COG 250: Introduction to Cognitive Science

Final Exam Study Handout

Monday April 20, 2026 | 9:00 AM – 12:00 PM | 3 hours, no aids

Exam Format: Section 1 — Short Answer: choose **4 of 10** terms (10 marks each = 40 marks). Section 2 — Essay: choose **1 of 3** prompts (60 marks). The exam is cumulative but weighted toward post-Test 2 content.

Strategy: Read the essay questions before starting Section 1. Choose terms that feed into your essay, then cite those terms in your essay response. For each term, aim to cover: a clear definition, why it matters theoretically, and a criticism or connection.

Novel Angle entries suggest connections or criticisms not explicitly covered in class — use these to demonstrate independent thinking and earn top marks.

Section 1: Short Answer Terms (Answer 4 of 10)

Group 1: Computational & Classical Theories of Mind

1. Computationalism

Definition: The view that cognition is computation: mental processes are formal symbol manipulations following logical rules. Encompasses both GOFAI (Good Old-Fashioned AI) and Computational Functionalism. Moved away from behaviourism by focusing on the functions of internal mental states rather than observable behaviour alone.

Relevance: Central to the debate over whether the mind is a computer. Underlies both GOFAI and Connectionism. Inspired by linguistics — the idea that cognition is rule-like and fundamentally language-like.

Novel Angle: *Large Language Models (e.g. GPT-4) are arguably the most sophisticated computationalist systems ever built, yet they still fail at reliable causal reasoning and grounded meaning — arguably confirming that symbol manipulation alone cannot produce genuine understanding. This connects to Searle's Chinese Room argument, which may not have been covered explicitly in class.*

You could mention these experiments/examples:

- **Turing Test (Turing, 1950):** the imitation game as an operational test for machine intelligence; widely used to illustrate what computationalism predicts and where it falls short.
- **Winograd's SHRDLU (1972):** an early GOFAI program that manipulated blocks using natural language. It worked in a tiny controlled "microworld" but collapsed when the domain expanded, illustrating the frame problem directly.
- **Searle's Chinese Room (1980):** a thought experiment showing that syntactic symbol manipulation does not produce semantic understanding — a direct challenge to computationalism's core claim.

2. Ghost in the Machine

Definition: Descartes' challenge to Hobbes: if the mind is purely mechanical, it cannot possess original meaning or genuinely reason toward truth. The "ghost" is the subjective, meaningful mind; the "machine" is the physical body or brain. The problem of original meaning: matter cannot have meaning intrinsically — it can only have derived meaning through association with mental states.

Relevance: Foundational challenge to computationalism and any purely physicalist theory of mind. Hobbes attempted to resolve it with intentionality, but Descartes argued this only pushes the problem back. Leads directly into the Hard Problem of Consciousness.

Novel Angle: *Gilbert Ryle (who coined the phrase) used it to critique Cartesian dualism, arguing the ghost/machine split is itself the mistake — a "category error." His critique applies equally well to computationalism. You could*

also connect to the AI alignment problem: we can build increasingly capable machines, but we cannot install genuine values into them. This is arguably a modern, applied version of the ghost-in-the-machine problem.

You could mention these experiments/examples:

- **Descartes vs. Hobbes (1640s):** the original philosophical exchange. Worth citing as the "founding case" that frames the debate. Not an experiment, but a historically grounded theoretical dispute you can use to anchor your answer.
- **Searle's Chinese Room (1980):** shows that syntactic manipulation without semantic grounding is exactly the ghost-in-the-machine problem in action.

3. Morphological Computation

Definition: The idea that the body's physical structure performs computational work, offloading cognitive labour from the brain onto the organism's morphology and its interactions with the environment. The body is not a passive carrier of the brain — it is part of the cognitive system.

Relevance: Challenges brain-centric views of cognition. Central to 4E (embodied, embedded, enacted, extended) cognition and dynamical systems approaches. Illustrates that cognitive processing is distributed across brain, body, and environment.

