Supplementary Online Material for: Testing Capacity-Constrained Learning

Andrew Caplin[1] Daniel Martin[2,*] Philip Marx[3]
Anastasiia Morozova[2] Leshan Xu[2]

[1]Department of Economics, New York University, New York, NY, USA.

[2]Department of Economics, University of California, Santa Barbara, Santa Barbara, CA, USA.

[3]Department of Economics, Louisiana State University, Baton Rouge, LA, USA.

[*]Corresponding author: danielmartin@ucsb.edu

ORCID: Caplin 0000-0002-1169-9027, Martin 0000-0001-6483-3923

August 22, 2026

A Relationship to Existing Tests

We now describe NIAS and NIAC, two existing tests of Bayesian learning, and explain their relationship to NIS. Overall NIS implies both NIAS and NIAC, but neither implies NIS, so NIS is strictly stronger than both. It is also strictly stronger than their combination. Throughout, we will continue to illustrate the tests using aggregate choice data from Experiment 1.2 of DN23.

A.1 NIAS: Optimal Actions Given Information

The key assumptions behind many, if not most, models of attention and perception are that decision-makers update beliefs correctly (using Bayes' Rule) and maximize expected utility given some unobservable learning about the state. In other words, they choose optimally given the (private) information that they have acquired.

Caplin and Martin (2015) show that, for utility function $u : X \to \mathbb{R}$ and some private information, SDSC data is consistent with choosing optimally given that information if and only if the *No Improving Action Switches (NIAS)* condition is satisfied. This condition requires that no wholesale switch in actions improves utility. Intuitively, NIAS requires that, given the learning revealed by observed choices, expected utility cannot be improved by simply switching actions. The decision-maker's behavior already reflects how they use their information to distinguish states, and relabeling actions cannot improve outcomes. Formally, utility function u satisfies the NIAS condition with respect to P if for every decision problem $i \in D$ and actions $a, \hat{a} \in A_i$,

$$\sum_{\omega \in \Omega} P_i(a,\omega) u(x_i(a,\omega)) \geq \sum_{\omega \in \Omega} P_i(a,\omega) u(x_i(\hat{a},\omega)), \tag{1}$$

NIS implies NIAS because NIAS is satisfied if and only if the DM satisfies NIS within each decision problem. That is, when $i = j$, the main-text NIS condition becomes:

$$\sum_{a \in A_i} \sum_{\omega \in \Omega} P_i(a,\omega) u(x_i(a,\omega)) = \sum_{a \in A_i} \max_{\hat{a} \in A_i} \sum_{\omega \in \Omega} P_i(a,\omega) u(x_i(\hat{a},\omega)), \tag{2}$$

which is satisfied if and only if NIAS holds.

A.1.1 Example from Dean and Neligh (2023)

To illustrate NIAS we turn to the decision problems, prizes, and aggregate SDSC data from Experiment 1.2 of DN23, which are given in the main text. Within each decision problem i of Experiment 1.2 of DN23, the NIAS condition reduces to

$$\begin{aligned} P_i(a_1,\omega_1) &\geq P_i(a_1,\omega_2), \\ P_i(a_2,\omega_2) &\geq P_i(a_2,\omega_1), \end{aligned}$$

which is clearly satisfied. Note that this condition would hold for any SDSC data in which the diagonal entry of the matrix is larger than the entries in each row, consistent with the incentive to match the action to the state.

A.2 NIAC: Optimal Costly Learning

Models of rational inattention also assume that decision-makers optimally acquire their (private) information based on the cost of that information and the expected utility of choosing in line with that information. Caplin and Dean (2015) characterize the optimal choice of costly information by pairing NIAS with the *No Improving Attention Cycles (NIAC)* condition, which states that utility cannot be improved by cycling choice probabilities between decision problems. In other words, NIAC ensures that a decision-maker cannot increase their gross utility simply by reassigning information structures across decision problems.

