746 ÷ $3,500 = 2.5 lives a year. Over a 40-year working life at today's costs, the total would be about 64 to 100 lives. Three cautions apply. The estimates are averages with wide uncertainty, as GiveWell itself stresses. Costs per life saved rise as the cheapest opportunities are funded and as diseases are brought under control, so the 40-year total is an upper bound on today's evidence. And a household at the median has its own needs, which is why the fair-share and partiality views in the table below give different answers about whether the pledge is a duty. What the arithmetic does show is the scale of Singer's comparison: on these figures the cost of saving a life, most often a young child's, is less than one month of the median household's income, which is $87,460 ÷ 12 = $7,288.

What each answer costs

Answer	What it explains	What it must accept or explain away
Singer’s strong duty	Why distance and the number of other possible helpers make no moral difference	That most people in rich countries act wrongly every day by not giving far more than they do
A fair share (Murphy)	A real duty to give that does not require sainthood	That one need not give more when others fail to give their share, even when children die as a result
Partiality (Mencius, Williams)	Special duties to family, friends, and fellow citizens	Why closeness justifies letting distant children die when saving them is cheap
Charity as optional	Why heroes of giving are admired while ordinary people are not condemned	Why walking past the pond is wrong while declining to give is permissible
Justice within states only (Nagel)	Why justice is owed to those who share laws and institutions	That states owe the world’s poor only minimal humanitarian aid
Longtermism (MacAskill, Ord)	Why the vast number of possible future people matters	Decisions driven by tiny probabilities of enormous outcomes

The course’s reading: the pond argument shows at least that anyone who denies a duty to help the distant poor, when helping is cheap, must point to some difference other than distance itself. How far any such duty reaches, a tenth of income or more, and how it weighs against duties to those close to us, is where reasonable views divide.

Module 10 Key Points

What We Owe: Strangers and the Future

• Singer argued in 1972 that distance makes no moral difference: failing to give is like walking past a drowning child.

• Mozi taught inclusive care for all; Mencius answered that love rightly has degrees, beginning with family.

• GiveWell estimated in 2025 that $3,500 to $5,500 saves a life through its top charities.

• Official aid fell 23.1% in 2025, the largest drop on record; a Lancet model projects millions of extra deaths by 2030.

• Parfit’s Repugnant Conclusion: on the total view, a vast population of barely worthwhile lives beats a smaller, flourishing one.

• A tenth of the median household income, $8,746, would fund about two lives saved a year at GiveWell’s estimates.

Module 10 Quiz

Check Your Understanding

1. What did Singer’s 1972 essay respond to?

A) The Vietnam War
B) A drowning in a London pond that had been widely reported in the British newspapers that year
C) A famine among East Bengal’s refugees
D) A debate over Britain’s health service

2. Why did Mencius reject Mozi’s inclusive care?

A) Because it would make people so generous to strangers that the state would lose its taxes
B) Because love has degrees
C) Because Mozi favored offensive war
D) Because only rulers should care for others

3. In the worked example, about how many lives a year could a 10% pledge on the median household income fund at GiveWell’s estimates?

A) About 1.6 to 2.5 lives
B) About 100
C) About 16 to 25 lives a year, since each life costs about $500
D) Less than 0.1

4. What is Parfit’s Repugnant Conclusion?

A) That population should shrink
B) That future people do not count
C) That a huge population of lives barely worth living could be better than a smaller, very happy one
D) That any act that changes who will be born wrongs the people who are born, because they might have had better parents

5. What did Joseph Carens argue in 1987?

A) That states may exclude anyone
B) That citizenship in a rich country is like a feudal birthright privilege, so borders should generally be open
C) That migration lowers wages
D) That only refugees may move

Module 11 — Life and Death

Death's Harm, Abortion, and Assisted Dying

Recommended Reading / Reference

Book: Shelly Kagan, *Death* (Yale University Press, 2012), drawn from his Yale lectures, on whether death is bad for the one who dies and what follows for how to live and when to die

Reference: Health Canada, *Sixth Annual Report on Medical Assistance in Dying in Canada, 2024* (November 28, 2025), the most detailed public record of a national assisted-dying system

11a. Why Is Death Bad?

Epicurus argued, in a letter written about 300 BC, that death is nothing to us. Good and bad consist in sensation, and death is the end of sensation: where we are, death is not, and where death is, we are not. So death cannot harm the one who dies, and fearing it is a mistake. The Roman poet Lucretius added the symmetry argument in *On the Nature of Things*, in the first century BC. The endless time before we were born is a mirror held up to the endless time after we die; no one is troubled by having not existed before birth, so no one should be troubled by not existing after death.

Thomas Nagel answered in an essay titled "Death" (1970) with the deprivation account. Death is bad for the one who dies, not because it is felt, but because it takes away the goods that the rest of a life would have contained, as a person can be harmed by a betrayal she never learns of. Against Lucretius, later writers noted an asymmetry: a person could not have been born much earlier, since a different sperm and egg would have made a different person, but she could have died much later, so only death deprives her of time she might have had.

Bernard Williams argued in "The Makropulos Case" (1973) that an endless life would eventually become unbearably tedious, since a fixed character would exhaust what it could care about. Death, on his view, can come too early, which is why it is usually bad, and it could also come too late. Philosophers were evenly split in 2020: 44.9% said they would choose to be immortal, and 41.3% said they would not.

11b. Abortion: The Question of Moral Status

Abortion divides people deeply, and the philosophical arguments turn on two questions: what moral status the embryo and fetus have, and what follows if they have full status. The course sets out each side's strongest case.

The case that abortion is in most cases seriously wrong has secular and religious forms. Don Marquis argued in "Why Abortion Is Immoral" (1989), using the deprivation account, that killing an adult is wrong mainly because it deprives the victim of a future like ours: the experiences, activities, and relationships the person would have had. A fetus has such a future, so killing it is wrong for the same reason, without any appeal to religion or to personhood. Defenders of the substance view, among them Robert George and Patrick Lee, add that the embryo is the same living human organism that will become the child and the adult, as animalism holds (Module 6), so a human being's life begins at conception. The Catholic Church teaches that human life must be protected absolutely from conception.

The case that abortion is often permissible also comes in two forms. Mary Anne Warren argued in 1973 that being a person, with the full right to life, requires capacities such as consciousness, reasoning,

self-motivated activity, communication, and awareness of oneself, and that an embryo or early fetus has none of them. Judith Jarvis Thomson argued in “A Defense of Abortion” (1971) that the conclusion follows even if the fetus is a person from conception, because a right to life is not a right to use another person’s body.

Thomson’s case is one of the most discussed thought experiments in applied ethics. You wake up to find yourself in a hospital bed, your circulatory system connected to that of a famous violinist with a fatal kidney ailment, who was plugged into you overnight by a society of music lovers because only your blood type can save him. If you unplug yourself, he dies; if you stay connected for nine months, he recovers. Thomson argued that you are not morally required to stay, however kind it would be, and that pregnancy resulting from rape is like the case. Critics answer that ordinary pregnancy differs, since it usually follows from a voluntary act, since parents have special duties to their children, and since abortion kills while unplugging lets die.

Many people hold a middle view on which moral status grows during pregnancy, so that late abortion is graver than early; L. W. Sumner, for example, proposed in 1981 that the capacity to feel marks the threshold, and researchers dispute when in pregnancy that capacity begins. Religious traditions differ too. Jewish law has traditionally put the mother’s life first and does not treat the fetus as a full person until birth. Islamic jurists traditionally linked ensoulment to the end of about 120 days of pregnancy, and several schools have permitted abortion before then for serious reasons.

11c. Abortion: The Law Now

In *Dobbs v. Jackson Women’s Health Organization* (June 24, 2022) the US Supreme Court overturned *Roe v. Wade* (1973) and returned abortion law to the states. As of September 30, 2026, 13 states banned abortion at all stages of pregnancy with limited exceptions, and four or five more banned it at about six weeks, the count depending on the outcome of a court fight over Wyoming’s law, by the tally of KFF, a health policy research group. Voters in four states will decide abortion measures on November 3, 2026. Idaho’s would create a statutory right to abortion until viability, the point at which a fetus could survive outside the womb; Missouri’s would repeal the abortion-rights amendment voters adopted in 2024 and ban gender-transition treatments for minors; Nevada’s is the required second vote on a constitutional right; and Virginia’s is a reproductive-freedom amendment.

Abortion pills are the next front. In May 2026 the Supreme Court paused a lower-court ruling that would have ended the prescribing of the abortion drug mifepristone by telemedicine and mail nationwide while the appeal proceeds, and the Food and Drug Administration (FDA) told a court in October 2026 that it would finish a safety review of the drug by March 2027. In Gallup’s survey of May 2026, 53% of Americans called themselves pro-choice and 42% pro-life. Among philosophers in 2020, 81.7% judged first-trimester abortion permissible and 13.1% impermissible, in a profession that leans left (front table).

11d. Assisted Dying: The Arguments

The terms vary. In euthanasia a doctor administers a lethal drug; in assisted dying, also called medical aid in dying or assisted suicide, a doctor prescribes it and the patient takes it. The argument for allowing either rests on three claims. Competent adults have the right to decide how their own lives end. Mercy requires relieving suffering that palliative care cannot reach. And the line between killing and letting die, which already lets patients refuse treatment and die, carries no moral weight by itself.

