A Brief Introduction to the Formal Foundations of Attention, Perception, and Rational Inattention

Prepared for UCSB Econ 278H, Spring 2026 (Daniel Martin)

Abstract

This article reviews the Econ 278H lecture sequence on attention, perception, and rational inattention. Subjective perception asks whether observed choices can be represented by Bayesian expected utility with private information. Rational inattention adds a prior stage in which the decision maker chooses their private information. The central revealed restrictions are NIAS for optimal action choice, NIAC for optimal attention choice with general costs, NIS for capacity-constrained learning, ILR for Shannon costs, and NIAC$^+$ for optimal choice of costly experiments. The applications include discrimination tests, AI oversight, rational inattention in games, and costly experiments.

Keywords: attention; subjective perception; rational inattention; revealed preference; NIAS; NIAC; Shannon entropy; AI oversight; costly experiments

1. Primitives: States, Actions, and Revealed Beliefs

Let Ω be a finite state space and let $\mu^* \in \Delta(\Omega)$ denote the objectively correct prior. A decision problem i consists of an action set A_i and a payoff function $u_i : A_i \times \Omega \to \mathbb{R}$. The observed data are state-dependent stochastic choice probabilities

$$P_i(a, \omega) = \mu^*(\omega) P_i(a \mid \omega), \qquad a \in A_i,\ \omega \in \Omega.$$

The marginal choice probability is $P_i(a) = \sum_\omega P_i(a, \omega)$.

Definition 1 (Revealed posterior)**.** *Whenever $P_i(a) > 0$, the revealed posterior associated with action a is*

$$\bar{\gamma}_i^a(\omega) = \Pr(\omega \mid a) = \frac{P_i(a, \omega)}{P_i(a)} = \frac{\mu^*(\omega) P_i(a \mid \omega)}{\sum_{\omega'} \mu^*(\omega') P_i(a \mid \omega')}.$$

Equivalently, the revealed information structure is

$$\bar{\pi}_i(\bar{\gamma}_i^a \mid \omega) = P_i(a \mid \omega), \qquad \bar{\pi}_i(\bar{\gamma}_i^a) = P_i(a).$$

These objects let one move from observed behavior to restrictions on unobserved perception.

The notation also clarifies why the prior in the lectures is written as μ^*. It is the correct state distribution from the analyst's perspective. A subjective prior, if different, is an additional behavioral object. In many revealed-preference tests the prior is either observed, elicited, or assumed correct, so the revealed posterior $\bar{\gamma}_i^a$ is pinned down from the joint distribution of states and actions. When the subjective prior is not observed, the same inequalities become partial-identification restrictions rather than point restrictions.

2. Subjective Perception as Bayesian Expected Utility

A Bayesian expected utility representation with private information consists of a utility function, a subjective prior μ_i, an information structure, and a decision rule. In the slide notation, the information structure is a map $\pi_i : \Omega \to \Delta(\Gamma_i)$, where $\pi_i(\gamma \mid \omega)$ is the probability of receiving posterior γ in state ω. Let

$$\pi_i(\gamma) = \sum_{\omega \in \Omega} \mu_i(\omega) \pi_i(\gamma \mid \omega)$$