Condition 1 (No Improving Attention Cycles (**NIAC**)). *Utility function u satisfies NIAC for P if for any sequence of decision problems $1, 2, \ldots, J \in D$ with convention $J + 1 = 1$,*

$$\sum_{j=1}^{J}\left(\sum_{a\in A_j}\max_{\hat{a}\in A_j}\sum_{\omega\in\Omega}P_j(a,\omega)u(x_j(\hat{a},\omega))\right)\geq$$
$$\sum_{j=1}^{J}\left(\sum_{a\in A_{j+1}}\max_{\hat{a}\in A_j}\sum_{\omega\in\Omega}P_{j+1}(a,\omega)u(x_j(\hat{a},\omega))\right)$$

Note that together, NIAS and NIAC imply that

$$\sum_{j=1}^{J}\left(\sum_{a\in A_j}\sum_{\omega\in\Omega}P_j(a,\omega)u(x_j(a,\omega))\right)\geq$$
$$\sum_{j=1}^{J}\left(\sum_{a\in A_{j+1}}\max_{\hat{a}\in A_j}\sum_{\omega\in\Omega}P_{j+1}(a,\omega)u(x_j(\hat{a},\omega))\right)$$

NIS requires that every possible switch (j to $j+1$) is non-improving, so NIS implies NIAC. Intuitively, if every binary comparison is non-improving, then all cycles must be non-improving.

A.2.1 Example from Dean and Neligh (2023)

To illustrate NIAC we again turn to the decision problems, prizes, and aggregate SDSC data from Experiment 1.2 of DN23, which are given in the main text. If Experiment 1.2 of DN23 consisted of just decision problems 1 and 2, in which the number of probability points was 5 and 40 respectively, the NIAC condition would boil down to determining whether cycling the choice probabilities between these decision problems improves utility. Here, *cycling* means that problem 1 uses the information structure revealed in problem 2 while problem 2 simultaneously uses the information structure revealed in problem 1. NIS strengthens this requirement by ruling out any improving switch in a single direction, not only profitable cycles.

$$P_1 = \begin{array}{cc} & \begin{array}{cc}\omega_1 & \omega_2\end{array} \\ \begin{array}{c}a_1\\a_2\end{array} & \begin{pmatrix}0.37 & 0.20\\0.13 & 0.30\end{pmatrix}\end{array} \qquad P_2 = \begin{array}{cc} & \begin{array}{cc}\omega_1 & \omega_2\end{array} \\ \begin{array}{c}a_1\\a_2\end{array} & \begin{pmatrix}0.38 & 0.17\\0.12 & 0.33\end{pmatrix}\end{array}$$

Under the maintained assumptions of expected utility monotonically increasing in prizes and normalized as discussed in the main text, the value of the revealed signal structure in decision problem 2 is

$$\begin{aligned}&\underbrace{P_2(a_1,\omega_1)u(x_2(a_1,\omega_1))}_{0.38\times 40}+\underbrace{P_2(a_1,\omega_2)u(x_2(a_1,\omega_2))}_{0.17\times 0}\\ &+\underbrace{P_2(a_2,\omega_1)u(x_2(a_2,\omega_1))}_{0.12\times 0}+\underbrace{P_2(a_2,\omega_2)u(x_2(a_2,\omega_2))}_{0.33\times 40}\\ &=28.4.\end{aligned}$$

Switching to the signal structure revealed in problem 1 requires re-optimizing for prizes in problem 2:

$$\begin{aligned}&\max_{a\in\{a_1,a_2\}}[P_1(a_1,\omega_1)u(x_2(a,\omega_1))+P_1(a_1,\omega_2)u(x_2(a,\omega_2))]\\ +\quad&\max_{a\in\{a_1,a_2\}}[P_1(a_2,\omega_1)u(x_2(a,\omega_1))+P_1(a_2,\omega_2)u(x_2(a,\omega_2))]\end{aligned}$$

However, given that NIAS is satisfied in each decision problem, the optimal actions remain the same and so the preceding term equals:

$$\begin{aligned}&\underbrace{P_1(a_1,\omega_1)u(x_2(a_1,\omega_1))}_{0.37\times 40}+\underbrace{P_1(a_1,\omega_2)u(x_2(a_1,\omega_2))}_{0.20\times 0}\\ &+\underbrace{P_1(a_2,\omega_1)u(x_2(a_2,\omega_1))}_{0.13\times 0}+\underbrace{P_1(a_2,\omega_2)u(x_2(a_2,\omega_2))}_{0.30\times 40}\\ &=26.8\end{aligned}$$

Because the optimal actions remain the same and choice probabilities on the diagonal decrease, this switch in learning results in a net loss ($\approx 26.8-28.4=-1.6$).