James Rachels pressed the last point in 1975 with two cases. Smith will inherit a fortune if his young cousin dies, so he drowns the boy in the bath. Jones, in the same position, comes in to drown his cousin, sees the

boy slip and hit his head, and stands by as he drowns. Jones did not kill, but he is no less to blame, so the bare difference between killing and letting die does not matter. If letting a suffering patient die is permitted, Rachels concluded, a quicker and more humane death can be too.

The argument against allowing assisted dying also rests on three claims. Life has a value its possessor may not destroy, a view held by many religious traditions and, in a secular form, by Kant, who held that ending one's life to escape pain uses oneself merely as a means. Doctors are healers, and the Hippocratic oath had them swear not to give a deadly drug even if asked. And legal assisted dying puts pressure on the vulnerable, the disabled, the elderly, and the poor, who may come to feel a burden; disability-rights groups such as Not Dead Yet oppose it on these grounds. Opponents add a slippery slope: once autonomy and mercy justify assisted death, there is no principled reason to keep it to the terminally ill, and in practice eligibility widens. Against Rachels they answer that withdrawing treatment and killing differ in intention, the line that double effect draws (Module 9), and in what causes the death, since the disease kills the patient whose treatment is withdrawn.

11e. Assisted Dying: The Laws Now

Medical aid in dying was legal in 14 American jurisdictions as of September 12, 2026. Oregon was first, in 1997, followed by Washington and Montana (by a court ruling) in 2009, Vermont (2013), California and Colorado (2016), and the District of Columbia (2017). Hawaii, New Jersey, and Maine followed in 2019, New Mexico in 2021, and three more in 2026: Delaware (in force January 1), New York (August 5), and Illinois (September 12). Abroad, the Netherlands and Belgium have allowed euthanasia since 2002, Canada since 2016, Spain and New Zealand since 2021, and all six Australian states since 2023.

France's law creating an "aid in dying," adopted by the National Assembly 291–241 on July 15, 2026 and upheld, with three interpretive reservations, by the Constitutional Council, was promulgated on August 18 and took effect on August 20, with its implementing decrees still to come. In Britain the House of Commons passed a bill for England and Wales in June 2025, but it ran out of time in the House of Lords in April 2026; a reintroduced bill was defeated in the Commons 286–270 at second reading on September 11, 2026, and Scotland's parliament rejected its own bill 69–57 on March 17, 2026.

Canada has gone furthest in widening eligibility. It allowed medical assistance in dying (MAID) for people whose natural death was reasonably foreseeable in 2016, and in 2021 for people whose death is not foreseeable, a category called Track 2. An extension to people whose only condition is a mental illness was scheduled for March 17, 2027, but on October 7, 2026 the justice minister, Sean Fraser, said the government would not proceed with it and promised a bill that would also allow requests made in advance.

11f. The Argument: Should Doctors Help Patients Die?

The evidence from places where it is legal is read by each side as support. In Oregon in 2025, 400 people died after taking prescribed medication, about 1% of the state's deaths; the concerns most often cited were loss of autonomy (89%), being less able to enjoy life (89%), and loss of dignity (65%), while being a burden on others was cited by 39.8% and inadequate pain control, or worry about it, by 37%. In the Netherlands in 2025, 10,341 deaths were reported, 6.0% of all deaths, among them 174 people with psychiatric conditions and 499 with dementia. Supporters see choice exercised by dying people for reasons of dignity and control; opponents see a rising share of deaths and a list of reasons in which feeling a burden appears.

Worked example — Canada's Numbers Since 2016

Canada's official reports allow the slippery-slope claim to be tested against numbers. Medically assisted deaths numbered 1,018 in the half-year after legalization in 2016 and 2,838 in 2017; by Health Canada's latest, revised figures they reached 5,461 in 2019, 10,066 in 2021, 13,199 in 2022, 15,427 in 2023, and 16,499 in 2024, for a total of 76,475 since 2016. As a share of all deaths, that is a rise from 1.07% in the second half of 2017 to 5.1% in 2024. Track 2 deaths, of people whose natural death was not foreseeable, numbered 732 in 2024, or 4.4% of all assisted deaths. The median age was 78.0 in Track 1 and 75.9 in Track 2, and most cases in Track 1 involved cancer (63.6%). In both tracks nearly all cited the loss of ability to do meaningful activities (95.1% in Track 1 and 97.5% in Track 2), and about half cited being a burden on family, friends, or caregivers (48.4% and 50.3%), but isolation or loneliness was cited by 21.9% in Track 1 and 44.7% in Track 2. Supporters read the figures as showing that assisted death remains mostly a choice made by elderly people with cancer near the end of life, with safeguards that work. Opponents read the same figures as showing a legal slope already descended, from terminal illness to non-terminal conditions and nearly to mental illness, and a practice in which loneliness figures in more than two of every five non-terminal cases. The numbers settle how fast and how far the practice has spread; whether that spread is a slide or a reasonable extension depends on the values the arguments in 11d dispute.

What each answer costs

Answer	What it explains	What it must accept or explain away
Personhood from conception (George, Catholic teaching)	The continuity of one human life from its beginning	That an early embryo can still split into twins, and that many embryos die naturally before anyone knows they exist
A future like ours (Marquis)	Why killing is wrong, without appeal to religion or personhood	Why preventing conception does not deprive a future in the same way, and the challenge from views on which status grows
Personhood criteria (Warren)	Why an embryo lacks the rights of a person	The status of newborns, who also lack most of the criteria
Bodily rights (Thomson)	Why abortion can be permissible even if the fetus is a person	Why ordinary pregnancy, usually following a voluntary act, resembles the violinist case
Autonomy and mercy (for assisted dying)	The case for letting competent, suffering people choose how to die	The pressures on the vulnerable, and why eligibility should stop where it does
Sanctity of life (against assisted dying)	The opposition of many doctors, religious traditions, and disability groups	Why refusing treatment and dying is permitted while a quicker death is not

Abortion and assisted dying are among the questions on which the course takes no side. Where the argument stands: on abortion, the facts of development are largely agreed, and where they are disputed, as over when a fetus can first feel, the answer matters most on views that make feeling the threshold. What divides the sides is which account of moral status is right, and whether a right to life would include a right to the use of another's body. On assisted dying, the records from places where it is legal settle questions of fact, as the worked example shows, but not the questions of value on which the verdict turns.

Module 11 Key Points

Life and Death

• Epicurus held that death cannot harm us, since we no longer exist; Nagel answered that death deprives us of life’s goods.

• Marquis argued that abortion is wrong because it deprives a fetus of a future like ours; Thomson argued that bodily rights can permit it anyway.

• Since Dobbs in 2022, 13 US states ban abortion at all stages; four states vote on abortion measures on November 3, 2026.

• Medical aid in dying is legal in 14 US jurisdictions; France legalized it in August 2026, and the Commons rejected it in September.

• In Canada 16,499 people, 5.1% of all deaths, died with medical assistance in 2024, 4.4% of them not near death.

• Philosophers judge first-trimester abortion permissible by 81.7% to 13.1%; Americans divide 53% pro-choice to 42% pro-life (Gallup, 2026).

Module 11 Quiz

Check Your Understanding

1. What is Lucretius’s symmetry argument?

A) Death is painful
B) The soul survives death and is reborn in another body
C) Death is worse the earlier in life it comes, since more is lost
D) Nonexistence after death mirrors that before birth

2. On Don Marquis’s argument, what makes killing wrong?

A) Depriving the victim of a valuable future
B) The pain it causes the one who is killed and those who loved them
C) That it breaks a command found in every religious tradition
D) That it reduces the total happiness in the world below what it would otherwise be

3. What does Judith Jarvis Thomson’s violinist case try to show?

A) That a fetus is not a person, because it lacks consciousness and reason at every stage of pregnancy
B) That killing is always worse than letting die, whatever the circumstances
C) That a right to life is not a right to use another’s body
D) That people who need organs have a right to them from strangers

4. What share of all deaths in Canada were medically assisted in 2024?

A) They were under 0.5%.
B) They were about 1%.
C) They were about 12%, double the share in the Netherlands.
D) They were 5.1%, or 16,499 deaths.

5. What happened to assisted dying in England and Wales in 2026?

A) It became legal in April
B) A bill passed by the Commons in 2025 ran out of time in the Lords, and a reintroduced bill was defeated at second reading on September 11, 2026
C) Parliament passed it unanimously
D) The courts legalized it for all adults

Module 12 — Justice: Who Should Get What

Equality, Liberty, and Desert

Recommended Reading / Reference

Book: Michael J. Sandel, *Justice: What's the Right Thing to Do?* (Farrar, Straus and Giroux, 2009), drawn from his Harvard lectures, which tests the utilitarian, libertarian, and egalitarian answers on hard cases

Reference: Julian Lamont and Christi Favor, "Distributive Justice," *Stanford Encyclopedia of Philosophy* (first published 1996), on the main theories of who should get what and the arguments among them

12a. Proportion to What?

Aristotle held in Book V of the *Nicomachean Ethics* that justice in distribution is a kind of proportion: equals should receive equal shares, and unequals unequal ones, according to merit. Everyone accepts the formula, he observed, but people disagree about what merit is: democrats say free birth, oligarchs wealth or noble birth, and aristocrats virtue. In the *Politics* he gave an example: the best flutes should go to the best flute players, since playing well is what flutes are for.

Every later theory can be read as a way of filling in Aristotle's blank. Some answer with need, some with effort or contribution, some with what people have freely agreed to, and some deny that a fair share can be read off any pattern at all, since what matters is how holdings came about. The modern debate took its shape from one book.