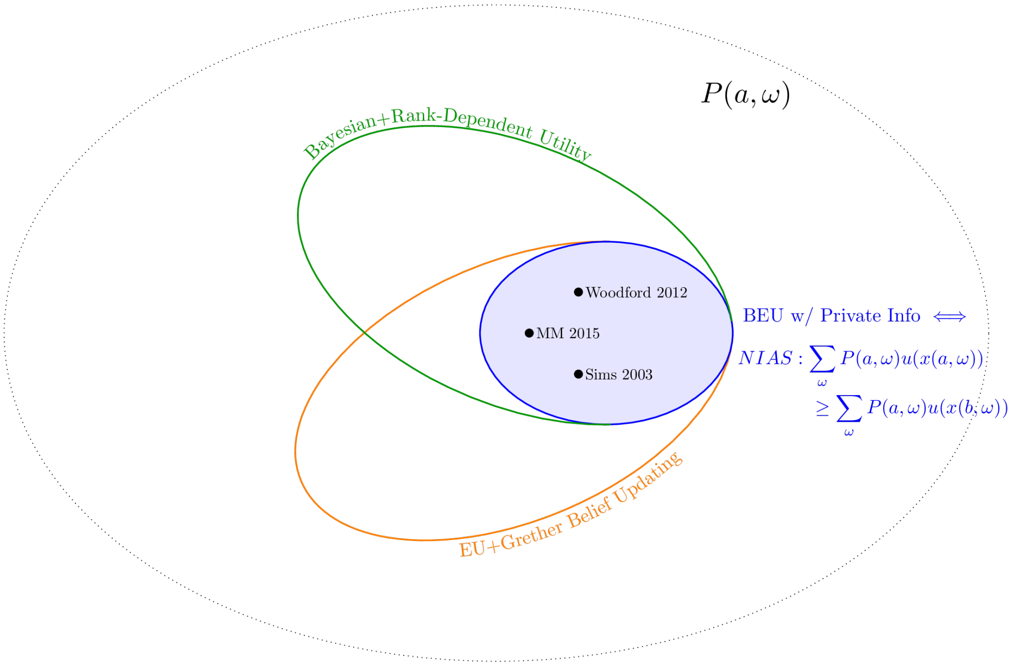

Figure 1: Subjective perception begins with the joint distribution of states and actions and asks whether the implied action-labeled posteriors can be rationalized by Bayesian expected utility.

denote the unconditional probability of posterior γ. A decision rule $\sigma_i : \Gamma_i \to \Delta(A_i)$ generates the data if

$$P_i(a,\omega) = \mu_i(\omega) \sum_{\gamma \in \Gamma_i} \pi_i(\gamma \mid \omega)\sigma_i(a \mid \gamma).$$

Bayesian updating requires that, whenever $\pi_i(\gamma) > 0$,

$$\gamma(\omega) = \frac{\mu_i(\omega)\pi_i(\gamma \mid \omega)}{\sum_{\upsilon \in \Omega} \mu_i(\upsilon)\pi_i(\gamma \mid \upsilon)}.$$

Maximization requires

$$\pi_i(\gamma) > 0 \text{ and } \sigma_i(a \mid \gamma) > 0 \quad \Rightarrow \quad a \in \arg\max_{b \in A_i} \sum_{\omega \in \Omega} u_i(b,\omega)\gamma(\omega).$$

The analyst need not observe π_i or σ_i. The reason is that optimal action choice imposes a direct inequality on the joint distribution of chosen actions and states.

Definition 2 (NIAS)**.** *State-dependent stochastic choice data P_i satisfy No Improving Action Switches if, for all $a, b \in A_i$,*

$$\sum_{\omega \in \Omega} P_i(a,\omega)\left(u_i(a,\omega) - u_i(b,\omega)\right) \geq 0.$$

NIAS says that the observations in which action a was chosen cannot be improved by changing all of them to a fixed alternative b. Equivalently, using the revealed posterior $\bar{\gamma}_i^a$,

$$\sum_{\omega} \bar{\gamma}_i^a(\omega)u_i(a,\omega) \geq \sum_{\omega} \bar{\gamma}_i^a(\omega)u_i(b,\omega) \quad \text{for all } b \in A_i.$$

Caplin and Martin (2015) show that this condition is not merely necessary. With finite states and actions, it is the core revealed-preference restriction for imperfect perception: if every chosen action is optimal at its revealed posterior, the data can be rationalized by a private information structure.

The discrimination application uses the same inequality with group-indexed data. Let $g \in G$ index groups and let $P_g(a, \omega)$ denote choices and later outcomes. A statistical-discrimination representation allows P_g to differ across groups because information structures or state distributions differ, but it holds fixed the relevant payoff function $u(a, \omega)$. Preference-based prejudice is revealed when no common payoff vector can satisfy the system of NIAS inequalities across groups and decision problems. This is the formal reason that simple group gaps are insufficient: a gap in $P_g(a)$ may reflect different objective state distributions, different signals, or different utilities. The robust test of Martin and Marx (2022) asks whether the utility explanation is forced by the inequalities.

3. General Rational Inattention

Rational inattention adds an information-acquisition stage. An information structure π_i is now chosen before the decision rule. Bayes plausibility is equivalently written as

$$\sum_{\gamma \in \Gamma_i} \pi_i(\gamma)\gamma = \mu_i.$$

The gross value of information structure π in decision problem i is

$$G_i(\pi) = \max_{\sigma_i} \sum_{\omega \in \Omega} \mu_i(\omega) \sum_{\gamma \in \Gamma} \pi(\gamma \mid \omega) \left(\sum_{a \in A_i} \sigma_i(a \mid \gamma) u_i(a, \omega) \right).$$

A general rational-inattention representation consists of a cost function $K : \Delta(\Omega) \times \Delta(\Delta(\Omega)) \to \mathbb{R}_+ \cup \{\infty\}$ such that the chosen information structure π_i solves

$$\pi_i \in \arg\max_{\pi} \left\{ G_i(\pi) - K(\mu_i, \pi) \right\}, \qquad \sum_{\gamma \in \Gamma} \pi(\gamma)\gamma = \mu_i.$$

The cost is usually normalized so that free inattention has zero cost: $K(\mu, \delta_\mu) = 0$. It is also natural to impose monotonicity in the Blackwell order: if π Blackwell dominates π', then $K(\mu, \pi) \geq K(\mu, \pi')$ (Blackwell, 1953).