Recall from the main-text NIS example that the value of the revealed signal structure in (decision) problem 1 is 3.34, whereas the value of the revealed signal structure in problem 2 but using the prize specification of problem 1 is 3.55. Switching from signal structure P_1 to P_2 in problem 1 results in a net gain of utility ($\approx 3.55-3.34=0.21$). As discussed in the main-text NIS example, this produces a violation of the NIS condition, but because there is no overall improvement ($-1.6+0.21<0$), the NIAC condition is satisfied.

B Characterizing Capacity-Constrained Learning

Our main goal is to characterize which learning strategies could have rationally generated the observed data P under the model of capacity-constrained learning. The decision models that we consider involve rational Bayesian learning and subsequent optimal choice. Our notation broadly follows Caplin, Martin, and Marx (2025b), henceforth CMM. As in Kamenica and Gentzkow (2011), we specify conceivable learning as a Bayes-consistent distribution Q of posteriors $\gamma \in \Delta(\Omega)$ with finite support $\Gamma(Q) \equiv \operatorname{supp} Q$. We refer to such distributions of posteriors as information structures, with their set given by:

$$\mathcal{Q} \equiv \{Q \in \Delta(\Delta(\Omega)) \text{ with } |\Gamma(Q)| < \infty \text{ and } \sum_{\gamma \in \Gamma(Q)} \gamma Q(\gamma) = \mu\}.$$

Once learning has taken place, the DM selects a mixed strategy over actions as a function of the posterior, $q(a|\gamma) \in \Delta(A)$. Define $P_{(Q,q)}$ as the hypothetical SDSC that any strategy (Q, q) would generate,

$$P_{(Q,q)}(a, \omega) \equiv \sum_{\gamma \in \Gamma(Q)} q(a|\gamma)Q(\gamma)\gamma(\omega). \tag{3}$$

Then a strategy (Q, q) generates the data P if:

$$P_{(Q,q)} = P \tag{4}$$

We will say that such a strategy *rationalizes* the data if it furthermore arises from optimal choice.

We model the DM's optimization problem in two stages, which we solve using backward induction. In the second stage, given an information structure Q and decision problem A, the DM chooses an action strategy to maximize expected utility. Specifically, given a prize specification x and a posterior γ, define the posterior expected utility as: [1]

$$U(a|\gamma, x) \equiv \sum_{\omega \in \Omega} \gamma(\omega)u(x(a, \omega)), \tag{5}$$

and define the gross expected utility of strategy (Q, q) given u as:

$$g(Q, q|x) \equiv \sum_{\gamma \in \Gamma(Q)} \sum_{a \in \mathcal{A}} Q(\gamma)q(a|\gamma)U(a|\gamma, x).$$

Then in the second stage as a function of learning Q, the DM chooses an action strategy to solve:

$$\underset{q:\Gamma(Q)\to\Delta(A)}{\operatorname{argmax}} \; g(Q, q|x) \tag{6}$$

In what follows, it will also be useful to define the resulting indirect expected utility of an information structure Q in decision problem A given utility function u as:

$$G(Q|A, x) \equiv \max_{q:\Gamma(Q)\to\Delta(A)} g(Q, q|x) \tag{7}$$

In the first stage, the DM chooses an information structure to maximize this indirect expected utility subject to a feasibility constraint $\mathcal{Q}^* \subset \mathcal{Q}$. That is, the DM chooses a learning strategy to solve:

$$\underset{Q \in \mathcal{Q}^*}{\operatorname{argmax}} \; G(Q|A, x) \tag{8}$$

[1] We depart slightly from Caplin, Martin, and Marx (2025b) by parametrizing expected utility as a function of prizes rather than Bernoulli utility u. This is because throughout the experiments in this paper, prizes are varied whereas a subject's utility of money is held fixed.

Example 1: Fixed-Capacity Rational Inattention. In Sims (2003), the decision-maker chooses how to allocate attention across states, but faces a constraint on information processing capacity. Formally, the decision-maker chooses state-conditional action probabilities $P(a|\omega)$ to maximize expected utility, subject to the constraint that mutual information between states and actions cannot exceed a fixed capacity C:

$$I(A;\Omega) = \sum_{a,\omega} P(a,\omega) \log \frac{P(a,\omega)}{P(a)\mu(\omega)} \leq C.$$

The capacity C is exogenously fixed—it does not depend on incentives or payoffs. This means the set of achievable joint distributions over actions and states is determined entirely by C and the prior μ, independent of the decision problem. As we show in Appendix D.1, this constraint defines a feasible set of information structures $\mathcal{Q}^*$, making fixed-capacity rational inattention a special case of capacity-constrained learning.

