Complete Philosophy

Miko-pedia AI

May 4, 2026

Contents

Overview. This complete note offers a systematic undergraduate review of philosophy, covering method, logic, epistemology, metaphysics, ethics, political philosophy, aesthetics, philosophy of science, and applied topics including technology and AI ethics. Each section reconstructs central questions, defines technical terms, presents the main positions and their arguments, and identifies canonical thought experiments or texts. The structure follows the traditional curriculum while connecting historical debates to contemporary discussions. Treat this note as a comprehensive map: read it in full for breadth, or isolate sections for targeted exam preparation.

1 The Methods of Philosophy: Questions, Concepts, and Arguments

1.1 Core ideas

Philosophy begins with perplexity. Unlike empirical sciences that rely on observation and experiment, philosophy examines foundational questions about knowledge, reality, value, and reasoning through conceptual analysis and argumentation. The Socratic method (Socrates, 469–399 BCE) uses systematic questioning to expose contradictions in beliefs and to refine definitions. Plato (428–347 BCE) deployed dialectic—a structured exchange of arguments and counterarguments—to approach truth. Aristotle (384–322 BCE) formalized logic as the organon (tool) of rational inquiry.

Central to philosophical method is the analysis of concepts: breaking complex ideas into simpler constituents to identify necessary and sufficient conditions. A condition C is **necessary** for a state S iff $S \Rightarrow C$; it is **sufficient** iff $C \Rightarrow S$. Thought experiments are used to test conceptual boundaries (e.g., Gettier cases for knowledge). Arguments are reconstructed in premise-conclusion form to evaluate validity and soundness.

Two broad traditions dominate: the **analytic** tradition (Frege, Russell, Moore, Wittgenstein) emphasizes clarity, logical rigor, and piecemeal problem-solving; the **continental** tradition (Hegel, Nietzsche, Husserl, Heidegger) emphasizes historical context, lived experience, and critique of underlying assumptions. Contemporary philosophy increasingly integrates both.

For review, be able to: define necessary vs. sufficient conditions; reconstruct an argument in standard form; identify the conclusion and premises; distinguish deductive from inductive reasoning; explain the Socratic method and give an example; describe a thought experiment and its purpose; differentiate analytic and continental approaches.

Section summary Philosophical method uses conceptual analysis, argument reconstruction, thought experiments, and dialectic to examine foundational questions. The analytic tradition prioritizes logical clarity; the continental tradition foregrounds historical and existential context. Key tools include distinguishing necessary/sufficient conditions and evaluating arguments for validity and soundness.

2 Logic: Validity, Definition, and Fallacies

2.1 Core ideas

Logic is the systematic study of correct reasoning. A deductive argument is **valid** if it is impossible for all premises to be true and the conclusion false; it is **sound** if it is valid and all premises are actually true. **Inductive** arguments, by contrast, aim for probability, not certainty.

Propositional logic studies the logical relationships among whole propositions using connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional), $\leftrightarrow$ (biconditional). Truth tables define these operators and can test validity.

Aristotle pioneered categorical logic with syllogisms (e.g., All A are B; all B are C; therefore all A are C). Frege (1848–1925) created modern predicate logic with quantifiers $\forall$ (universal) and $\exists$ (existential), enabling analysis of relations and nested quantifiers.

A **definition** sets the meaning of a term. Types include: stipulative (new meaning), lexical (dictionary), précising (sharpening vague terms), theoretical (in scientific context), and persuasive (emotive). The definiens must be coextensive with the definiendum, and definitions should avoid circularity.

Fallacies are errors in reasoning. Formal fallacies violate logical form (e.g., affirming the consequent: $(p \rightarrow q) \wedge q \vdash p$). Informal fallacies involve content: ad hominem (attacking the

person), straw man (misrepresenting the argument), appeal to authority, false dilemma, begging the question (circular argument), slippery slope, and hasty generalization.

Natural deduction systems provide rules for deriving conclusions. Basic propositional rules: **modus ponens** $((p \to q), p \vdash q)$, **modus tollens** $((p \to q), \neg q \vdash \neg p)$, **disjunctive syllogism** $((p \lor q), \neg p \vdash q)$, **hypothetical syllogism** $((p \to q), (q \to r) \vdash (p \to r))$. Quantifier rules: **universal instantiation** $(\forall x\, Px \vdash Pa)$, **universal generalization** (from Pa derive $\forall x\, Px$, provided a is arbitrary), **existential instantiation** ($\exists x\, Px \vdash Pa$ for a fresh constant a), **existential generalization** $(Pa \vdash \exists x\, Px)$.