The revealed-preference restriction is an Afriat-style cycle condition. Suppose the analyst recovers the revealed information structures $\{\bar{\pi}_i : i \in D\}$. If the true π_i is optimal in problem i, and $\bar{\pi}_i$ is the revealed structure generated by the data, then the restriction can be written using the $\bar{\pi}_i$'s:

$$G_i(\bar{\pi}_i) - K(\mu, \bar{\pi}_i) \geq G_i(\bar{\pi}_j) - K(\mu, \bar{\pi}_j),$$

when the subjective prior is common across problems. Summing these inequalities around any cycle eliminates the unobserved cost levels.

Definition 3 (NIAC)**.** *For a common subjective prior μ, observed information structures satisfy No Improving Attention Cycles if, for every finite cycle $i_1, \ldots, i_m, i_{m+1} = i_1$,*

$$\sum_{t=1}^{m} \left(G_{i_t}(\bar{\pi}_{i_t}) - G_{i_t}(\bar{\pi}_{i_{t+1}}) \right) \geq 0.$$

Caplin and Dean (2015) show that NIAS and NIAC characterize costly information acquisition under broad assumptions. NIAS handles the second-stage action choice; NIAC handles the first-stage attention choice. This separation is conceptually useful. A dataset can fail because agents misuse their information, or because the information choices cannot be made optimal by any stable cost function.

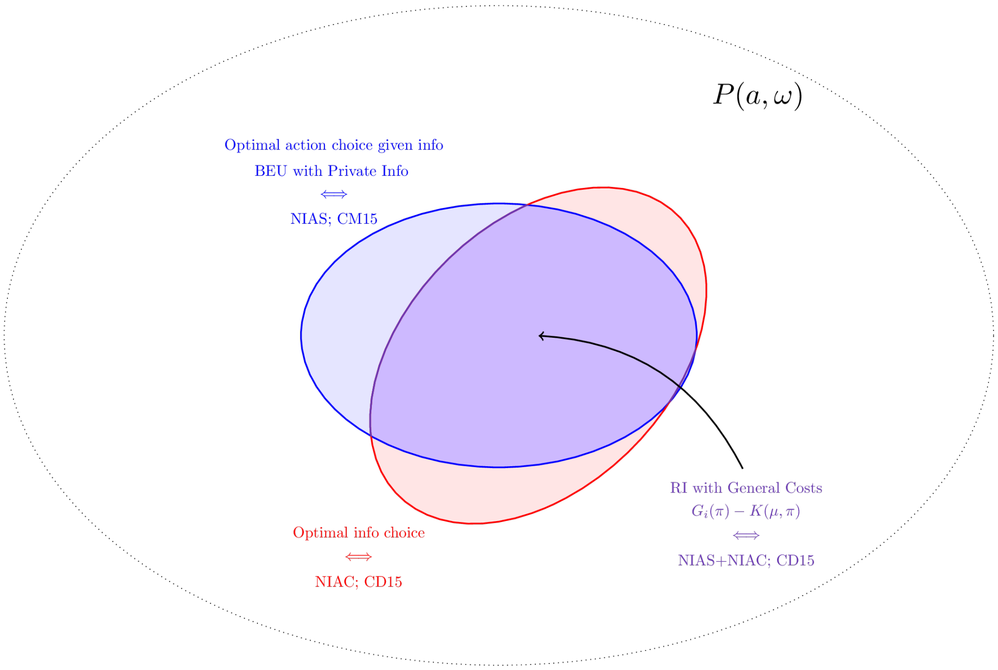

Figure 2: General rational inattention adds optimal choice of information to the Bayesian expected-utility representation.