Example 2: Efficient Coding. In the efficient coding models of Woodford (2012), Khaw, Li, and Woodford (2021), and Frydman and Jin (2022), the decision-maker also faces a fixed bound on information processing, but the constraint is formulated differently. Rather than bounding mutual information evaluated at the true prior μ, efficient coding bounds the *channel capacity*, the maximum mutual information achievable over all possible input distributions. Specifically, with state attributes ω_n indexed by n, the constraint is:

$$\sum_n \max_{\pi_n} I_{\pi_n}(A_n;\Omega_n) \leq C,$$

where the maximization over input distributions π_n means the constraint limits the *potential* rate of information transmission, not just the actual rate under the true prior. This makes it costly to maintain fine discrimination among rare states, a key distinction from the rational inattention model. As in that model, however, the capacity C is exogenously fixed and does not vary with incentives. As we show in Appendix D.2, this again defines a feasible set of information structures $\mathcal{Q}^*$, making efficient coding a special case of capacity-constrained learning.

B.1 Representation Theorem

We say that a collection of SDSC data sets $P = (P_i)_{i\in D}$ has a capacity-constrained representation if there exists a set of decision-problem-dependent strategies (Q_i, q_i) for each $i \in D$ solving expected utility maximization (6) and optimal learning (8), given action set A_i, prize specification x_i, and a common feasible set of learning $\mathcal{Q}^*$.

Theorem 1. *Data set P has a capacity-constrained learning representation if and only if P satisfies NIS.*

The proof is relegated to Appendix C. Intuitively, the proof consists of two steps. The first step is to show that we can restrict the search for a capacity-constrained representation (CCR) to *revealed* experiments Q_i^P, defined in (14), for each decision problem i and corresponding SDSC P_i. More specifically, then, the first step is to conclude that the existence of any CCR implies existence with feasible learning set $\mathcal{Q}^* \equiv \cup_i Q_i^P$ and the condition:

$$G(Q_i^P|A_i, x_i) \geq G(Q_j^P|A_i, x_i), \tag{9}$$

for all pairs of decision problems $i, j \in D$. The second step is to establish equivalence of the preceding condition (9) with the NIS condition.

Theorem 1 provides a complete characterization of capacity-constrained learning in terms of observable choice data. While necessity is immediate—optimizing over a fixed feasible set precludes improving switches—sufficiency is the key contribution. It ensures that satisfying NIS is not only consistent with capacity-constrained learning, but that a rationalizing model can always be constructed. This completeness is essential for empirical application. When NIS fails, we can confidently reject the entire class of capacity-constrained models, not just specific parametric variants. Conversely, when NIS holds, we know the data admits a capacity-constrained interpretation, providing a meaningful positive result rather than merely a failure to reject.

C Proof of Theorem 1

Proof. Two established properties of the revealed experiment and action strategy defined in (13), (14), and (15) are key:

1. Caplin and Martin (2015) show that revealed experiment Q_i^P together with the implied mixed action strategies q_i^P generate the data:

$$P_{(Q_i^P, q_i^P)} = P_i$$

2. Caplin and Dean (2015) show that Q_i^P is uniquely the least Blackwell informative experiment that generates the data.

The first point to note is that if a CCR exists, then there exists a CCR in which only the revealed strategies are feasible, $\mathcal{Q}^* \equiv \cup_i Q_i^P$, and in which the revealed strategies are optimal for the corresponding decision problem. To see this suppose that a CCR exists and trim the feasible set of experiments to a set of size no more than the number of decision problems by associating with each decision problem i an optimal experiment Q_i with the defining characteristic of a CCR,

$$G(Q_i|A_i, x_i) \geq G(Q_j|A_i, x_i) \tag{10}$$

for all $i, j \in D$.