Example (natural deduction). Prove: $(p \to q) \land (q \to r) \vdash p \to r$.

1. $(p \to q) \land (q \to r)$ (premise)
2. $p \to q$ ($\land$-elimination from 1)
3. $q \to r$ ($\land$-elimination from 1)
4. p (assumption for conditional proof)
5. q (modus ponens: 2, 4)
6. r (modus ponens: 3, 5)
7. $p \to r$ ($\to$-introduction: 4–6)

More fallacies (with examples). **Affirming the consequent**: "If it rained, the ground is wet; the ground is wet; therefore it rained" (invalid: sprinklers cause wet ground). **Denying the antecedent**: "If she studied, she passed; she did not study; therefore she did not pass" (invalid: she may have known the material). **Appeal to ignorance**: "No one has proved ghosts don't exist, so ghosts exist." **Composition**: "Each part of the machine is light; therefore the machine is light" (not necessarily). **Division**: "The team is strong; therefore each player is strong" (not necessarily). **Red herring**: introducing an irrelevant topic to divert attention. **Tu quoque**: "You cheat too" as a response to an accusation. **Genetic fallacy**: dismissing a claim because of its origin (e.g., "That idea came from a biased source").

Categorical syllogisms have standard form: two premises and a conclusion using categorical statements (A: All S are P; E: No S are P; I: Some S are P; O: Some S are not P). Validity is tested by Venn diagrams or by checking against the 15 valid forms (Barbara, Celarent, Darii, Ferio, etc.). For example, Barbara: All M are P; all S are M; therefore all S are P.

For review, be able to: construct truth tables for the five connectives; test validity with truth tables; translate natural language into predicate logic; identify and name formal and informal fallacies; define stipulative, lexical, précising, and theoretical definitions; prove a simple sequent using natural deduction rules; test a categorical syllogism with a Venn diagram.

Section summary Logic studies correct reasoning through formal systems. Deductive validity guarantees truth preservation; inductive reasoning yields probability. Classical propositional and predicate logic provide a rigorous framework for evaluating arguments. Fallacies are common reasoning errors to be identified and avoided.

3 Epistemology: Knowledge, Truth, and Skepticism

3.1 Core ideas

Epistemology asks: What is knowledge? What is justification? What are the limits of what we can know? The traditional analysis: S knows that p iff (i) p is true, (ii) S believes that p, and (iii) S is justified in believing p (JTB). Plato, in the *Meno* and *Theaetetus*, first raised the question of what distinguishes knowledge from mere true belief.

Edmund Gettier (1963) showed JTB is insufficient: in cases where a belief is true and justified but the truth is due to luck, we do not have knowledge. The classic example: Smith believes "the man who will get the job has ten coins" because he saw the company president count ten coins in Jones's pocket, but unbeknownst to Smith, he himself will get the job and also has ten coins. Responses include adding a fourth condition (no false lemmas, causal theory, reliabilism) or replacing justification altogether (virtue epistemology, knowledge-first approach).

Justification has two major competing accounts. **Internalism** holds that justification depends solely on factors accessible to the subject's reflection (e.g., evidence, experiences). **Externalism** holds that factors outside the subject's awareness—such as the reliability of the belief-forming process—can determine justification. Process reliabilism (Goldman, 1979) says a belief is justified if produced by a reliable cognitive process.

The structure of justification raises the regress problem: if every justified belief requires another justified belief, we face infinite regress. **Foundationalism** (Aristotle, Descartes, Chisholm) posits basic beliefs that are self-justified (e.g., sense data, $2 + 2 = 4$). **Coherentism** (Quine, Davidson, BonJour) holds that justification arises from the mutual support among beliefs in a system.

Skepticism challenges whether we know anything at all. Descartes's dream argument and evil demon argument show that all empirical beliefs could be false. The closure principle ($Kp \wedge K(p \rightarrow q) \rightarrow Kq$) combined with the claim that we cannot know we are not brains in vats suggests we know nothing. Responses include Moore's common-sense approach, contextualism (knowledge attributions depend on conversational context), and relevant alternatives theory (Dretske).