4. Capacity-Constrained Learning

A capacity-constrained learner faces a hard feasibility set $\Pi \subseteq \Delta(\Delta(\Omega))$ and zero marginal cost inside the set:

$$K(\mu, \pi) = \begin{cases} 0, & \pi \in \Pi, \\ \infty, & \pi \notin \Pi. \end{cases}$$

The decision maker chooses the best feasible experiment,

$$\pi_i \in \arg\max_{\pi \in \Pi} G_i(\pi).$$

If the same feasible set applies across problems, each observed experiment must beat the other observed experiments in its own problem. This is the NIS condition:

$$G_i(\bar{\pi}_i) \geq G_i(\bar{\pi}_j) \qquad \text{for all } i, j \in D,$$

with the appropriate revealed or recovered information structures. In words, once one observed experiment is feasible, transporting it to another problem cannot improve the decision maker's payoff there.

This restriction is sharper than NIAC. NIAC allows a costly but more valuable experiment to be chosen in one problem and a cheaper, less valuable experiment in another. A hard capacity model does not.

5. Shannon Costs and Invariant Likelihood Ratios

The Shannon specification replaces the unknown cost K with mutual information. For SDSC $P(a, \omega)$, the mutual information between states and actions is

$$I_\mu(A; \Omega) = \sum_{\omega \in \Omega} \mu(\omega) \sum_{a \in A} P(a \mid \omega) \log \frac{P(a \mid \omega)}{P(a)}.$$

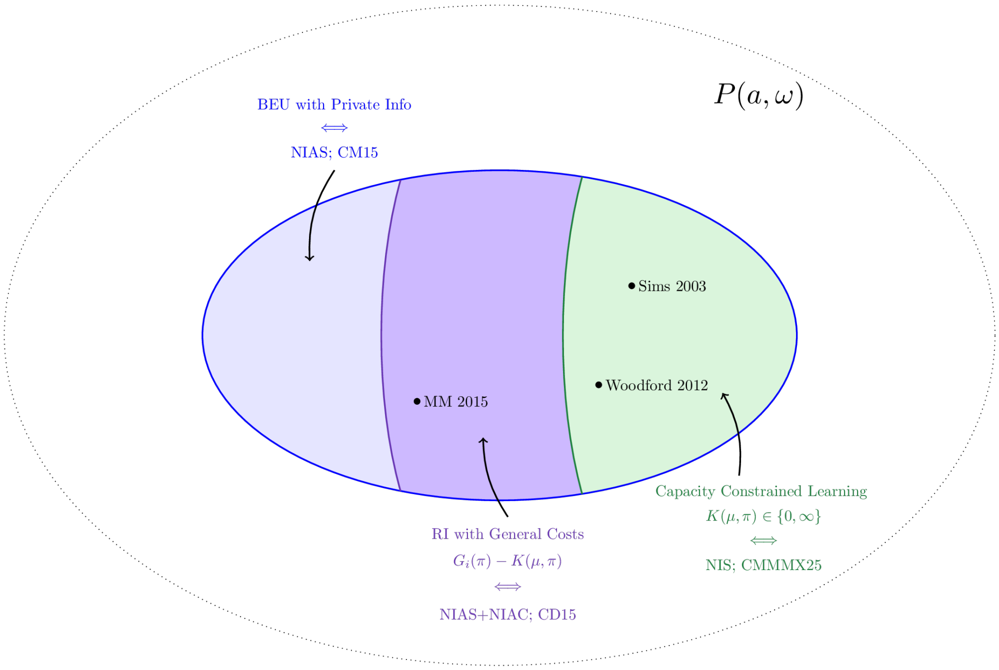

Figure 3: Capacity-constrained learning is a hard-feasibility version of rational inattention and yields sharper testable implications.

Equivalently,

$$I_\mu(A;\Omega) = \sum_\omega \mu(\omega) \sum_a P(a \mid \omega) \log P(a \mid \omega) - \sum_a P(a) \log P(a).$$

With subjective prior μ, the linear Shannon-cost problem is

$$\max_{\{P(\cdot|\omega)\}_{\omega\in\Omega}} \sum_\omega \mu(\omega) \sum_a P(a \mid \omega) u(a,\omega) - \kappa I_\mu(A;\Omega),$$

subject to $P(\cdot \mid \omega) \in \Delta(A)$ for every ω.

For active actions, the first-order conditions imply the familiar logit form

$$P(a \mid \omega) = \frac{P(a)\exp\{u(a,\omega)/\kappa\}}{\sum_{b\in A} P(b)\exp\{u(b,\omega)/\kappa\}}.$$

Unlike a random-utility logit, the unconditional probabilities $P(a)$ are endogenous attention weights. They must be consistent with the prior and the conditional choice rule. This fixed-point structure is why Blahut-Arimoto methods are natural computational tools (Blahut, 1972; Arimoto, 1972).