Novel Angle: *Octopus cognition provides a compelling empirical case: octopi have roughly 500 million neurons, about two-thirds distributed across their arms rather than centralized in the brain. Each arm can solve sensorimotor problems semi-autonomously, without waiting for instructions from the central nervous system. This is a biological example of morphological computation that goes beyond robots and grounds the argument empirically.*

You could mention these experiments/examples:

- **Passive Dynamic Walkers (McGeer, 1990):** mechanical legs that walk down slopes using only gravity and body shape, with no motor control or neural input. A clean demonstration that morphology alone can "compute" locomotion.
- **Subsumption Architecture Robots (Brooks, 1986):** robots that navigate complex environments using layered behavioural processes and real-time body-environment feedback, without centralized representation. Informed the design of the Roomba.
- **Phantom Limb Syndrome:** changes in body schema alter proprioceptive and cognitive processing, showing that morphology is constitutive of cognition, not merely supportive of it. Ramachandran's mirror box therapy is a well-known experimental case.

Group 2: Predictive Processing Framework

4. Bayesian Brain Hypothesis

Definition: The brain functions as a Bayesian inference engine: it builds probabilistic models (priors) of the world and continuously updates them based on incoming sensory evidence to reduce uncertainty. The brain is top-down first — it predicts, then checks against sensory input, rather than passively receiving the world.

Relevance: Foundational to Predictive Processing and the Free Energy Principle (Friston). Directly challenges classical bottom-up models of perception. The bacteria-in-water example: a bacterium resists dissipation by effectively "predicting" its own location, illustrating the principle at a biological systems level.

Novel Angle: *Placebo and nocebo effects provide a striking empirical case that is rarely framed this way in introductory courses: the brain's prior expectations about a treatment can override actual sensory input, with clinically measurable consequences. A sugar pill relieves pain because the predictive system expects it to. This shows Bayesian priors overriding bottom-up signals in a real-world medical context.*

You could mention these experiments/examples:

- **Hollow Mask Illusion:** observers consistently perceive the concave back of a mask as convex — a top-down prior ("faces are convex") overrides bottom-up depth cues. Attenuated in individuals with schizophrenia, who

show weaker priors, supporting the Bayesian account.

- **Binocular Rivalry Studies:** when different images are shown to each eye, conscious perception oscillates between them. PP explains this as hypothesis selection between competing predictions rather than passive sensory integration.
- **Schachter & Singer Emotion Study (1962):** participants given adrenaline without explanation constructed different emotional experiences depending on a confederate's behaviour (happy or angry). Illustrates how prior contextual interpretation shapes affective experience — later reframed by PP theorists as support for a one-factor (rather than two-factor) account of emotion.

5. Error Dynamics

Definition: A predictive system is sensitive not just to the amount of prediction error, but to the *rate* at which error is reduced over time, relative to prior expectations (Miller et al., 2023). Organisms optimize error reduction trajectories, not just current error levels. Two strategies: (1) change your predictive model — data compression, selecting for what is invariant; (2) change your behaviour to reduce sensory uncertainty — data particularization, introducing variation for error correction.

Relevance: The central mechanism within Predictive Processing for explaining adaptive behaviour, curiosity, and motivation. Explains why organisms tolerate short-term uncertainty spikes: because doing so produces larger long-term reductions in error.

Novel Angle: *Addiction can be reframed as a failure of error dynamics: addictive substances produce artificially steep, rapid reductions in prediction error. The predictive system learns to expect this rate of reduction, and ordinary experience comes to feel insufficiently rewarding by comparison. This frames addiction as a computational pathology that follows directly from PP's core mechanism — not a framing typically introduced in class.*

You could mention these experiments/examples:

- **Miller et al. (2023) "Morbid Curiosity" paper:** the primary course reading that introduces the error dynamics framing. The horror film and true crime paradigms are the running empirical cases.
- **Rescorla-Wagner Model (1972):** foundational learning model showing that animals track prediction error rather than reward per se. An early precursor to the PP account of learning.
- **Schultz et al. Dopamine Study (1997):** demonstrated that dopamine neurons respond to prediction errors, not rewards themselves. Strong neuroscientific evidence that the brain tracks error dynamics as described by PP.