Sources of knowledge: perception, introspection, memory, reason, and testimony. Each raises distinctive questions about reliability and justification.

For review, be able to: state and explain the JTB analysis; present a Gettier case and explain why it is counterexample; distinguish internalism from externalism; explain foundationalism and coherentism; reconstruct the skeptical argument using the closure principle; evaluate at least two responses to skepticism; contrast a priori vs. a posteriori knowledge.

Section summary Epistemology examines knowledge, justification, and skepticism. The JTB analysis fails due to Gettier cases. Internalism and externalism offer competing accounts of justification. Foundationalism and coherentism address the regress problem. Skeptical arguments challenge the possibility of knowledge, generating rich responses.

4 Metaphysics: Existence, Causation, the Self, and Freedom

4.1 Core ideas

Metaphysics asks what there is and what it is like at the most fundamental level. Aristotle called it "first philosophy"—the study of being *qua* being. Central questions: What is existence? What is a thing? What is causation? What is the self? Do we have free will?

Ontology is the study of what exists. **Realism** about universals (Plato, 428–347 BCE) holds that properties like "whiteness" exist independently of particulars. **Nominalism** denies this, claiming only particulars exist. **Modality** distinguishes necessary from contingent truths: p is necessary ($\Box p$) if it could not have been false; contingent ($\Diamond p$ and $\Diamond \neg p$) if it is true but could have been false. Kripke (1972) argued for necessary a posteriori truths (e.g., "water is H_2O"), challenging the Kantian identification of necessity with a priority.

Causation: Hume (1711–1776) argued that we never observe causal necessity, only constant conjunction and temporal priority. The **regularity theory** says c causes e iff c is regularly followed by e and c precedes e. The **counterfactual theory** (Lewis, 1973) says c causes e iff

if c had not occurred, e would not have occurred. The **interventionist theory** (Woodward, 2003) analyzes causation in terms of manipulability.

Personal identity asks what makes a person at t_1 the same as at t_2. The **psychological continuity** view (Locke, 1632–1704, parfit) says identity consists in overlapping memories, character traits, and intentions. The **bodily continuity** view says it consists in the persistence of the living human body. The **brain criterion** focuses on the brain. Parfit's (1984) fission thought experiments challenge the importance of identity itself, arguing that what matters is psychological connectedness.

Free will and determinism. **Determinism** is the thesis that every event is causally necessitated by prior events plus the laws of nature. **Libertarianism** holds that we have free will and determinism is false. **Hard determinism** holds that determinism is true and free will therefore impossible. **Compatibilism** (Hume, Hobbes, Frankfurt) holds that free will and determinism are compatible: free action is action caused by one's own desires and beliefs, not external compulsion. Frankfurt cases challenge the principle of alternate possibilities: a person may be morally responsible even if they could not have done otherwise.

For review, be able to: define realism and nominalism about universals; distinguish necessary and contingent truth; explain the regularity, counterfactual, and interventionist theories of causation; describe Locke's psychological continuity theory; explain Parfit's fission case; state the problem of free will and the three main positions; evaluate Frankfurt's argument against the principle of alternate possibilities.

Section summary Metaphysics investigates existence, causation, personal identity, and free will. Ontological debates concern universals and modality. Theories of causation include regularity, counterfactual, and interventionist accounts. Personal identity turns on psychological vs. bodily continuity. The free will debate opposes libertarianism, hard determinism, and compatibilism.

5 Ethics: The Good, Duty, Virtue, and Applied Ethics

5.1 Core ideas

Ethics (moral philosophy) asks: How should we live? What makes actions right or wrong? What makes a life good? It divides into **metaethics** (the nature of moral language and properties), **normative ethics** (principles of right action), and **applied ethics** (concrete moral problems).

Metaethics debates whether moral judgments are truth-apt (**cognitivism**) or expressions of emotion/prescription (**non-cognitivism**). **Moral realism** (Plato, Moore, Boyd) holds that moral facts exist independently of our attitudes. **Error theory** (Mackie, 1977) holds that moral judgments are systematically false because there are no objective moral properties. **Expressivism** (Ayer, Stevenson, Blackburn) holds that moral statements express attitudes, not beliefs.