The invariant likelihood ratio restriction, ILR, is the corresponding testable implication. If actions a and b are chosen with positive probability in states ω and ω', Shannon rational inattention implies

$$\log \frac{P(a \mid \omega)}{P(a \mid \omega')} - \log \frac{P(b \mid \omega)}{P(b \mid \omega')} = \frac{(u(a,\omega) - u(a,\omega')) - (u(b,\omega) - u(b,\omega'))}{\kappa}.$$

ILR is powerful because it removes the endogenous marginals $P(a)$ and state normalizations. It says that posterior likelihood ratios depend on payoff differences and the cost parameter κ, not directly on the prior. The prior still matters through Bayes plausibility, action probabilities, and which actions are active, so this is not a claim that optimal behavior is prior-free.

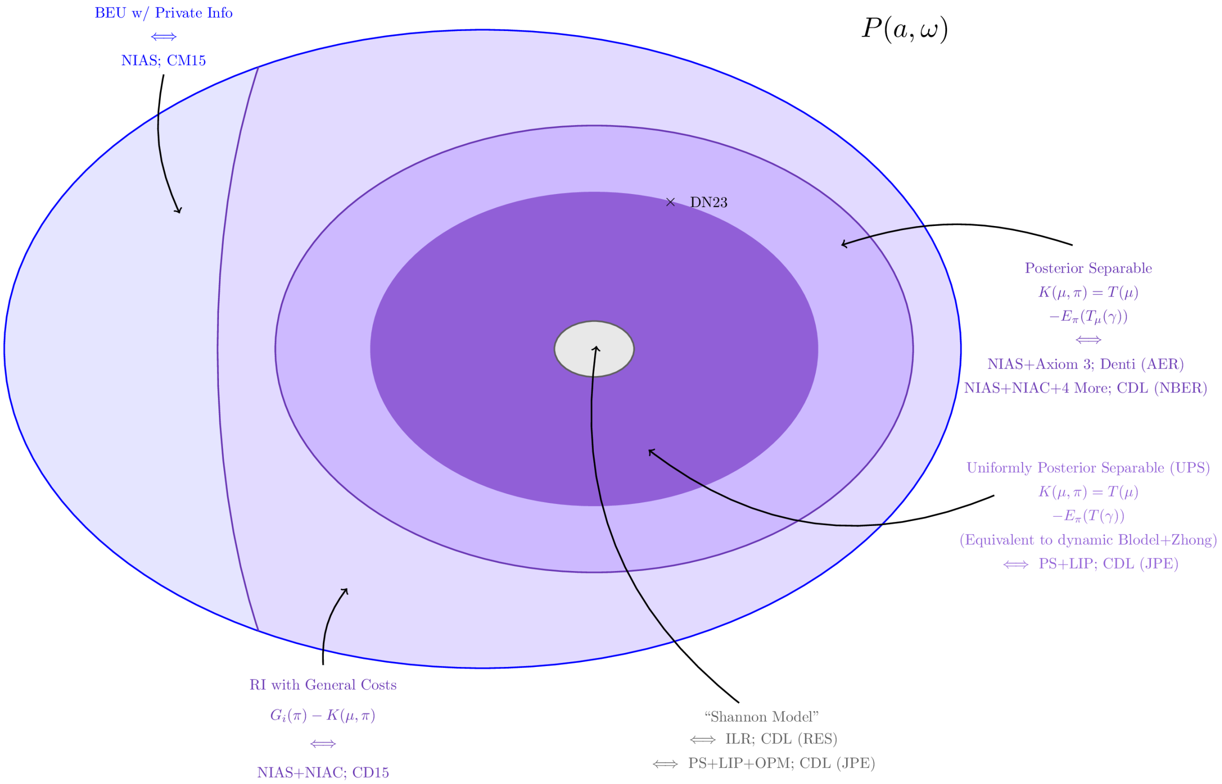

Figure 4: Shannon costs sit inside the broader posterior-separable and uniformly posterior-separable classes.