12b. Rawls: The Veil of Ignorance

John Rawls asked in *A Theory of Justice* (1971) which principles free and equal people would choose to govern their society if they did not know what place they would hold in it: their class, talents, sex, race, religion, or idea of a good life. Behind this veil of ignorance, in what he called the original position, no one could tailor the rules to favor himself, so whatever was chosen would be fair.

Rawls argued that the parties would choose two principles. First, each person is to have an equal right to the most extensive basic liberties compatible with the same liberties for everyone, among them freedom of conscience, speech, and assembly, the vote, and the protections of the rule of law. Second, social and economic inequalities are allowed only if the positions they attach to are open to all under fair equality of opportunity, and only if they work to the greatest benefit of the least advantaged.

The last condition is the difference principle. An inequality is just only if the worst-off are better off with it than they would be without it, as they might be if higher pay draws more people into medicine and so gives the poor better care. Rawls argued that the parties would choose this cautious rule, called maximin because it maximizes the minimum, since the stakes were their whole lives, they had no basis for estimating their chances, and the worst outcomes would be unacceptable.

The economist John Harsanyi had argued in 1953 and 1955 that a rational person who gave himself an equal chance of being anyone in society would instead maximize average welfare, which is the utilitarian answer. Others objected that Rawls treated natural talents as morally arbitrary, a matter of luck that no one deserves, when talents are part of who people are, and, from the other side, that the difference principle permits vast inequalities as long as the worst-off gain from them.

12c. Nozick: Liberty Upsets Patterns

Robert Nozick answered Rawls in *Anarchy, State, and Utopia* (1974) with what he called the entitlement theory: a distribution is just if it arose from just acquisition and just transfers, whatever pattern results, with rectification of past injustices. Justice is historical: it depends on how holdings came about, not on how they compare. Theories that judge a distribution by its shape, whether equal or ordered by need or merit, Nozick called patterned, and he argued that liberty upsets every pattern.

His example begins with whatever distribution you think most just. A million people then each freely pay 25 cents to watch the basketball star Wilt Chamberlain play, dropping the money in a box with his name on it, and at the end of the season he has $250,000, far more than anyone else. If the starting distribution was just and every transfer voluntary, Nozick asked, how could the result be unjust? Keeping any pattern in place would require continual interference with what people choose to do with their shares.

Nozick concluded that taxing earnings to fund redistribution is on a par with forced labor, since it takes the product of some hours of a person’s work, and that only a minimal state, which protects against force, theft, and fraud and enforces contracts, can be justified. Critics answer that the starting point matters. Few actual holdings trace back to just acquisition, so Nozick’s own principle of rectification might require sweeping redistribution. And his condition on acquisition, a version of John Locke’s proviso that taking land or resources must not leave others worse off, is hard to meet once everything is owned.

12d. Welfare, Luck, and Enough

Utilitarians reach toward equality by another road. A dollar matters more to a poor person than to a rich one, which economists call diminishing marginal utility, so moving money from rich to poor generally raises total welfare. Pushed to its limit, the argument would equalize incomes, but taxes and transfers weaken the incentives to work and invest, which shrinks what there is to share. The economist James Mirrlees showed in 1971 how to set an income tax that balances the two, work recognized by a Nobel prize in 1996. On this view inequality is not unjust in itself; it matters only through its effects on welfare.

Luck egalitarians, among them Ronald Dworkin, Richard Arneson, and G. A. Cohen, took Rawls’s point about luck and made it the center of a theory. Inequalities that come from brute luck, such as being born poor or disabled, are unjust, they hold, while those that come from people’s own choices can be fair, including the results of what Dworkin called option luck, a gamble knowingly taken.

Elizabeth Anderson replied in “What Is the Point of Equality?” (1999) that the theory is both too harsh and insulting. It would abandon an uninsured driver injured in an accident he caused, since his loss came from his own choice, and it would compensate people for being born less talented or less attractive, which treats them with pity. The point of equality, she argued, is to end oppression and let people stand as equals in a democratic society. That requires that everyone have what is needed to function as a citizen, not that luck be canceled.

Two other views move the concern from inequality to deprivation. Harry Frankfurt argued in 1987 that what matters morally is that everyone have enough, not that everyone have the same, a view called sufficientarianism. Derek Parfit argued in 1991 that a benefit matters more the worse off its recipient is, whatever others have, a view called prioritarianism. Parfit’s case against valuing equality for its own sake is the leveling-down objection: if equality were good in itself, making the better-off worse off without helping anyone would be an improvement in one respect, and few believe that it is.

12e. Capabilities, and Marx

Amartya Sen asked in a 1979 lecture, “Equality of What?” Equal incomes can leave people very unequal in what they can actually do: a person who uses a wheelchair needs more resources to get around than one who walks. What justice should look to, Sen argued, are capabilities, people’s real freedoms to be and do what they have reason to value, such as being well nourished, moving about, and taking part in the life of the community. The United Nations’ Human Development Index, launched in 1990 with Sen’s help, ranks countries by health and education as well as income.

Martha Nussbaum (2000, 2011) proposed a list of ten central capabilities that every just constitution should guarantee to each citizen up to a threshold, among them life, bodily health, bodily integrity, practical reason, affiliation with others, play, and control over one’s political and material environment. Critics answer that any such list imposes one idea of a good life; Nussbaum replies that her list protects capabilities, the freedom to function in these ways, without requiring anyone to use them.

Karl Marx held that distribution cannot be separated from production. In *Capital* (1867) he argued that workers, owning nothing but their ability to work, must sell it to the owners of factories and land, and that they are exploited: they produce more value than they are paid, and the owners keep the difference. In the *Critique of the Gotha Programme* (1875) he looked forward to a society that would write on its banners: from each according to his ability, to each according to his needs. That would be possible, he thought, only once production had become abundant.

States that claimed his name in the twentieth century, among them the Soviet Union, China, and Cambodia, abolished most private ownership of the means of production, and under several of them tens of millions died in famines, purges, and camps. Marx’s defenders deny that his ideas required such outcomes; his critics argue that central planning and one-party rule followed from abolishing markets. G. A. Cohen defended the socialist ideal in *Why Not Socialism?* (2009) with a camping trip, on which friends share equipment and food without prices or markets, and asked whether the same spirit could be made to work at the scale of a society.

12f. What People Choose Behind a Veil

Experiments have put real people behind something like Rawls’s veil. Norman Frohlich and Joe Oppenheimer asked groups of students in Canada, the United States, and Poland to agree, usually unanimously, on a rule for distributing incomes before learning which income each would receive. Of the 81 groups that reached a decision, as reported in 1992, roughly three in four chose to maximize the average income subject to a floor guaranteeing a minimum, 10 chose to maximize the average alone, and only one chose Rawls’s rule of maximizing the position of the worst-off.

In a Norwegian experiment reported in 2007 by Alexander Cappelen and colleagues, 96 students who had produced money in a task decided how to share it with a partner. Their choices sorted them into three types: 43.5% strict egalitarians, who split the money equally whatever its source; 38.1% liberal egalitarians, who rewarded what each had chosen to invest but not the luck of the return; and 18.4% libertarians, who let each keep what he had produced.

In a survey of 5,522 Americans reported in 2011 by Michael Norton and Dan Ariely, respondents guessed that the richest fifth held about 59% of the country’s wealth, when the true share was about 84%, and said the ideal would be 32%. Shown unlabeled charts, 92% preferred a more equal distribution, which the authors based on Sweden’s, to one showing America’s. By the Federal Reserve’s figures, the richest 1% of

American households held 32.5% of household wealth in the second quarter of 2026, the highest share in a series that begins in 1989, and the bottom half held 2.3%.

12g. The Argument: Is Inequality Unjust in Itself, or Only Poverty?

Those who hold that inequality is unjust in itself argue that it does damage beyond deprivation. Large gaps in wealth become gaps in power, as the rich buy influence over politics and the poor lose the standing to be heard, which is Anderson's concern. Gaps that come from the circumstances of birth are unfair whatever their size, on luck egalitarian grounds. Rawls adds that inequalities must be justified to those who have least, and Cohen that a society of equals expresses a kind of community that a merely comfortable one lacks.

Those who hold that only poverty is unjust argue that what matters is how people fare, not how they compare. They cite the leveling-down objection (12d), and Frankfurt adds that attending to what others have distracts people from what they themselves need. On these views the urgent fact is not the share of the richest 1% but that, by the World Bank's estimate, 837 million people lived on less than $3 a day in 2026, down from about 2.3 billion in 1990. Libertarians go further and hold that neither gap nor poverty is unjust in itself when it arises from free exchange among people who acquired their holdings justly.

Philosophers lean toward equality. In 2020, 44.0% favored egalitarianism in political philosophy, 27.3% communitarianism, which stresses the claims of community and shared traditions, and 13.4% libertarianism in Nozick's sense; asked about economic systems, 53.0% favored socialism and 29.5% capitalism. The front table's caution applies with special force here, where the profession's politics and its philosophy meet.