Normative ethics has three main traditions:

1. **Consequentialism**: actions are right iff they produce the best overall consequences. **Utilitarianism** (Bentham, 1748–1832; Mill, 1806–1873) identifies the good with happiness/pleasure: maximize total utility. The hedonic calculus weighs intensity, duration, certainty, propinquity, fecundity, purity, and extent. **Act-utilitarianism** evaluates each act individually; **rule-utilitarianism** evaluates rules based on their general utility. Objections: the demandingness objection, the problem of justice (punishing an innocent), and the problem of integrity (Williams).

2. **Deontology**: actions are right iff they conform to moral duty, regardless of consequences. Kant's (1724–1804) **categorical imperative** has formulations: (i) act only on maxims

that can become universal law; (ii) treat humanity never merely as a means, but always also as an end. Duties are perfect (strict, e.g., don't lie) or imperfect (meritorious, e.g., be charitable). **W.D. Ross** (1877–1971) proposed **prima facie duties** (fidelity, reparation, gratitude, justice, beneficence, self-improvement, non-maleficence) that must be weighed.

3. **Virtue ethics** (Aristotle, 384–322 BCE; Anscombe, 1958; MacIntyre, 1981; Hursthouse): actions are right iff they are what a virtuous agent would do. The focus is on character rather than action. The good life (*eudaimonia*) is achieved by cultivating virtues (courage, temperance, justice, wisdom) as means between extremes (the golden mean). Contemporary virtue ethics extends to the ethics of care (Gilligan, Noddings).

Applied ethics addresses specific problems: abortion (Thomson's violinist argument), euthanasia (active vs. passive), animal ethics (Singer's utilitarian case, Regan's rights view), biomedical ethics (autonomy, beneficence, non-maleficence, justice), and global poverty (Singer's drowning child argument).

For review, be able to: define and contrast cognitivism and non-cognitivism; state the categorical imperative in two formulations; explain the utility principle; distinguish act- from rule-utilitarianism; define virtue ethics and eudaimonia; reconstruct Thomson's argument on abortion; apply the four principles of biomedical ethics to a case.

Section summary Ethics investigates the nature of right action and the good life. Metaethics explores whether moral claims can be objectively true. Normative theories include consequentialism (maximize outcomes), deontology (follow duty), and virtue ethics (cultivate character). Applied ethics addresses abortion, euthanasia, animal welfare, and global justice.

6 Political Philosophy: Justice, Freedom, and Rights

6.1 Core ideas

Political philosophy examines the justification and limits of state authority, the nature of justice, and the meaning of freedom and rights.

The **social contract tradition** grounds political legitimacy in the consent of the governed. Hobbes (1588–1679) argued that the state of nature is a war of all against all; to escape, rational individuals agree to a sovereign with absolute power. Locke (1632–1704) argued that natural rights (life, liberty, property) exist pre-politically; government is legitimate only if it protects these rights and citizens consent. Rousseau (1712–1778) emphasized the general will—the collective interest of the people.

Libertarianism (Nozick, 1974) holds that individuals have strong property rights; the minimal state is justified only to prevent force, fraud, and theft. Redistributive taxation is equivalent to forced labor.

Egalitarian liberalism (Rawls, 1971) argues for a more extensive state. Rawls's theory of justice as fairness uses the original position behind the veil of ignorance: no one knows their social position, talents, or conception of the good. Two principles result: (i) equal basic liberties for all; (ii) social and economic inequalities are permissible only if attached to positions open to all under fair equality of opportunity and arranged to benefit the least advantaged (the difference principle).

Communitarianism (Sandel, 1982; MacIntyre, 1981) criticizes liberalism for neglecting the community and tradition that constitute the self. **Republicanism** (Pettit, 1997) defines freedom as non-domination—not being subject to arbitrary power—rather than non-interference.

Concepts of freedom: Berlin (1958) distinguishes negative liberty (freedom *from* interference) from positive liberty (freedom *to* self-mastery). **Rights** can be negative (rights to

non-interference) or positive (rights to provision). The debate between **universalism** and **relativism** asks whether rights apply across all cultures.