6. Posterior-Separable and UPS Costs

Shannon mutual information is a special case of posterior-separable information costs. In the slide convention, let $T : \Delta(\Omega) \to \mathbb{R}$ be the potential. At prior μ, a posterior-separable cost is

$$K(\mu,\pi) = T(\mu) - \int_{\Delta(\Omega)} T(\gamma)\, d\pi(\gamma).$$

For Shannon, $T(\gamma) = \kappa H(\gamma)$, so

$$K(\mu,\pi) = \kappa\left(H(\mu) - \int H(\gamma)\, d\pi(\gamma)\right),$$

where $H(\gamma) = -\sum_\omega \gamma(\omega) \log \gamma(\omega)$. Caplin et al. (2022) characterize rationally inattentive behavior using this class and its uniformly posterior-separable generalizations; Bloedel and Zhong (2025) provide an optimization foundation for related UPS-style costs.

The additional structure sharpens identification. General NIAC asks whether some cost function rationalizes the data. Posterior separability asks whether a single potential function T can rationalize the data. Shannon asks whether that potential is negative entropy up to scale. Formally, the hierarchy is Shannon $\subset$ UPS/posterior-separable $\subset$ general RI $\subset$ BEU with private information. Each containment adds empirical content. Moving down the hierarchy makes rationalization easier; moving up makes the model more predictive and computationally disciplined.

7. AI Oversight as a Structural Application

In the AI oversight lecture, the human actions are a_I = call in and a_O = call out, and the states are ω_I = ball in and ω_O = ball out. Distance from the line defines the empirical sample, but it is not a state in the attention problem. The data are joint probabilities $P_t(a,\omega)$, where $t \in \{\text{No AI}, \text{AI}\}$, with revealed posteriors

$$\gamma_{a,t}(\omega) = \frac{P_t(a,\omega)}{P_t(a)}.$$

Correct calls are normalized to utility zero. With Hawk-Eye review, the payoff matrix is

$$U_{\text{AI}}(a,\omega) = \begin{array}{c|cc} & \omega_I & \omega_O \\ \hline a_I & 0 & -1+\eta_O(1+c_O) \\ a_O & -1+\eta_I(1+c_I) & 0 \end{array},$$

where η_I and η_O are challenge rates conditional on incorrect calls, and c_I and c_O are the psychological or reputational penalties from being corrected. Thus $c = 0$ means correction has no direct psychological downside, $c = -1$ exactly offsets the ordinary benefit of correction, and $c < -1$ means being overruled is worse than simply being wrong.

This specification turns observed pre- and post-oversight choice rules into a structural decomposition. Because correct calls are normalized to 0 and ordinary incorrect calls to -1, pre-oversight calls identify the state-specific marginal attention costs κ_I and κ_O under the maintained stable attention technology $K(\mu,\pi)$, not a separate baseline mistake-cost vector. Post-oversight calls identify the oversight penalties c_I and c_O. The key Shannon restrictions are the invariant likelihood-ratio equations

$$\frac{\gamma_{a_I,t}(\omega_I)}{\gamma_{a_O,t}(\omega_I)} = \exp\left(\frac{U_t(a_I,\omega_I) - U_t(a_O,\omega_I)}{\kappa_I}\right), \qquad \frac{\gamma_{a_O,t}(\omega_O)}{\gamma_{a_I,t}(\omega_O)} = \exp\left(\frac{U_t(a_O,\omega_O) - U_t(a_I,\omega_O)}{\kappa_O}\right).$$

Almog et al. (2024) use this logic to show that oversight can improve average accuracy while shifting the composition of errors. The formal point is general: AI oversight changes the payoff matrix faced by humans, and a rationally inattentive agent responds by reallocating attention and errors.

8. Rational Inattention in Games

Strategic settings add endogenous priors. In Martin (2017), a seller has quality $\theta \in \{\theta_L, \theta_H\}$ and chooses price p. A buyer observes price and then chooses how much attention to pay to quality before buying. If the seller strategy is $\rho(p \mid \theta)$ and the prior probability of high quality is μ_0, the interim belief after price p is

$$\mu_p = \frac{\mu_0 \rho(p \mid \theta_H)}{\mu_0 \rho(p \mid \theta_H) + (1-\mu_0)\rho(p \mid \theta_L)}.$$

The price is therefore both a payoff-relevant term and a signal. Given p, the buyer's Shannon attention problem over actions $a \in \{B, N\}$ implies

$$\log \frac{P(B \mid H, p)}{P(B \mid L, p)} - \log \frac{P(N \mid H, p)}{P(N \mid L, p)} = \frac{(u(B,H,p) - u(B,L,p)) - (u(N,H,p) - u(N,L,p))}{\kappa}.$$

The seller anticipates the induced demand schedule. A low-quality seller may mimic a high-quality seller if the buyer's posterior and attention response make the high price profitable. Equilibrium therefore solves two fixed points at once: the seller strategy determines μ_p, and μ_p determines the buyer's attention and demand.