6. Morbid Curiosity

Definition: The drive to engage with threatening, dark, or death-related content. In the PP framework, this is an epistemic strategy: we tolerate short-term uncertainty in order to build generative models for high-stakes, harm-relevant scenarios in a low-risk context (Miller et al., 2023). Also an emotion-regulation strategy and a driver of metacognition and interoceptive awareness.

Relevance: Resolves the PP curiosity paradox: why would an uncertainty-minimizing system seek out uncertainty? Because prioritizing long-term reductions in uncertainty over short-term ones means it is sometimes adaptive to face horror in order to learn regularities we cannot easily access in everyday life.

Novel Angle: *Connect to Buddhist maranasati (mindfulness of death): if horror films function as controlled exposure to mortality-related uncertainty, then death meditation is the deliberate, structured version of the same epistemic strategy. Mark Miller has published work exploring this exact connection between PP and contemplative practice — citing this bridge shows genuine integration of ideas beyond the lecture material.*

You could mention these experiments/examples:

- **Scrivner et al. Haunted House Study (2021):** participants in a haunted house who reported enjoying the experience showed better-calibrated fear responses and faster physiological recovery. Supports the emotion-regulation and interoceptive-tuning account of morbid curiosity.

- **True Crime Consumption Research:** surveys show true crime listeners are disproportionately women who report using the genre to learn threat-detection and safety cues — consistent with the "epistemic grip" hypothesis (i.e. horror as a low-stakes arena for learning about high-stakes scenarios).
- **Litman Dysphoric Curiosity Scale (2005):** measures individual differences in tolerance for aversive uncertainty. Scores predict engagement with morbid content, linking personality research to the PP account.

Group 3: Relevance Realization

7. Relevance Realization

Definition: A framework explaining how organisms operating in large, ill-defined worlds selectively attend to and act on what matters for their survival and goals (Vervaeke; Jaeger et al., 2024). Grounded in three opponent processes: **Autopoiesis** (the system maintains itself by seeking conditions for its own survival), **Anticipation** (internal predictive models orient the agent toward goals), and **Integrated Adaptation** (the agent-arena relationship continuously tightens through action and feedback).

Relevance: Argues that agency is fundamentally non-computational: unlike machines that operate in pre-specified "small worlds," organisms navigate open-ended problem spaces where relevance cannot be formalized in advance. Connects to PP via precision weighting — what you find salient is what reduces the most uncertainty. Relevance is not calculated; it emerges from caring about one's own survival.

Novel Angle: *The frame problem in classical AI (Dreyfus, 1972): early AI systems could not determine which facts remained relevant when the world changed, because relevance was never formalizable. Relevance Realization can be framed as cognitive science's answer to the frame problem — and the reason humans don't have this problem isn't that we run a better algorithm, it's that we are autopoietic, self-organizing systems for whom relevance emerges from our own existence.*

You could mention these experiments/examples:

- **Jaeger et al. (2024):** the primary theoretical paper for this section. Introduces autopoiesis, anticipation, and integrated adaptation as the three opponent processes grounding RR in biological agency.
- **Dreyfus's Critique of AI ("What Computers Can't Do," 1972):** documented systematic failures of GOFAI on open-ended, real-world tasks. Empirically grounds the "large world" problem that RR is designed to address.
- **Gibson's Affordance Research (1979):** ecological psychology experiments showing that perceived action possibilities are shaped by body dimensions and skill level, not just objective environmental features. Supports the agent-arena relational view central to RR.

Group 4: Consciousness

8. Hard Problem of Consciousness

Definition: Chalmers' distinction between the "easy" and "hard" problems: the easy problems (tractable in principle) concern which brain systems perform which cognitive functions. The Hard Problem asks *why* physical processes give rise to subjective, first-person experience at all — why there is "something it is like" to be conscious.