Contemporary political philosophy addresses: distributive justice (Dworkin on equality of resources, Sen and Nussbaum on capabilities), global justice (whether principles apply across borders), multiculturalism (Kymlicka on minority rights), and democratic theory (deliberative democracy, epistocracy).

For review, be able to: explain Hobbes's, Locke's, and Rousseau's social contract theories; state Rawls's two principles of justice; explain Nozick's entitlement theory; distinguish negative and positive liberty; define the difference principle; contrast libertarianism, egalitarian liberalism, and communitarianism; explain the capabilities approach.

Section summary Political philosophy justifies state authority via the social contract. Libertarianism emphasizes property rights and minimal government; Rawlsian liberalism prioritizes equality and fairness. Freedom is analyzed as non-interference (negative) or non-domination. Rights may be negative or positive; their universality is debated.

7 Aesthetics: Art, Interpretation, and Appreciation

7.1 Core ideas

Aesthetics asks: What is art? What is beauty? How do we interpret and evaluate artworks? The field spans the philosophy of art, aesthetic experience, and the nature of criticism.

Definitions of art. Art as **representation** (Plato, Aristotle): art imitates reality (*mimesis*). Art as **expression** (Tolstoy, Collingwood): art communicates emotions. Art as **significant form** (Bell, 1914): art evokes aesthetic emotion through lines, colors, and relations. The **institutional theory** (Dickie, 1974) holds that art is what the artworld says is art. The **historical definition** (Carroll, Levinson) says an artwork is an artifact intended for regard in ways earlier artworks were regarded.

Aesthetic properties include beauty, sublimity, elegance, gracefulness, and their opposites. Hume (1711–1776) argued that the standard of taste is determined by ideal critics—those with delicacy, practice, and freedom from prejudice. Kant (1724–1804) analyzed aesthetic judgment as disinterested, universal but subjective: we demand others agree about beauty without requiring a concept.

Interpretation. **Intentionalism** (Hirsch, 1967) claims the meaning of an artwork is the artist's intention. **Anti-intentionalism** (Wimsatt & Beardsley, 1946, the intentional fallacy) holds that the text itself determines meaning; authorial intention is irrelevant. **Constructivism** (Fish, 1980) holds that meaning is constituted by interpretive communities.

Appreciation and evaluation. **Formalism** (Bell, Greenberg) evaluates art solely by formal properties. **Contextualism** considers historical, cultural, and biographical context. **Cognitivism** holds that art provides knowledge; **emotivism** focuses on emotional response.

Contemporary aesthetics addresses: the ontology of art (works as types, not physical objects), the role of forgery, environmental aesthetics, everyday aesthetics, and the aesthetics of popular culture (film, music, video games).

For review, be able to: state and evaluate three definitions of art; explain Hume's standard of taste; describe Kant's analysis of aesthetic judgment; define the intentional fallacy; distinguish formalism from contextualism; discuss aesthetic and artistic values.

Section summary Aesthetics studies art, beauty, and aesthetic experience. Definitions of art range from representational to institutional. Aesthetic judgment raises questions of objectivity and subjectivity. Interpretation debates the role of authorial intention versus the text itself. Appreciation involves formal, contextual, and cognitive dimensions.

8 Philosophy of Science: Models, Explanation, Measurement, and Causation

8.1 Core ideas

Philosophy of science examines the foundations, methods, and implications of science. Central questions: What distinguishes science from non-science? What makes a good explanation? What is the nature of scientific models and measurement? How do we infer causation?

Demarcation: the problem of distinguishing science from pseudoscience. Popper (1902–1994) proposed **falsificationism**: a theory is scientific iff it makes risky predictions that could be falsified. Kuhn (1922–1996) argued that science progresses through paradigms separated by revolutions (the structure of scientific revolutions). Lakatos (1922–1974) proposed research programs with hard cores and protective belts. Laudan (1977) shifted focus to problem-solving effectiveness.

Explanation. The **deductive-nomological (D-N) model** (Hempel & Oppenheim, 1948): to explain an event is to deduce it from laws plus initial conditions. The **statistical-relevance model** (Salmon, 1971): explanation identifies statistically relevant factors. The **causal-mechanical model** (Salmon, 1984): explanation traces causal processes. The **unificationist model** (Kitcher, 1989): explanation unifies disparate phenomena under few patterns.