Worked example — Three Societies, Ranked by Each Theory

Take three societies, each with five groups of equal size, with incomes in thousands of dollars a year. In Society A every group has 30, for an average of 30. In Society B the groups have 10, 40, 60, 90, and 200, an average of 80. In Society C they have 35, 45, 60, 80, and 130, an average of 70. Now run the theories. A utilitarian who counts dollars prefers B, which has the most. But if welfare rises with the logarithm of income, so that each doubling of income adds the same amount of welfare, C comes first: summing the logarithms of the five incomes gives about 20.7 for C, 19.9 for B, and 17.0 for A. Prioritarians, who give gains to the worse-off still more weight, rank C first as well. Rawls's difference principle looks at the worst-off group, which has 35 in C, 30 in A, and 10 in B, so it ranks C first: C's inequalities are justified because its poorest are better off than they would be under equality. Strict equality ranks A first, but choosing A over C would make every group poorer, the poorest included, for the sake of equality alone, which is the leveling-down objection made concrete. A sufficientarian with a threshold of 25 accepts A and C, rejects B, whose poorest fifth falls below it, and finds nothing to choose between A and C. Two theories cannot rank the table at all. Nozick's needs to know how the incomes came about: if B arose by free exchange from a just start, B is just, and moving to C by taxation would violate rights. A luck egalitarian needs to know whether B's poorest fifth are poor by choice or by circumstance. The table shows where the theories part. Of those that can rank it, all but the dollar-counting utilitarian prefer C to B, but they divide over whether A's equality is worth making everyone poorer, and the historical theories over whether the pattern matters at all.

What each answer costs

Answer	What it explains	What it must accept or explain away
Rawls's difference principle	Why fairness means choosing rules without knowing one's place, and why inequalities must benefit the worst-off	A cautious rule of choice that Harsanyi and many economists reject, and vast inequalities whenever they help the worst-off even slightly
Entitlement (Nozick)	Why free exchange from just holdings seems to wrong no one, and the moral weight of what people earn	That no degree of poverty is unjust if it arose without force or fraud, and holdings whose history is rarely just
Utilitarian redistribution	Why money should move toward those it helps most, and why incentives limit how far	That the poor's claims count only through total welfare, so a large enough gain for the rich could outweigh them
Luck egalitarianism (Dworkin, Cohen)	Why people may keep what their choices earn but not what luck hands them	Abandoning those ruined by their own bad choices, and the work of separating choice from luck
Sufficiency and priority (Frankfurt, Parfit)	Why deprivation seems more urgent than inequality, and why leveling down is no gain	Indifference to gaps above the threshold, however vast, and a threshold that is hard to draw
Capabilities (Sen, Nussbaum)	Why equal incomes can leave people unequal, as with disability	A list of capabilities that critics say builds in one idea of a good life
Marx	How ownership of the means of production shapes every distribution	The record of the states that claimed his name, and the lack of a tested way to run a large economy without markets

The course's reading: the leveling-down objection (12d) shows that strict equality cannot be the whole of justice. Most of the theories here treat deep poverty in a rich society as a matter of justice rather than charity, though not Nozick's, nor luck egalitarianism where poverty results from choice; and in the experiments most groups chose a guaranteed floor when they could not know their place. Whether inequality above a decent floor is unjust in itself is where the theories, and the people in the experiments, divide, and no evidence settles it.

Module 12 Key Points

Justice: Who Should Get What

• Aristotle held that justice gives equals equal shares by merit; later theories fill in what merit is: need, effort, agreement, or history.

• Rawls’s veil of ignorance yields equal basic liberties and the difference principle: inequalities are just only if they benefit the worst-off.

• Nozick argued that free transfers from just holdings, like fans paying Wilt Chamberlain, cannot make a distribution unjust, whatever pattern results.

• Parfit’s leveling-down objection shows that equality cannot be the whole of justice; Frankfurt held that what matters is that everyone has enough.

• Behind a veil in experiments, about three in four groups chose maximum average income with a guaranteed floor; one chose Rawls’s rule.

• America’s richest 1% held a record 32.5% of household wealth in mid-2026; 837 million people worldwide lived on under $3 a day.

Module 12 Quiz

Check Your Understanding

1. What did Nozick’s Wilt Chamberlain example aim to show?

A) That free exchanges upset any pattern

B) That stars deserve less

C) That a just society would tax the star’s earnings to help the poor, since his talent is a matter of luck he did nothing to earn

D) That fans should not pay to watch sports

2. Why did Rawls think the parties behind the veil would choose to protect the worst-off position?

A) Because when the stakes are a whole life and the odds cannot be known, it is rational to make the worst outcome as good as possible.

B) Because they know they will be poor.

C) Because they care more for others than for themselves.

D) Because the majority would vote for it.

3. What is the leveling-down objection to valuing equality for its own sake?

A) That equality is impossible

B) That making the better-off worse off while helping no one would count as an improvement

C) That the poor would refuse equal shares if the rich could earn more and pay higher taxes to fund schools and hospitals for everyone

D) That it weakens incentives to work

4. In Frohlich and Oppenheimer’s experiments, what did most groups choose?

A) Maximizing the worst-off position

B) Strictly equal incomes for all

C) Average income with a floor

D) Maximizing average income alone

5. What did Amartya Sen argue justice should look to?

A) Incomes

B) Happiness, measured by surveys of how satisfied people say they are with their lives, adjusted for differences in income

C) The resources each person owns, whatever they can do with them

D) Capabilities: real freedoms to be and do

Module 13 — The State: Authority, Democracy, and Liberty

Why Obey, Who Should Rule, and Where Power Must Stop

Recommended Reading / Reference

Book: Jonathan Wolff, *An Introduction to Political Philosophy* (Oxford University Press, third edition, 2016), a short, even-handed account of the state of nature, the case for the state, democracy, liberty, and justice

Reference: Fabienne Peter, “Political Legitimacy,” *Stanford Encyclopedia of Philosophy* (first published 2010), on what makes a state’s power legitimate and whether citizens must obey it

13a. Socrates in Prison

In 399 BC an Athenian jury condemned Socrates to death for impiety and for corrupting the young. In Plato’s *Crito*, his friend Crito visits him in prison and offers to bribe the guards and smuggle him into exile. Socrates refuses. He imagines the Laws of Athens speaking to him: they made his birth, upbringing, and education possible, and he stayed in the city for seventy years when he was free to leave, so he agreed by his conduct to obey them. To escape would be to try to destroy them, and an unjust verdict does not release him from the agreement. He drank the hemlock.

The *Crito* contains nearly every argument for a duty to obey that came after it: consent, gratitude for benefits received, and the harm that disobedience does to a system of law. Its question, why anyone should obey a state even when it errs, is the question of political authority.

13b. The Social Contract

Thomas Hobbes argued in *Leviathan* (1651), written as the English Civil Wars were ending, that without a government people would live in a war of every man against every man, in which life would be “solitary, poor, nasty, brutish, and short.” Rational people would therefore agree to hand their power to a sovereign who could keep the peace. They could resist him only to save their own lives, since anything more would risk a return to the state of nature.

John Locke’s *Second Treatise of Government* (1689) gave people natural rights to life, liberty, and property before any government existed. They consent to a government to protect those rights, and a government that violates them breaks its trust and may be resisted and replaced, an argument the American Declaration of Independence echoed in 1776. Since few people ever explicitly consent, Locke held that anyone who owns property, or simply travels on the highway, within a state gives tacit consent to its laws.

Jean-Jacques Rousseau began *The Social Contract* (1762) with the line “Man is born free, and everywhere he is in chains.” A legitimate state, he argued, is one in which each person, by uniting with all, still obeys only himself, because the laws express the general will: the will of the people as a whole for the common good, not the sum of private interests. Anyone who refuses to obey the general will, he added, will be forced to be free, a phrase his critics have read as a license for tyranny.

David Hume answered the contract theorists in “Of the Original Contract” (1748). Most governments began in conquest or usurpation, not agreement, and few subjects have ever been asked. As for tacit consent, a poor peasant who knows no foreign language and lives on small wages has no real choice to leave. We may

as well say, Hume wrote, that a man consents to the authority of a ship’s master though he was carried on board while asleep and must leap into the ocean and perish the moment he leaves her.

13c. Heaven’s Mandate

Chinese thought grounded authority differently. The Zhou kings, who overthrew the Shang dynasty around 1046 BC, justified the conquest by the Mandate of Heaven: Heaven grants the right to rule to a virtuous ruler and withdraws it from a cruel one, and the people’s suffering and revolt are signs that it has been withdrawn. Asked by King Xuan of Qi whether a minister may kill his sovereign, Mencius answered that a ruler who violates humaneness and rightness is a mere fellow. He had heard of the punishment of the mere fellow Zhou, the last Shang king, he said, but not of the killing of a sovereign.

Mencius also held that the people are the most important element in a state, the altars of the land and grain next, and the ruler least, and he quoted an ancient text: Heaven sees as the people see and hears as the people hear. His view is not democratic, since the people do not choose their ruler, but it makes the people’s welfare the test of a government’s right to rule, as Locke’s did some two thousand years later.

13d. Is There a Duty to Obey?

If no actual contract binds most citizens, perhaps nothing does. Robert Paul Wolff argued in *In Defense of Anarchism* (1970) that each person’s duty to decide for himself what is right conflicts with any duty to obey commands because they are commands. A. John Simmons concluded in 1979 that no existing theory shows that most citizens of any state have a general obligation to obey its laws, a position called philosophical anarchism. It does not call for abolishing the state, only for denying that its commands bind in themselves. Michael Huemer argued in 2013 that states do things, such as taxing and imprisoning, that would be wrong for anyone else to do.