Relevance: No mechanistic account, including Predictive Processing, fully closes the explanatory gap. PP can explain attention, precision weighting, and the functional architecture of consciousness very well, but it cannot explain why any of this is *felt*. Chalmers argues the easy problems are tractable; the Hard Problem is of a different order entirely.

Novel Angle: *Chalmers' meta-problem of consciousness (2018 — likely not covered in class): instead of asking why we are conscious, ask why we think there is a Hard Problem at all. A PP system would predict its own puzzlement about consciousness, since introspective reports are themselves just more predictions. This either dissolves the Hard Problem or reveals that our intuitions about qualia are constructed rather than foundational — a genuinely novel reframing.*

You could mention these experiments/examples:

- **Mary's Room Thought Experiment (Jackson, 1982):** Mary knows every physical fact about colour but has never seen red. When she does, does she learn something new? If yes, qualia are not fully captured by physical or functional descriptions.
- **Neural Correlates of Consciousness (NCC) Research (Koch et al.):** fMRI and EEG studies identify brain correlates of conscious perception, but correlates are not explanations. Empirically illustrates the easy/hard distinction: we can map the neural mechanisms without explaining why they feel like anything.
- **Global Workspace Theory Experiments (Baars; Dehaene):** masking and dual-task studies reveal a sharp threshold between conscious and non-conscious processing ("ignition" in frontoparietal networks). Addresses the easy problem while leaving the hard problem untouched.

9. Qualia

Definition: The subjective, felt qualities of conscious experience — "what it is like" to see red, feel pain, or taste coffee. First-person properties that resist full third-person, objective description. Descartes: qualia are secondary properties (colour, taste, smell) that cannot be measured mathematically despite being epistemically foundational — you cannot reason toward truth without knowing what truth *means*.

Relevance: Central to both the Ghost in the Machine problem and the Hard Problem. The existence of qualia is what makes Descartes' challenge to computationalism so difficult: if meaning requires qualia, and qualia cannot be reduced to symbol manipulation, then no purely mechanical system can genuinely reason.

Novel Angle: *Synesthesia provides a useful empirical angle: synesthetes involuntarily experience cross-modal qualia (e.g. hearing music triggers the experience of colour). Neuroimaging confirms cross-modal activation. This suggests qualia are not fixed outputs of sensory processing but are partly constructed — which aligns with PP (qualia as the felt texture of the brain's best hypothesis) while complicating the claim that qualia are epistemically foundational or incorrigible.*

You could mention these experiments/examples:

- **Mary's Room (Jackson, 1982):** the primary thought experiment for qualia (see Hard Problem above). Mary's new knowledge upon seeing red for the first time is argued to be qualitative, not physical.
- **Inverted Spectrum Thought Experiment (Locke; Chalmers):** could two people's internal colour experiences be systematically inverted while their behaviour remains identical? If conceivable, qualia are not reducible to functional role.
- **Synesthesia Research (Cytowic, 1989; Ward et al.):** neuroimaging studies confirm genuine cross-modal activation in synesthetes. Demonstrates that the felt quality of experience (e.g. colour) can be triggered by an unrelated stimulus (sound), empirically supporting the view that qualia are partly constructed by the cognitive system.

10. Philosophical Zombies

Definition: A thought experiment: a being physically and behaviourally identical to a human but with no inner subjective experience. The zombie reports consciousness, behaves normally, and passes every functional test — but there is nothing it is like to be it. It is, as one might say, "all machine and no ghost."

Relevance: If zombies are *conceivable*, consciousness may be epiphenomenal (causally inert and not essential for behaviour). If zombies are *impossible*, consciousness does genuine causal work and cannot be separated from the physical. Directly challenges functionalist and computationalist theories: a functionalist must either deny zombie conceivability or accept that consciousness does nothing.