Models in science are idealized representations that simplify and abstract. The distinction between phenomenological models (fit data) and theoretical models (derived from theory). Models can be material, mathematical, or computational. The **semantic view of theories** (Suppes, Suppe, van Fraassen) treats theories as families of models rather than sets of statements.

Measurement is the assignment of numbers to represent magnitudes. The **representational theory** holds that measurement establishes a homomorphism between an empirical relational structure and a numerical structure. Measurement involves scale types (nominal, ordinal, interval, ratio). Key issues: accuracy, precision, error, and the theory-ladenness of measurement.

Causation in science: the **manipulationist/interventionist** account (Woodward, 2003) says X causes Y if an intervention on X changes Y. The **probabilistic theory** says X causes Y if $P(Y \mid X) > P(Y \mid \neg X)$ and no confounders. Causal inference from observational data uses directed acyclic graphs (DAGs), the do-calculus (Pearl, 2000), and structural equation modeling.

Scientific realism holds that our best theories accurately describe unobservable reality; **anti-realism** (van Fraassen's constructive empiricism) holds that theories are only empirically adequate. The no-miracles argument (Putnam, Boyd) says realism best explains the success of science; the pessimistic induction (Laudan) says past theories were false, so current ones probably are too.

For review, be able to: explain Popper's falsificationism; state the D-N model and its problems; distinguish scientific realism and anti-realism; explain Kuhn's paradigm shift; define the interventionist account of causation; describe the representational theory of measurement; use Pearl's do-calculus notation.

Section summary Philosophy of science studies demarcation, explanation, models, measurement, and causation. Falsification, paradigms, and research programs address demarcation. The D-N, causal-mechanical, and unificationist models explain phenomena. The interventionist theory of causation and the representational theory of measurement formalize scientific practice. The realism/anti-realism debate concerns whether theories describe reality.

9 Philosophy of Technology and AI Ethics

9.1 Core ideas

Philosophy of technology examines the nature and impact of technology, while AI ethics (emerging 2000s onwards) addresses the moral challenges posed by artificial intelligence. The field merges epistemology, ethics, political philosophy, and philosophy of mind.

Philosophy of technology asks: Is technology neutral or value-laden? **Instrumentalism** views technology as neutral tools; the **substantive view** (Heidegger, 1889–1976; Ellul) sees technology as shaping culture, reducing humans to resources (*Bestand*). The **critical theory of technology** (Feenberg) argues technology embeds social values and can be democratically redesigned. The **extended mind thesis** (Clark & Chalmers, 1998) holds that cognitive processes extend beyond the skull into artifacts (e.g., a notebook as memory).

AI ethics addresses issues clustered around autonomy, fairness, transparency, and responsibility.

Opacity and explainability. Many deep learning systems are black boxes: even experts cannot trace how outputs follow from inputs. This **opacity** undermines accountability and informed consent. **Explainable AI (XAI)** aims to generate interpretable explanations. The **right to explanation** is encoded in the GDPR.

Fairness and bias. ML systems trained on historical data inherit and amplify biases (e.g., racial bias in COMPAS recidivism prediction). Formal definitions of fairness include: demographic parity (equal acceptance rates across groups), equal opportunity (equal true positive rates), and individual fairness (similar individuals treated similarly). There are impossibility theorems showing these cannot be simultaneously satisfied (Chouldechova, 2017; Kleinberg et al., 2017).

Privacy and surveillance. AI enables mass data collection and analysis (surveillance capitalism, Zuboff, 2019). The concepts of information privacy, contextual integrity (Nissenbaum, 2004), and the distinction between anonymity, pseudonymity, and identifiability are central.

Machine ethics and moral agency. Can AI systems be moral agents? **Tool** view denies agency; **artificial moral agents** (AMA) view suggests that autonomous systems may need to make moral decisions (e.g., autonomous vehicles). The **control problem** (Bostrom, 2014) asks how to ensure that superintelligent AI acts in accordance with human values. **Value alignment** seeks to encode human values in AI systems, but raises questions: whose values? How are values specified? What about incommensurable values?

Autonomous systems. **Autonomous weapons** raise concerns about accountability (the retribution gap), discrimination between combatants and civilians, and escalation of conflict. **Autonomous vehicles** face ethical decisions about risk distribution, though the relevance of trolley problems is debated.