By Caplin, Martin, and Marx (2025a), any form of learning that optimally generates the data is an *optimality preserving spread* of the revealed experiment. Hence replacing Q_i, the experiment that is chosen in decision problem i (given choice set A_i and prize specification x_i) in the given CCR, with Q_i^P leaves expected utility unchanged:

$$G(Q_i|A_i, x_i) \quad = \quad G(Q_i^P|A_i, x_i).$$

Next, Caplin and Dean (2015) show that the revealed information structure is uniquely the least Blackwell informative experiment that generates the data. As a result, for any decision problem, the value of a revealed information structure Q_i^P is no more than its corresponding original information structure Q_i. That is for all $j \neq i$:

$$G(Q_j|A_i, x_i) \geq G(Q_j^P|A_i, x_i).$$

We conclude that (9) holds, i.e.,

$$G(Q_i^P|A_i, x_i) \geq G(Q_j^P|A_i, x_i),$$

for all $i, j \in D$, and hence that existence of a CCR implies existence with feasible learning set $\mathcal{Q}^* \equiv \cup_i Q_i^P$.

To complete the proof requires us now to show that equation (9) is equivalent to u satisfying NIS for P. To begin, we can replace the RHS in the NIS condition with that in (9) because:

$$\begin{aligned}
\sum_{a \in A_j} \max_{\hat{a} \in A_i} \sum_{\omega \in \Omega} P_j(a, \omega) u(x_i(\hat{a}, \omega)) &\overset{(1)}{=} \sum_{a \in A_j} \max_{\hat{a} \in A_i} \sum_{\omega \in \Omega} P_j(a) \gamma_j^a(\omega) u(x_i(\hat{a}, \omega)) \\
&\overset{(2)}{=} \sum_{a \in A_j} P_j(a) \max_{\hat{a} \in A_i} \sum_{\omega \in \Omega} \gamma_j^a(\omega) u(x_i(\hat{a}, \omega)) \\
&\overset{(3)}{=} \sum_{\gamma \in \Gamma(Q_j^P)} Q_j^P(\gamma) \max_{\hat{a} \in A_i} \sum_{\omega \in \Omega} \gamma(\omega) u(x_i(\hat{a}, \omega)) \\
&\overset{(4)}{=} \sum_{\gamma \in \Gamma(Q_j^P)} Q_j^P(\gamma) \max_{\hat{a} \in A_i} \sum_{\omega \in \Omega} U(\hat{a}|\gamma, x_i) \\
&\overset{(5)}{=} G(Q_j^P|A_i, x_i)
\end{aligned}$$

where the first equality follows by definition (13) of revealed posteriors γ_j^a, the second by rearrangement, the third by collecting actions $\{a \in A_j : \gamma_j^a = \gamma\}$ and the definition (14) of the revealed experiment Q_j^P, and the fourth by definition of posterior expected utility (5). Finally the fifth equality follows because (by the fourth equality), the RHS in the NIS condition is a restriction of (7) to pure strategies $q(a|\gamma) \in \{0, 1\}$, which establishes an upper bound; and yet, an optimal solution to linear program (7) in pure strategies always exists, which establishes equality. For the LHS, we note further that:

$$\sum_{a\in A_i} \max_{\hat{a}\in A_i} \sum_{\omega\in\Omega} P_i(a,\omega)u(x_i(\hat{a},\omega)) = \sum_{a\in A_i} \sum_{\omega\in\Omega} P_i(a,\omega)u(x_i(a,\omega))$$

by (2), i.e., NIAS or NIS restricted to case $i = j$. Substitution on both sides of equation (9) then establishes the validity of the NIS condition and with it the proof. □

D Examples of Capacity-Constrained Models

D.1 Example: Fixed-Capacity Rational Inattention

In Sims (2003), the DM solves:

$$\max_P \sum_{a\in A} \sum_{\omega\in\Omega} \mu(\omega)P(a|\omega)u(x(a,\omega))$$

subject to:

$$\forall a \in A, \omega \in \Omega : P(a|\omega) \in [0,1], \quad \forall\, \omega \in \Omega : \sum_{a\in A} P(a|\omega) = 1,$$
$$\sum_{a\in A} \sum_{\omega\in\Omega} \mu(\omega)P(a|\omega) \log\left(\frac{P(a|\omega)}{\sum_{\omega\in\Omega} \mu(\omega)P(a|\omega)}\right) \leq C$$

where C is the Shannon capacity.[2]