This application also clarifies the phrase "beliefs are the channel." Players do not directly choose opponents' posteriors. They choose payoff-relevant actions that change the objective distribution of types conditional on observable histories. Attention then maps those induced priors into posterior beliefs and actions. The formal equilibrium object is a strategy profile together with attention policies that are sequentially optimal at every induced prior.

9. Costs of Experiments Across Priors

The last lecture separates costs of posterior distributions from costs of experiments. An experiment is a stochastic matrix $S(s \mid \omega)$. Given prior μ, it induces signal probabilities

$$m_{\mu,S}(s) = \sum_{\omega} \mu(\omega) S(s \mid \omega)$$

and posteriors

$$\gamma^{s}_{\mu,S}(\omega) = \frac{\mu(\omega) S(s \mid \omega)}{\sum_{\omega'} \mu(\omega') S(s \mid \omega')}.$$

The same experiment S therefore induces different posterior distributions when the prior changes. Conversely, different experiments can induce the same posterior distribution under a fixed prior.

A cost-of-experiments model assigns cost to the likelihood matrix S, not merely to the induced distribution of posteriors. The experiment itself is not observed, so the empirical restriction must be written in terms of the data. For observed data P_j, define the value in problem i of using the conditional choice probabilities from problem j:

$$G_i^{+}(P_j) = \sum_{a \in A_j} \max_{\hat{a} \in A_i} \sum_{\omega \in \Omega} \mu_i(\omega) P_j(a \mid \omega) u_i(\hat{a}, \omega).$$

This rotates the conditional probabilities $P_j(a \mid \omega)$, not the joint distributions $P_j(a, \omega)$.

Definition 4 (NIAC^{+}). *Observed data satisfy* NIAC^{+} *if, for every finite cycle* $i_1, \ldots, i_m, i_{m+1} = i_1$*,*

$$\sum_{t=1}^{m} \Big(G_{i_t}^{+}(P_{i_t}) - G_{i_t}^{+}(P_{i_{t+1}}) \Big) \geq 0, \qquad i_{m+1} = i_1.$$

If costs attach to posterior distributions, changing the prior changes the object being costed. If costs attach to experiments, the same likelihood matrix can be evaluated across priors. Work on costly experiments and strategic communication makes this distinction central (Denti et al., 2022; de Clippel and Rozen, 2021). It is also important for empirical design: to distinguish posterior costs from experiment costs, one needs variation in priors holding the underlying signal technology fixed. Figure 5 locates this costly-experiment model within the revealed-preference hierarchy.

10. Summary

The course's models differ in their restrictions, but they share the same mathematical structure. Observed behavior is the joint distribution $P(a, \omega)$. Private information is a Bayes-plausible distribution over posteriors. Optimal action choice gives NIAS. Optimal attention choice gives NIAC. A hard common feasible set gives NIS. Shannon entropy gives mutual-information costs and ILR. Posterior separability replaces entropy with a convex potential. Costs of experiments replace posterior distributions with likelihood matrices and lead to NIAC^{+}.

The value of the formalism is not abstraction for its own sake. It explains what can be learned from data. In subjective perception, it identifies when choices are compatible with Bayesian expected utility under private information. In discrimination, it distinguishes preference-based prejudice from statistical discrimination. In AI oversight, it converts changes in error patterns into changes in perceived payoff costs. In games, it shows how prices and actions change priors before attention is chosen. In costly experiments, it isolates whether costs attach to beliefs or to the signal technology itself. The approach treats attention as an economic primitive that is observable through its restrictions even when the underlying mental process is not.