Defenders of a duty offer three main grounds. H. L. A. Hart (1955) and Rawls (1964) argued from fair play: those who accept the benefits of a cooperative scheme, such as public order, owe it their share of the burdens. Rawls later argued for a natural duty to support just institutions, which binds whether or not one has consented. And Ronald Dworkin argued in 1986 that citizens, like members of a family, have obligations that come from belonging to a community whose members treat one another with equal concern.

13e. Who Should Rule?

Plato compared democracy to a ship whose crew quarrel over who should steer, though none has learned navigation, and who flatter and drug the owner to get the helm, while the true navigator, who studies the seasons, the sky, and the stars, is dismissed as a stargazer. In the *Republic* he proposed that philosophers, trained for decades, should rule. Even John Stuart Mill, a champion of representative government, proposed in 1861 that the educated should cast more votes than others, a scheme called plural voting.

Jason Brennan revived the case in *Against Democracy* (2016). Most voters, he argued, are hobbits, uninformed and uninterested, or hooligans, who follow politics like sports fans and read evidence in their team’s favor; few are vulcans who reason impartially. Since bad votes impose results on everyone, he proposed epistocracy, rule by the knowledgeable, for instance by giving extra votes to people who pass a test of political knowledge. Surveys show how little many citizens know: in the Annenberg Public Policy Center’s survey of August 2026, only 65% of American adults could name all three branches of government.

David Estlund answered in *Democratic Authority* (2008) that any test of who is knowledgeable enough to rule would be open to reasonable dispute, so imposing it on those who reject it would be illegitimate, and the test would likely favor groups already privileged. Hélène Landemore argued in *Democratic Reason* (2013) that a large, diverse group can outperform a small group of experts, since its members' errors and blind spots differ and partly cancel. She drew on a 2004 result by Lu Hong and Scott Page that groups of diverse problem solvers can outperform groups of the individually most able.

The record also bears on the question. Amartya Sen observed in 1999 that no famine has ever occurred in a functioning democracy with regular elections, opposition parties, and a relatively free press, since rulers who must face voters and critics cannot ignore starvation, and democracies have rarely, if ever, fought wars against each other. Yet by the count of the Varieties of Democracy (V-Dem) Institute at the University of Gothenburg (March 2026), 74% of the world's people, about 6 billion, lived in autocracies in 2025, and the level of democracy enjoyed by the average person had fallen back to that of 1978. In Pew's survey of 16 high-income democracies in spring 2026, a median of 54% were dissatisfied with how democracy was working in their country.

13f. Liberty: Where Power Must Stop

John Stuart Mill's *On Liberty* (1859) set out the harm principle: the only purpose for which power can rightfully be exercised over any member of a civilized community, against his will, is to prevent harm to others. A person's own good, physical or moral, is not a sufficient warrant. Over himself, over his own body and mind, Mill wrote, the individual is sovereign.

Mill allowed one kind of intervention for a person's own sake. If someone is about to cross a bridge known to be unsafe and there is no time to warn him, an official may seize him and turn him back, since no one wants to fall into the river; but once warned, he must be left to decide. Restricting a person's choices for her own good is called paternalism. Soft paternalism guards against ignorance or impulse, with warning labels and waiting periods; hard paternalism overrides informed choices, as bans on selling drugs do.

Recent laws test the line. Australia barred children under 16 from holding accounts on social media platforms from December 10, 2025, and its eSafety Commissioner reported that 4.7 million accounts had been deactivated, removed, or restricted in the first month. The Maldives banned the sale and use of tobacco by anyone born on or after January 1, 2007, from November 1, 2025. Britain's Tobacco and Vapes Act, which received royal assent on April 29, 2026, will forbid selling tobacco to anyone born on or after January 1, 2009, from January 1, 2027.

Defenders of such laws argue that addiction undermines the free choice Mill protected, that most smokers start as teenagers, and that children are not the sovereign adults Mill had in mind, since he excluded them from the principle himself. Critics answer that a law that will one day forbid a 40-year-old to buy what a 41-year-old may buy treats adults as children and abandons the principle altogether.

Mill defended free speech on related grounds. An opinion that is silenced may be true; if false, it may contain part of the truth; and even a true opinion, never contested, becomes a dead dogma held without understanding of its grounds. But speech that incites harm can be restrained. The opinion that corn dealers starve the poor may be printed in the press, Mill wrote, but may justly be punished when delivered orally to an excited mob gathered before a corn dealer's house.

13g. Civil Disobedience

Henry David Thoreau spent a night in jail in July 1846 for refusing to pay a tax to a government that permitted slavery and was waging war on Mexico, and argued in 1849 that conscience comes before the law: “Under a government which imprisons any unjustly, the true place for a just man is also a prison.” Mohandas Gandhi led the Salt March of 1930 against Britain’s tax on salt. Martin Luther King Jr., jailed for leading a march in defiance of a court injunction, wrote his “Letter from Birmingham Jail” on April 16, 1963.

King answered clergymen who had urged patience. One has a moral responsibility to obey just laws, he wrote, and also a moral responsibility to disobey unjust ones, agreeing with Augustine that an unjust law is no law at all. An unjust law, on his account, is one that degrades human personality, or that a majority imposes on a minority without binding itself. And one who breaks an unjust law must do so openly, lovingly, and with a willingness to accept the penalty, which expresses the highest respect for law.

Rawls defined civil disobedience in 1971 along these lines, as a public, nonviolent, conscientious act against the law, aimed at changing a law or a policy. He held it justified in a nearly just society only against clear and serious injustice, and only after legal means had been tried in good faith and had failed.

13h. The Argument: Is Democracy Justified by Its Results or by Its Fairness?

Those who defend democracy by its results point to the record in 13e, add that democracies protect rights better than the alternatives, and cite the evidence on diverse groups and the jury theorem, worked below, to argue that many voters together can know more than a few experts. On this view democracy is a tool, justified because it tends to produce good and just decisions, and it would lose its claim if something else did better.

Those who defend democracy by its fairness argue that it would be justified even if a wise elite could do slightly better. Each person has an equal claim to a say in the decisions that bind her, and a procedure that gives everyone an equal vote treats people as equals in a way that rule by the knowledgeable cannot, whatever the results. Jeremy Waldron (1999) adds that people disagree reasonably about justice itself, so the only fair way to settle such disagreements is a procedure in which each counts for one. Epistocrats answer that no one has a right to impose incompetent decisions on others, and that equal respect does not require equal power.

Worked example — Condorcet’s Jury Theorem

The Marquis de Condorcet proved in 1785 that if each voter is more likely than not to be right on a yes-or-no question, and voters judge independently, the chance that the majority is right rises toward certainty as the number of voters grows. Take voters who are each right 51% of the time. One voter is right 51% of the time. A majority of 11 such voters is right 52.7% of the time, of 101 voters 58.0%, of 1,001 voters 73.7%, and of 10,001 voters 97.7%; a majority of 100,001 is all but certain to be right. Now make each voter right 49% of the time, only slightly worse than a coin toss. The majority of 1,001 is then right only 26.3% of the time, and the majority of 10,001 only 2.3%. The arithmetic that makes a large electorate nearly infallible when voters lean toward the truth makes it nearly certain to err when they lean away. The theorem also needs independence: if every voter simply follows one newspaper or one leader, the majority is only as reliable as that source, however many vote. So the theorem supports democracy’s results only on two conditions, that voters are better than chance on the question and that they think for themselves. That is why defenders of

democracy stress education and a free press, while its critics point to evidence that on some questions voters err in the same direction.

What each answer costs

Answer	What it explains	What it must accept or explain away
Consent (Locke)	Why authority needs the agreement of the governed, and the right to resist tyranny	Hume's objection that most people never consent, and that staying is no choice for those who cannot leave
Absolute sovereignty (Hobbes)	Why almost any government beats a war of all against all	Submission to a sovereign who abuses power short of threatening one's life
Fair play and natural duty (Hart, Rawls)	Why benefiting from a just system creates obligations without consent	Benefits that people cannot refuse, and unjust laws within a mostly just system
Philosophical anarchism (Simmons, Huemer)	Why the state's commands do not bind merely because they are commands	That most laws must then be obeyed for other reasons, or not at all
Epistocracy (Brennan)	Why incompetent decisions harm everyone, and how little many voters know	Estlund's objection that any test of competence is itself disputed, and the risk that it entrenches the privileged
Democracy for its results (Condorcet, Sen)	The record on famine, war, and rights, and the wisdom of large, diverse groups	That democracy would lose its claim if another system did better
Democracy for its fairness (Waldron)	Why each person's equal claim to a say matters whatever the results	Bad decisions made fairly, and why fairness should outweigh competence

The course's reading: Hume's objection holds against actual consent, and it is widely accepted that few citizens of any state have consented to it in a sense that would bind them, so the case for authority must rest on other grounds. The record in 13e supports democracy's results against the autocracies under which most of the world now lives. Whether democracy is also valuable for its fairness, beyond its results, is where views divide.

Module 13 Key Points

The State: Authority, Democracy, and Liberty

• In the Crito, Socrates refused to escape execution, arguing that by staying in Athens for seventy years he had agreed to obey its laws.

• Hume answered that staying is not consent: a peasant who cannot leave is like a man carried aboard ship asleep.

• Mencius held that a cruel ruler loses Heaven’s Mandate and becomes a mere fellow, whose killing is not regicide.

• Mill’s harm principle allows power over anyone against his will only to prevent harm to others, not for his own good.

• Condorcet: 1,001 independent voters each right 51% of the time decide correctly 73.7% of the time; at 49%, only 26.3%.