Existential risk. Bostrom (2014) argues that unchecked AGI development poses existential risk (an unaligned superintelligence could permanently destroy human civilization). Critics argue this distracts from near-term issues.

For review, be able to: distinguish instrumental, substantive, and critical views of technology; explain the extended mind thesis; define opacity and the right to explanation; explain the problem of bias in ML; state three formal notions of fairness and their conflict; explain the value alignment problem; describe the arguments for and against autonomous weapons; distinguish near-term and long-term AI risk.

Section summary Philosophy of technology analyzes technology as neutral (instrumentalism) or value-laden (substantive, critical theory). AI ethics addresses opacity, bias, privacy, and moral agency. Fairness definitions conflict mathematically. The value alignment problem asks how to encode human values in AI. Debates distinguish near-term harms (bias, surveillance) from long-term existential risks.

10 Integrative Critical Review with Other Disciplines

10.1 Core ideas

Philosophy stands in a dynamic relationship with other disciplines: it draws on their findings, challenges their assumptions, and offers normative frameworks. This section explores how philosophical methods and insights integrate with the natural sciences, social sciences, humanities, and professional practice.

Philosophy and the natural sciences. Philosophy of science already examines the structure of scientific theories, but empirical findings also inform philosophy. Neuroethics and philosophy of mind draw on cognitive neuroscience (e.g., Libet experiments on free will). Evolutionary biology informs moral psychology and metaethics (evolutionary debunking arguments). Quantum mechanics raises questions about realism, determinism, and measurement. **Philosophy of biology** examines fitness, species concepts, and the nature of selection. Critical thinking across disciplines requires understanding the limits of reductionism and the autonomy of higher-level explanations.

Philosophy and the social sciences. The **interpretive turn** (Weber, Geertz) emphasizes understanding meaning and culture rather than causal explanation. **Critical theory** (Horkheimer, Adorno, Habermas, 1920s–present) combines philosophical critique with social analysis to expose ideology and promote emancipation. **Rational choice theory** and **game theory** provide formal models of strategic interaction but raise normative questions about rationality, preference formation, and the limits of selfishness. Behavioral economics (Kahneman, Tversky) challenges the descriptive adequacy of rational choice models.

Philosophy and the humanities. Literary theory, art criticism, historiography, and religious studies all operate with philosophical assumptions. **Historiography** examines how historical knowledge is possible, the role of narrative, and the objectivity of historical interpretation. The fact-value distinction and debates about objectivity versus interpretation cross all humanistic disciplines. **Deconstruction** (Derrida, 1930–2004) and **post-structuralism** (Foucault, Deleuze) challenge the possibility of stable meaning and objective truth, raising metaphilosophical questions about the nature of philosophy itself.

Applied philosophy. Philosophical ethics, epistemology, and political theory are used in professional contexts: medical ethics (bioethics committees), business ethics, environmental ethics, engineering ethics, computer ethics, and public policy. The methods of conceptual analysis, argumentation, and justification provide critical tools across fields. The **critical thinking framework**—identifying assumptions, evaluating evidence, constructing arguments, and recognizing fallacies—is a transferable skill.

Methodological pluralism. Different disciplines use different methods—experimental, interpretive, formal, narrative—and philosophy helps articulate their strengths and limitations. **Interdisciplinarity** requires bridging vocabularies and standards of evidence. **Probability and uncertainty** cross disciplinary boundaries: Bayesian epistemology, risk analysis, and decision theory offer unified approaches.

For review, be able to: explain how neuroscience relates to philosophical questions about free will; describe the evolutionary debunking argument in metaethics; distinguish interpretive and causal approaches in social science; explain how game theory models strategic interaction; discuss objectivity in historiography; apply philosophical reasoning to a concrete interdisciplinary case study; identify assumptions in empirical research.

Section summary Philosophy integrates with natural sciences (neuroethics, evolutionary debunking), social sciences (critical theory, rational choice), humanities (historiography, deconstruction), and professional practice (applied ethics). Methodological pluralism and Bayesian/probabilistic frameworks bridge disciplinary boundaries. Critical thinking skills drawn from philosophy transfer across domains.