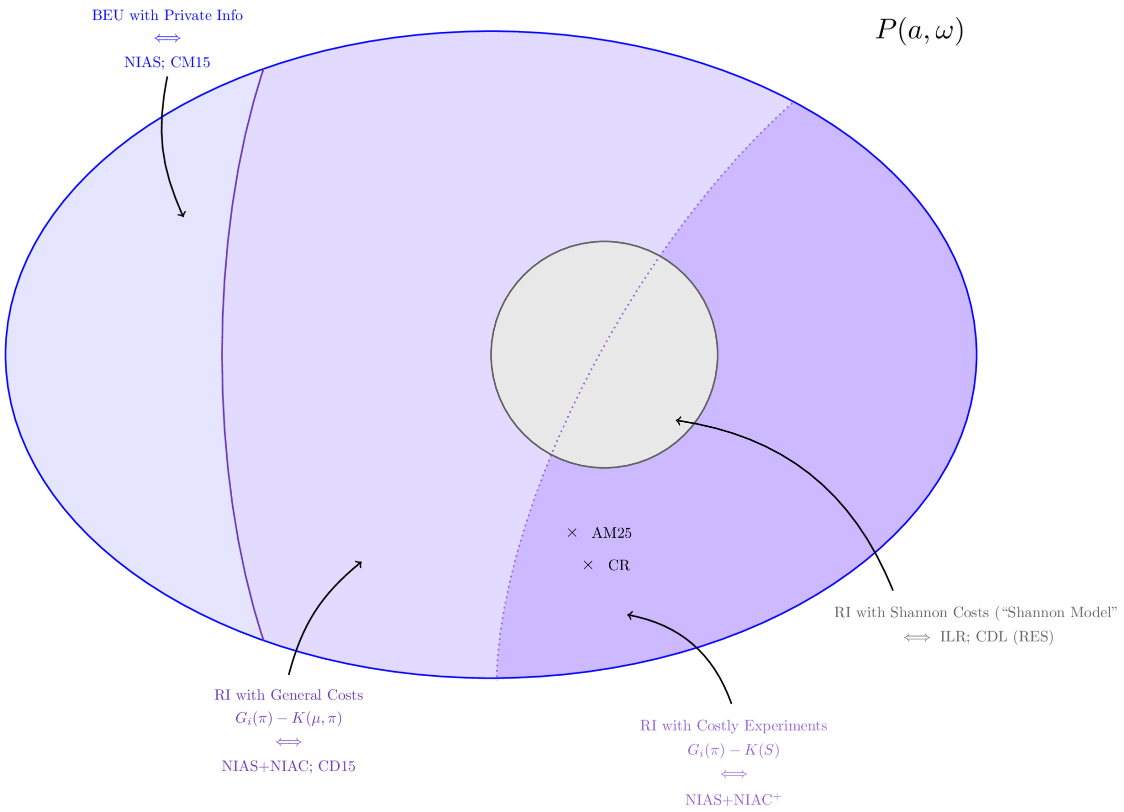

Figure 5: Costly-experiment models assign costs to the likelihood matrix S, yielding NIAC$^+$ restrictions that differ from posterior-cost rational inattention.

References

David Almog, Romain Gauriot, Lionel Page, and Daniel Martin. AI oversight and human mistakes: Evidence from centre court. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 103–105, New York, NY, 2024. Association for Computing Machinery. doi: 10.1145/3670865.3673481. Full working paper available as arXiv:2401.16754, revised February 2025.

Suguru Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1):14–20, 1972. doi: 10.1109/TIT.1972.1054753.

David Blackwell. Equivalent comparisons of experiments. *Annals of Mathematical Statistics*, 24(2): 265–272, 1953. doi: 10.1214/aoms/1177729032.

Richard E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473, 1972. doi: 10.1109/TIT.1972.1054855.

Alexander W. Bloedel and Weijie Zhong. The cost of optimally acquired information. arXiv:2511.05466, working paper, 2025.

Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203, 2015. doi: 10.1257/aer.20140117.

Andrew Caplin and Daniel Martin. A testable theory of imperfect perception. *The Economic Journal*, 125(582):184–202, 2015. doi: 10.1111/ecoj.12130.

Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing Shannon entropy. *Journal of Political Economy*, 130(6):1676–1715, 2022. doi: 10.1086/719276.

Andrew Caplin, Daniel Martin, Philip Marx, Anastasiia Morozova, and Leshan Xu. Testing capacity-constrained learning. Working paper; revised March 6, 2026; revise and resubmit at Experimental Economics; arXiv:2502.00195, 2026.

Geoffroy de Clippel and Kareen Rozen. Communication, perception and strategic obfuscation. Working paper, revised February 2021, 2021.

Mark Dean and Nathaniel Neligh. Experimental tests of rational inattention. *Journal of Political Economy*, 131(12):3415–3461, 2023. doi: 10.1086/725174.

Tommaso Denti, Massimo Marinacci, and Aldo Rustichini. Experimental cost of information. *American Economic Review*, 112(9):3106–3123, 2022.

Daniel Martin. Strategic pricing with rational inattention to quality. *Games and Economic Behavior*, 104:131–145, 2017. doi: 10.1016/j.geb.2017.03.007.

Daniel Martin and Philip Marx. A robust test of prejudice for discrimination experiments. *Management Science*, 68(6):4527–4536, 2022. doi: 10.1287/mnsc.2022.4396.