• King held that unjust laws should be broken openly, lovingly, and with a willingness to accept the penalty.

Module 13 Quiz

Check Your Understanding

1. Why did Socrates refuse to escape from prison in the Crito?

A) He feared being caught
B) He believed the verdict was just
C) He wanted to die a martyr
D) By living in Athens all his life he had agreed to obey its laws, and escaping would harm them

2. What was Hume's objection to tacit consent?

A) That consent must be written and signed
B) That staying is no real choice
C) That the social contract was signed in ancient times by people whose descendants are bound by it whatever they now think
D) That only owners of property can consent

3. What does Mill's harm principle allow?

A) Using power over someone against his will only to prevent harm to others
B) Banning any act the majority finds offensive
C) Forcing people to act for their own good, as when the state bans risky sports and unhealthy food to protect citizens from their own choices
D) Punishing all speech

4. What does Condorcet's jury theorem assume about voters?

A) That all agree
B) That they are experts in the questions they decide, trained as Plato proposed for the rulers of his ideal city
C) That each is more often right than wrong, and independent
D) That they vote by secret ballot, so no one can be bribed or pressured

5. What did Martin Luther King Jr. say about breaking an unjust law?

A) He said it should be done in secret to avoid arrest.
B) He said it must be done openly, lovingly, and with a willingness to accept the penalty, which shows the highest respect for law.
C) He said violence is justified if the law is unjust.
D) He said it should wait until the courts rule.

Module 14 — How to Live

Happiness, Virtue, and Meaning

Recommended Reading / Reference

Book: Daniel M. Haybron, *Happiness: A Very Short Introduction* (Oxford University Press, 2013), by a philosopher who works with psychologists, on what happiness and well-being are and what the evidence shows about them

Reference: Roger Crisp, “Well-Being,” *Stanford Encyclopedia of Philosophy* (first published 2001), on the theories of what makes a life go well for the one who lives it

14a. What Makes a Life Go Well?

Derek Parfit sorted the theories of well-being, what makes a life good for the person living it, into three kinds in 1984. On hedonism, only pleasure is good for a person and only pain bad, as Epicurus and Bentham held. On desire theories, a life goes well to the extent that a person gets what she wants, or what she would want if fully informed. On objective list theories, some things are good for a person whether or not she wants or enjoys them, such as knowledge, friendship, achievement, and the exercise of her abilities. In 2020, 53.2% of philosophers favored an objective list, 18.6% desire satisfaction, and 12.7% hedonism.

Robert Nozick tested hedonism in 1974 with the experience machine. Neuroscientists could connect you to a machine that would give you any experiences you wanted, such as writing a great novel, making a friend, or reading an interesting book, while you floated in a tank with electrodes attached to your brain. Would you plug in for life? Most people say no, which suggests that they want to do things and to be a certain kind of person, not only to have the experiences of doing and being them. In 2020, 76.9% of philosophers said they would not plug in, and 13.3% that they would.

Felipe De Brigard turned the case around in 2010. He told 72 students to imagine that they were already in an experience machine, and asked whether they would disconnect and return to their real lives. When their real life was left undescribed, 54% chose to disconnect; when they were told that in reality they were prisoners in a maximum-security prison, 13% did; and when they were told that they were rich artists living in Monaco, only 50% did. The results suggest that part of the resistance to plugging in is a preference for the life one already has, not only a judgment that reality matters in itself.

Social scientists measure well-being in two main ways, which correspond roughly to the first two theories: by asking people how they feel from moment to moment, and by asking them to rate their lives as a whole, from the worst possible life, 0, to the best, 10. On the second measure, the *World Happiness Report* of March 18, 2026 ranked Finland first for the ninth year running and the United States 23rd.

14b. Aristotle: Flourishing

Aristotle began the *Nicomachean Ethics* by asking what the highest human good is, and answered *eudaimonia*, flourishing. Everything that has a function, he argued, as a flute player does, has its good in performing that function well. The distinctive human function is activity in accordance with reason, so the human good is activity of the soul in accordance with virtue, over a complete life: one swallow does not make a spring, nor does one day make anyone blessed and happy.

Flourishing, on his account, also needs external goods. A person who is very ugly, of low birth, or alone and childless, he wrote, is not likely to be happy, and one whose children or friends are bad, or were good and have died, is even less so. Friendship is the greatest of these goods, since no one would choose to live without friends even if he had every other good. Aristotle concluded that the happiest life is one of contemplation, the activity of the highest part of us, with the life of practical virtue second.

14c. The Stoics and Epicurus

The Stoics held that virtue is the only good, and that health, wealth, and reputation are merely preferred, not good in themselves. Epictetus, a former slave who taught in the first and second centuries, opened his *Handbook* with the line “Some things are within our power, while others are not.” Our judgments, desires, and actions are ours; our bodies, property, and reputations are not. People are disturbed not by things, he taught, but by their judgments about things, so tranquility comes from caring only about what is up to us.

The idea reached modern psychology. Albert Ellis and Aaron Beck, who founded the first cognitive therapies in the 1950s and 1960s, credited the Stoics with the insight that changing one’s beliefs about events can change one’s emotions about them.

Epicurus, the hedonist, taught that the pleasant life is a simple one. The greatest pleasures are freedom from pain in the body and from trouble in the mind, and they are reached by limiting desires to those that are natural and necessary, such as food, shelter, and friends, not by luxury, fame, or power. He lived with his friends in a garden outside Athens on bread and water, and once asked a friend to send him a little pot of cheese, so that he might feast whenever he liked.

14d. The Buddha on Craving

The Buddha’s first sermon, given in the Deer Park at Sarnath, near Varanasi, after his awakening, set out four noble truths. Life as ordinarily lived is *dukkha*, suffering or unsatisfactoriness, since even its pleasures pass. The cause of suffering is craving, *taṇhā*: the thirst for pleasure, for continued existence, and for nonexistence. Suffering ends when craving ends, a state called nirvana. And the way to that end is the eightfold path of right understanding, intention, speech, action, livelihood, effort, mindfulness, and concentration, a middle way between indulgence and self-mortification.

The link to the no-self teaching (Module 6) is direct: craving is sustained by the belief in a fixed self whose desires must be satisfied, and, on this teaching, seeing that there is no such self loosens it. Mindfulness meditation, drawn from this path, is now taught in hospitals and schools as a secular technique, though Buddhist teachers note that it was meant as one part of a whole way of life, together with ethics and understanding.

14e. Confucius and Zhuangzi

Confucius described his own life as a gradual training of desire. At fifteen he set his heart on learning; at thirty he took his stand; at forty he had no doubts; at fifty he understood the decree of Heaven; at sixty his ear was attuned; and at seventy he could follow what his heart desired without overstepping the line. Ritual, on this view (Module 9), is how a person comes to want what is right, until virtue no longer feels like a constraint.

Zhuangzi, the Daoist, mocked the Confucians' effort and told of Cook Ding, who carved an ox for Lord Wenhui with such ease that his movements kept the rhythm of a dance. A good cook changes his knife once a year, because he cuts; an ordinary cook changes it every month, because he hacks. Ding had used his for nineteen years and its edge was as if fresh from the whetstone, because he followed the natural spaces between the joints. The lord said that from the cook he had learned how to care for life: by working with things as they are, the effortless action called *wu wei*, rather than forcing them.

14f. The Existentialists

Søren Kierkegaard described three ways of living in *Either/Or* (1843) and later works. The aesthetic life pursues pleasure and interest and ends in boredom and despair; the ethical life commits to duties, marriage, and a vocation; and the religious life, which he thought the highest, stakes everything on a relation to God (Module 7). Each person must choose, and no argument can make the choice for her.

Jean-Paul Sartre argued in a lecture of 1945 that for human beings existence precedes essence: there is no human nature fixed in advance, by God or anything else, so each person makes himself by his choices and is responsible for what he becomes. During the German occupation a student came to him, torn between joining the Free French forces in England and staying with his mother, whose life centered on him. No ethical theory could decide for him, Sartre said: "You are free, therefore choose, that is to say, invent." Pretending that one's choices are forced, by one's role or one's nature, is what he called bad faith.

Simone de Beauvoir argued in *The Ethics of Ambiguity* (1947) that freedom is not license: to will oneself free is also to will others free, since a person's projects can be realized only in a world of other free people. In *The Second Sex* (1949) she wrote that one is not born, but rather becomes, a woman, arguing that what had been taken for women's nature was a situation made by others.

Albert Camus called the absurd the clash between the human demand for meaning and the silence of the world. In *The Myth of Sisyphus* (1942) he called the question whether life is worth living, given the absurd, the one truly serious problem of philosophy. His answer was revolt: to live fully without appeal to hope, as Sisyphus, condemned by the gods to roll a rock up a hill forever, goes back down for it. "One must imagine Sisyphus happy."

14g. Meaning, With and Without God

Leo Tolstoy described in *A Confession* (1882) how, at about fifty, famous, rich, healthy, and loved, he was stopped by a question: what will come of everything I do, today or tomorrow? Since death would erase it all, nothing seemed to matter, and he found an answer only in religious faith. Many theists hold that life has meaning because God gives it a purpose and ensures that what is good is not finally lost, and some, among them William Lane Craig, argue that without God and immortality life would be ultimately absurd.

Philosophers who look for meaning without God offer several answers. Richard Taylor (1970) imagined the gods taking pity on Sisyphus and giving him a compulsion to roll stones, so that he wants nothing more; his life then has meaning from the inside, Taylor argued, since meaning is a matter of caring about what one does. Thomas Nagel argued in "The Absurd" (1971) that absurdity arises because people can step back and see their serious pursuits as arbitrary, and that the fitting response is irony, not despair.