Novel Angle: *A PP-based critique: the zombie argument assumes we have reliable introspective access to our own qualia. But if PP is right, introspective reports are themselves just more predictions — not ground-truth access to experience. If we cannot trust introspection, the conceivability argument loses much of its force. This connects the zombie debate back to the PP framework covered in the course and challenges the argument on its own terms.*

You could mention these experiments/examples:

- **Chalmers, "The Conscious Mind" (1996):** the canonical philosophical defence of zombie conceivability and the primary source for this argument.
- **Libet et al. Readiness Potential Study (1983):** brain activity predicts voluntary movement approximately 300–500 ms before participants report the conscious intention to move. Raises empirical questions about whether conscious experience is causally efficacious, touching the epiphenomenalism worry directly.
- **Change Blindness Studies (Simons & Levin, 1998):** participants fail to notice large changes in a scene during a brief interruption, suggesting conscious experience is far sparser than introspection implies. Undermines the reliability of introspective reports — relevant to the PP critique of the zombie argument.

Section 2: Essay Questions (Answer 1 of 3)

Each essay should have: **(1)** a clear thesis statement, **(2)** a roadmap of your argument at the end of the first paragraph, **(3)** theoretical and empirical support, **(4)** critical evaluation of the strongest opposing view. You may cite your Section 1 short answers in your essay — but be sure to clearly explain how you are using each concept.

Q1: Is the Mind a Computer? Critically evaluate the computational theory of mind. In what ways does computationalism succeed or fail in explaining human cognition?

Cover: GOFAI & Connectionism, the Paradox of Mechanical Reasoning, original meaning, Ghost in the Machine, and at least one alternative framework (Embodied Cognition, Dynamical Systems, or PP). Novel angle: use LLMs or the AI alignment problem as a contemporary test case for computationalism's limits.

Q2: Predictive Processing Framework. What is Predictive Processing, and how does it challenge traditional models of perception and cognition?

Cover: Bayesian Brain Hypothesis, Free Energy Principle, top-down vs. bottom-up processing, precision weighting, Error Dynamics, and at least one connection to consciousness or morbid curiosity. Novel angle: placebo/nocebo effects as an empirical case, or addiction as a failure of error dynamics.

Q3: Future of Cognitive Science. Which model of cognition provides the best framework for understanding the mind? Defend your position with examples from the readings.

Cover: Computationalism, Connectionism, Dynamical Systems, Embodied Cognition, Predictive Processing. Pick one to defend and engage seriously with its rivals. Novel angle: ground your argument in a specific empirical case — the hollow mask illusion (PP), passive dynamic walkers (Embodied), or LLM failures (Computationalism).

Conceptual Connections Quick Reference

Term	Connect to ...
Computationalism	Ghost in the Machine → Qualia → Hard Problem → Phil. Zombies. *Novel:* LLMs, Chinese Room, AI alignment.
Ghost in the Machine	Qualia, original meaning, Hobbes vs. Descartes, Hard Problem. *Novel:* Ryle's category error, AI alignment.
Morphological Computation	4E Cognition, Dynamical Systems, Subsumption Architecture. *Novel:* octopus cognition, phantom limb.
Bayesian Brain / PP	Error Dynamics, Free Energy Principle, Morbid Curiosity, Consciousness. *Novel:* placebo/nocebo.
Error Dynamics	Morbid Curiosity, Precision Weighting, curiosity paradox. *Novel:* addiction as hijacked error dynamics.
Morbid Curiosity	Error Dynamics, PP curiosity paradox, emotion regulation. *Novel:* Buddhist *maranasati* / death meditation.
Relevance Realization	PP (precision weighting), Autopoiesis, Affordances, 4E. *Novel:* Dreyfus frame problem.
Hard Problem	Qualia, Ghost in the Machine, Phil. Zombies. *Novel:* Chalmers' meta-problem.
Qualia	Hard Problem, Ghost in the Machine, PP-as-construction. *Novel:* synesthesia research.
Philosophical Zombies	Hard Problem, Qualia, Epiphenomenalism. *Novel:* PP introspection-as-prediction undermines conceivability.