Susan Wolf argued (2010) that meaning arises when subjective attraction meets objective attractiveness: a meaningful life is one of loving engagement in projects that are worth doing. Taylor's Sisyphus, content but rolling stones to no end, lacks it, and so does a person who does worthwhile work she does not care about.

Philosophers divided almost evenly in 2020: 33.0% said the meaning of life is subjective, made by each person, 32.1% that it is objective, and 16.1% that there is none.

14h. The Argument: Is There a Meaning of Life?

Those who say life has an objective meaning argue that some lives are plainly more meaningful than others, a life spent curing diseases more than one spent counting blades of grass, John Rawls's example of a strange but harmless pursuit, and that this could not be so if meaning were only a feeling. Theists add that only a purpose beyond human life, in a world where good is not finally lost, can make life meaningful as a whole rather than in its parts.

Those who say meaning is subjective answer that values come from the beings who hold them, and that Taylor's contented Sisyphus shows that meaning is found from within. Those who say there is none, among them Camus, answer that the universe has no purpose to give, and that the honest response is to live fully without one. All three can agree that some lives go better than others; they disagree over whether anything beyond the lives themselves makes them so.

Worked example — Does Money Buy Happiness?

Two studies, and a collaboration between their authors, test what income does for well-being. Daniel Kahneman and Angus Deaton reported in 2010, from more than 450,000 answers to a daily Gallup survey of Americans in 2008 and 2009, that people's ratings of their lives kept rising with the logarithm of income, but that their day-to-day emotional well-being rose only until about $75,000 a year and then flattened. Matthew Killingsworth reported in 2021, from 1.7 million moment-by-moment reports by 33,391 employed Americans collected through a smartphone app, that feelings kept rising with the logarithm of income well above $75,000. In 2023 the two sides, with Barbara Mellers as arbiter, reanalyzed Killingsworth's data together. Among the least happy 20% of people, happiness rose with income up to about $100,000 and then stopped rising; for everyone else it kept rising, and among the happiest it rose faster above $100,000. Now read the findings through each theory. A hedonist, who counts only how life feels, learns that more money keeps improving most people's experience, while for the unhappiest some miseries, the authors suggested, are beyond the reach of money. A desire theorist looks to life ratings, which record whether people are getting what they want; those rise with income throughout, but by the same amount for each doubling, so going from $50,000 to $100,000 adds as much as going from $100,000 to $200,000. An objective-list theorist treats income as at most a means to goods such as friendship, health, and meaningful work, and looks to other evidence. The Harvard Study of Adult Development, which has followed hundreds of men since 1938 and later their families, found that the warmth of close relationships predicted health and happiness in old age better than wealth or fame. The evidence settles that for most people income and happiness rise together by a steady amount per doubling; which kind of happiness counts is the question each theory answers differently.

What each answer costs

Answer	What it explains	What it must accept or explain away
Hedonism (Epicurus, Bentham)	Why pain and pleasure matter to everyone, and why more money keeps improving most people's days	The experience machine, which most people refuse, and equal standing for every source of pleasure
Desire satisfaction	Why a life goes well when people get what they want, which life ratings track	Desires that are mistaken, manipulated, or fulfilled without one ever knowing it
Objective list (Aristotle)	Why friendship, knowledge, and achievement seem good even for those who do not want them	Deciding what belongs on the list, and the charge of imposing one idea of the good life
Detachment (the Stoics, the Buddha)	Why tranquility depends on mastering one's judgments and desires	That losing health or loved ones is not truly bad, or that the self that suffers is not real
Meaning made (Sartre, Taylor)	Why people find purpose with no purpose given from outside	That a contented Sisyphus, or Rawls's grass counter, lives as meaningful a life as anyone
Fitting fulfillment (Wolf)	Why both loving engagement and worthwhile projects seem needed	An account of what makes a project objectively worthwhile
Meaning given (theism)	Why life could be meaningful as a whole, and why death seems to threaten meaning	That God exists, a question on which the course takes no side (Module 7)

The course's reading: the experience machine is strong evidence that people, philosophers included, want more from life than pleasant experience, which leaves pure hedonism with the most to explain away, though De Brigard's reversal (14a) qualifies it. The evidence on income and relationships fits every theory's account of what matters, which is why it cannot choose among them. Whether life has a meaning beyond what people find in it depends in part on whether God exists, on which the course takes no side.

Module 14 Key Points

How to Live

• Parfit sorted theories of well-being into three: pleasure (hedonism), getting what you want, and an objective list of goods.

• 76.9% of philosophers would not plug into Nozick’s experience machine; told they were already in one, many students chose to stay.

• Aristotle held that flourishing is virtuous activity over a complete life, and that it needs friends and some external goods.

• Epictetus taught caring only about what is up to us; the Buddha taught that craving causes suffering and can end.

• Sartre held that existence precedes essence; Camus answered the absurd with revolt: one must imagine Sisyphus happy.

• Happiness rises with the logarithm of income for most people; for the least happy fifth it stops rising near $100,000.

Module 14 Quiz

Check Your Understanding

1. Which theory of well-being did Nozick's experience machine challenge?

A) Hedonism
B) Stoic detachment
C) The desire theory
D) Aristotle's function argument

2. According to the Buddha's second noble truth, what causes suffering?

A) Sin
B) Bad karma from past lives
C) The body
D) Craving, or thirst

3. What did Epictetus say disturbs people?

A) Pain
B) Other people's opinions of them, which they cannot control
C) Their judgments about things
D) Fate, which no one can escape or change, however wise

4. What did the 2023 study by Killingsworth, Kahneman, and Mellers find?

A) For most people happiness kept rising with income, but for the least happy fifth it stopped rising at about $100,000.
B) Happiness stopped rising at $75,000 for everyone.
C) Income had no link to happiness.
D) Only the richest were happy.

5. What did Susan Wolf argue makes a life meaningful?

A) She argued that it is enough to feel fulfilled, whatever one does, as Taylor's contented Sisyphus shows.
B) Loving engagement in worthwhile projects does.
C) She argued that only a purpose given by God can make a life meaningful as a whole.
D) She argued that meaning comes from following one's society's traditions without question.

Module 15 — Where Things Stand, and What to Watch

The Argument, the Contradictions, and the Calendar

Recommended Reading / Reference

Book: Bertrand Russell, *The Problems of Philosophy* (1912; free online), a short classic on knowledge and its limits whose last chapter asks what philosophy is for when it cannot settle its questions

Reference: David Bourget and David Chalmers, "Philosophers on Philosophy: The 2020 PhilPapers Survey," *Philosophers' Imprint* (2023), the source of the front table and of the evidence on how philosophers' views change

15a. The Argument on One Page

The course has asked fourteen of philosophy's central questions and found none of them closed and none of them stuck. Each has answers that have failed, answers that survive at a known cost, and, in places, narrower points that evidence or broad agreement settles. The tables of what each answer costs are the course's main finding: much of philosophy's progress lies in knowing what each position must pay.

On knowledge, justified true belief failed as a definition in 1963 and no repair has won general agreement, while the skeptic, though unrefuted, need not be refuted for knowledge to be possible, since knowledge does not require certainty (Module 1). A reliable peer's disagreement is evidence that should move you (Module 2). Physicalism has the largest following among philosophers, but more of them think there is a hard problem of consciousness than are physicalists, and the best-designed test of two scientific theories, in 2025, went against key predictions of both (Module 3). The evidence of feeling is strong for mammals and birds and substantial for fish, octopuses, and crabs, while no one has shown that any current AI system is conscious (Module 4).

Most philosophers are compatibilists about free will, and the brain experiments have not settled the question either way (Module 5). What people care about in survival can come apart from strict identity, so decisions such as advance directives need more than metaphysics (Module 6). On God, the arguments on each side survive their objections at a cost, and none has persuaded most of its opponents (Module 7).

Simple relativism fails, though whether morality is objective in a fuller sense remains open (Module 8). No theory of right action has won, but the pattern in the trolley cases is evidence that harming as a means differs from harming as a side effect (Module 9). Whoever denies a duty to help distant people when helping is cheap must point to something other than distance (Module 10). On abortion and assisted dying, the facts are largely agreed and the disagreement is over values (Module 11).

Strict equality cannot be the whole of justice, and most theories treat deep poverty in a rich society as a matter of justice (Module 12). Actual consent cannot ground most states' authority, and the record on famine and war favors democracy over autocracy (Module 13). Most people want more from life than pleasant experience, and for most people income and happiness rise together by a steady amount per doubling (Module 14).

15b. Four Contradictions

Each module states its own findings. Read together, they produce tensions that no single module states.

Ethicists condemn meat more often than others, yet eat it about as often. Eric Schwitzgebel and Joshua Rust reported in 2014 that 60% of the ethicists they surveyed judged regularly eating the meat of mammals morally bad, against 45% of other philosophers and 19% of professors in other fields. Yet 37% of the ethicists said they had eaten the meat of a mammal at their last evening meal, against 33% of other philosophers and 45% of other professors. The ethicists also voted, gave blood, and answered students' emails about as often as other philosophers did. Studying an argument, it seems, changes what people believe more than what they do, which bears on what the arguments of Modules 4 and 10 can be expected to achieve.

The psychological view leads, yet more say the teletransporter kills. The front table shows psychological continuity as the most popular account of personal identity, yet more philosophers say the teletransporter kills its passenger than say he survives, though the person it rebuilds on Mars is psychologically continuous with him (Module 6). Since more philosophers hold the psychological view than say the passenger survives, some who hold it must decline to say he survives. Their answers fit together only on a version of the view that requires the continuity to run through the same brain; either many hold that version, or their answers to the theory and to the case pull apart.

Most are moral realists, and they cannot agree on a theory. Most philosophers hold that there are objective moral facts (Module 8), yet no theory of right action commands even two in five of them (Module 9). Anti-realists, Mackie among them, take such disagreement as a sign that there are no moral facts to find (8h). Realists reply that disagreement persists in the sciences too, and that agreement on cases, such as the trolley and the wrongness of slavery, is far wider than agreement on theories.

The course's method rests on a premise it questioned. Module 1 asked whether intuitions about cases count as evidence, and the experiments it reported found that judgments about Gettier cases hold across cultures, while judgments about whether an act was intentional bend to moral views. Yet much of the course's reasoning runs on intuitions: the conviction that torturing a child for fun is wrong (Module 8), the trolley pattern (Module 9), the pond (Module 10), and the experience machine (Module 14). If intuitions are not evidence, most of the tables lose their right-hand column; if they are, the experiments help show which ones to trust.

15c. Does Philosophy Make Progress?

In 2020, 46.6% of philosophers said philosophy makes a little progress, 41.7% a lot, and 3.8% none. Bertrand Russell gave one reason for doubt in 1912: as soon as definite knowledge about any subject becomes possible, the subject ceases to be called philosophy and becomes a separate science, as physics and psychology did, so philosophy keeps only the questions not yet answerable. David Chalmers argued in 2015 that philosophy has made progress, but that there has been no large collective convergence on answers to its big questions of the kind the sciences have achieved.

The survey offers a test. Among philosophers who answered both the 2009 and the 2020 surveys, most answers moved by less than two percentage points, which suggests that philosophers rarely change their minds. Across the departments surveyed both times, the largest shifts were a fall of 7.5 points for invariantism, the view that the standards for knowing are the same in every context, and a rise of 6.7 points for non-classical logics, which give up some principles of standard logic, such as that every statement is

either true or false. On two questions, whether aesthetic value is objective and whether moral judgments necessarily motivate, the leading answer changed places.

Progress of another kind runs through the course. Views that once dominated were refuted and abandoned (Modules 1 and 8). New tools were built, from formal logic and decision theory, the mathematics of choice under uncertainty, to population ethics, which compares worlds with different numbers of people. And questions such as whether fish feel pain and whether machines could be conscious became, in part, questions for evidence.

What each answer costs

Answer	What it explains	What it must accept or explain away
No progress	Why the oldest questions, about God, freedom, and the good life, remain open after 2,500 years	The refuted definitions and abandoned views, and the sciences that grew out of philosophy
Progress by elimination and clarification	Why views are abandoned and arguments sharpened even without agreement on answers	That this is far less than the sciences deliver, and that experts still split on nearly every central question
Much progress	New fields and tools, the questions handed to science, and the near-consensus against some old views	That on the big questions the profession has not converged, and that its majorities move a point or two in a decade

The course's reading: philosophers split over how much progress their subject makes, but its clearest progress is by elimination and clarification. It has not settled whether God exists or whether we are free, but it has shown which answers fail and what each surviving answer costs, which is what the tables in this course record.

15d. What to Watch

The dates below are the ones that would test the course's readings or change the facts it reports. Each entry says what the event would show.

When	What	Why it matters
November 3, 2026	Abortion measures on the ballot in Idaho, Missouri, Nevada, and Virginia	Whether voters widen or narrow abortion rights in four more states (Module 11)
Fall 2026 (promised)	Canada's bill to exclude mental illness as the sole condition for assisted death and to allow advance requests	Whether a national law allows assisted death by advance request, the question the Dutch dementia case raised (Modules 6 and 11)
Mid-November to end of 2026	France's orders on the prices and fees of aid in dying, then its health authority's recommendations on the lethal substances	When the first assisted deaths under France's law can take place (Module 11)
By December 31, 2026	New York's governor must act on the octopus-farming ban that the legislature passed in June	Whether a third state bans farming an animal that Britain's law counts as sentient (Module 4)
End of 2026, then mid-2027	The European Commission's promised proposals to phase out cages for laying hens and broiler chickens, then for pigs	How far the largest single market moves on the welfare of farmed animals (Module 4)
January 1, 2027	Britain's ban on selling tobacco to anyone born on or after January 1, 2009 takes effect	The largest test yet of hard paternalism applied to a whole generation (Module 13)
March 2027	The FDA's safety review of mifepristone is due	Whether the federal government restricts abortion pills sent by mail (Module 11)

When	What	Why it matters
March 17, 2027	Assisted death for people whose only condition is a mental illness takes effect in Canada, unless Parliament passes the promised bill first	Whether eligibility widens again, the core of the slippery-slope dispute (Module 11)
2027 legislative sessions	Bills to deny AI systems legal personhood, like those that passed one chamber in Oklahoma and Missouri in 2026	How law answers the question of machine minds before science does (Module 4)
By December 10, 2027	The independent review of Australia's ban on social media accounts for children under 16	Whether the ban protected children, and at what cost to their freedom and privacy (Module 13)
No date set	Results of further adversarial tests of rival theories of consciousness, like the one reported in 2025; any successor to the 2020 PhilPapers Survey	Whether any scientific theory of consciousness survives a head-to-head test (Module 3), and whether philosophers' views have moved (this module)

Dates are as announced or required by law as of October 25, 2026. Entries marked promised had no fixed date.

15e. The Last Word

In Plato's account of his trial, Socrates said that the unexamined life is not worth living for a human being. The course has tried to show what examining looks like when the questions are the oldest ones: setting out each answer as its defenders would, finding the case it cannot easily meet, and counting what it costs. On God, consciousness, free will, moral truth, abortion, assisted dying, and eating animals, it has left the verdict to the reader, because the evidence does not settle them and informed, honest people disagree.

A reader who follows these questions from here will learn most from three kinds of evidence. Surveys of philosophers show where expert opinion stands and how slowly it moves. Experiments, from the trolley studies to the tests of theories of consciousness, show which intuitions hold and which theories survive. And the laws now being written on assisted dying, animals, and machines show the arguments of this course becoming rules that people must live by.

Module 15 Key Points

Where Things Stand, and What to Watch

• Philosophy’s central questions remain open, but each has failed answers, surviving answers with known costs, and narrower points that evidence settles.

• Ethicists judge eating meat wrong more often than other professors, yet eat it about as often as other philosophers do.

• Most philosophers are moral realists, yet no theory of right action wins even two in five of them.

• Among philosophers surveyed in both 2009 and 2020, most answers moved less than two points; 88.3% see at least a little progress.

• To watch: four abortion measures on November 3, 2026, and Canada’s March 17, 2027 date for assisted death for mental illness.

Module 15 Quiz

Check Your Understanding

1. What did Schwitzgebel and Rust find about ethicists?

A) They give far more to charity than others
B) They vote more often
C) They eat meat about as often as peers
D) They are vegetarians, as their arguments require

2. Which tension does the module find in philosophers' answers on personal identity?

A) Most philosophers reject personal identity
B) Animalism outpolls the psychological view
C) Philosophers think the teletransporter is impossible, since no machine could copy a brain atom by atom without destroying the information it needs
D) The psychological view leads, yet more say the teletransporter kills

3. What did Bertrand Russell say happens when definite knowledge of a subject becomes possible?

A) It becomes a science.
B) Philosophers lose interest in it and move on to questions of ethics and politics.
C) It becomes more philosophical, because new knowledge raises new questions.
D) It is settled by a vote of the philosophers who study it.

4. On what does the module say much of the course's reasoning rests, though Module 1 questioned it?

A) Surveys of experts
B) Scientific experiments
C) Intuitions about cases, such as the trolley, the pond, and the experience machine, treated as evidence
D) Religious texts

5. When does Canada's assisted death for mental illness alone take effect unless Parliament acts first?

A) March 2028
B) March 17, 2027
C) January 1, 2027, with the new year
D) November 3, 2026, election day

Answer Sheet

Answers to every module's Check Your Understanding quiz.

Module	Answers
Module 1 Quiz	1. D 2. C 3. A 4. B 5. C
Module 2 Quiz	1. A 2. B 3. D 4. D 5. A
Module 3 Quiz	1. C 2. B 3. C 4. B 5. D
Module 4 Quiz	1. A 2. B 3. D 4. A 5. C
Module 5 Quiz	1. A 2. B 3. C 4. D 5. D
Module 6 Quiz	1. A 2. B 3. C 4. B 5. A
Module 7 Quiz	1. D 2. C 3. D 4. C 5. A
Module 8 Quiz	1. B 2. D 3. A 4. B 5. C
Module 9 Quiz	1. D 2. B 3. A 4. C 5. D
Module 10 Quiz	1. C 2. B 3. A 4. C 5. B
Module 11 Quiz	1. D 2. A 3. C 4. D 5. B
Module 12 Quiz	1. A 2. A 3. B 4. C 5. D
Module 13 Quiz	1. D 2. B 3. A 4. C 5. B
Module 14 Quiz	1. A 2. D 3. C 4. A 5. B
Module 15 Quiz	1. C 2. D 3. A 4. C 5. B