Letting $\mathcal{P}_\mu$ denote the set of *joint* distributions P over $A \times \Omega$ with marginal distribution $\sum_{a\in A} P(a,\omega) = \mu(\omega)$ for each $\omega \in \Omega$, we can equivalently write the problem as:

$$\max_{P\in\mathcal{P}_\mu} \sum_{a\in A} \sum_{\omega\in\Omega} P(a,\omega)u(x(a,\omega)) \tag{11}$$

subject to a Shannon capacity constraint on mutual information:

$$\sum_{a\in A} \sum_{\omega\in\Omega} P(a,\omega) \log\left(\frac{P(a,\omega)}{P(a)\mu(\omega)}\right) \leq C \tag{12}$$

where $P(a) \equiv \sum_{\omega\in\Omega} P(a,\omega)$ and $\mu(\omega)$ denote marginal probabilities over action a and state ω, respectively. It is now without loss of generality to restrict to the subset $\mathcal{P}^*_\mu$ of joint

[2]Shannon capacity refers to the maximum amount of information (in bits) that can be processed or transmitted through a channel without error, as formalized in information theory by Claude Shannon. In Sims (2003), this constraint limits the mutual information between states and actions, reflecting a bound on the DM's attention or cognitive processing.

distributions $P \in \mathcal{P}_\mu$ that additionally satisfy NIAS (1), [3] since wholesale action switches can only relax the constraint (12), and therefore distributions P violating NIAS cannot be optimal in (11) subject to (12). For each $P \in \mathcal{P}_\mu^*$ satisfying NIAS, Caplin and Martin (2015) guarantee the existence of a revealed experiment Q^P and action strategy $q^P : \Gamma(Q^*) \to \Delta(A)$ satisfying expected utility maximization (6). Namely, for any action $a \in A$ that is chosen, $P(a) > 0$, define revealed posterior γ^a by:

$$\gamma^a(\omega) = \frac{P(a,\omega)}{P(a)}; \tag{13}$$

define the revealed experiment Q^P by:

$$Q^P(\gamma) = \sum_{\{a \in A | \gamma^a = \gamma\}} P(a); \tag{14}$$

and define the action strategy by:

$$q^P(a|\gamma) = \begin{cases} \frac{P(a)}{Q^P(\gamma)} & \text{if } \gamma^a = \gamma; \\ 0 & \text{if } \gamma^a \neq \gamma. \end{cases} \tag{15}$$

Taking the feasible set $\mathcal{Q}^*$ to be the set of revealed experiments associated with the set of distributions P_μ^* yields the fixed-capacity model as an instance of our capacity-constrained model.

D.2 Example: Efficient Coding

In the fixed-capacity models of Woodford (2012), Khaw, Li, and Woodford (2021), and Frydman and Jin (2022), the state ω is a vector of length N in a bounded subset $\Omega \subset \mathbb{R}^N$, with attributes ω_n indexed by n. The action $a \in A$ is also an element of a bounded subset in $\mathbb{R}^N$ with attributes a_n. For consistency with our approach and notation, we restrict to the case where the state and action spaces are furthermore finite. The DM solves: [4]

$$\max_P \sum_{a \in A} \sum_{\omega \in \Omega} \mu(\omega) P(a|\omega) u(a,\omega) \tag{16}$$

[3]Note that P here is a choice variable, rather than observed data as in Caplin and Martin (2015), but the NIAS condition can be defined identically.

[4]In Woodford (2012) and Khaw, Li, and Woodford (2021), the DM's objective is to minimize the expected MSE, that is $u(a,\omega) = -(a-\omega)^2$. In Frydman and Jin (2022), the DM's objective is to maximize the expected financial gain of a lottery choice question, so $u(a,\omega) = p \cdot \omega_X \cdot \mathbb{1}\{c(a) = \text{Risk}\} + \omega_C \cdot \mathbb{1}\{c(a) = \text{Certain}\}$, with choice function

$$c(\cdot)\colon \mathcal{A} \mapsto \{\text{Risk}, \text{Certain}\} = \arg\max_{c(\cdot)} \Big(\iint p \cdot \omega_X \cdot \mathbb{1}\{c(a) = \text{Risk}\} P(a|\omega) \cdot \mu(\omega), da, d\omega \\ + \iint \omega_C \cdot \mathbb{1}\{c(a) = \text{Certain}\} \cdot P(a|\omega) \cdot \mu(\omega), da, d\omega \Big)$$

where (as in the previous example) P is a state-conditional action probability subject to the following additional constraints. Since all attributes are assumed independent, $P(a|\omega) = \prod_{n=1}^{N} P(a_n|\omega_n)$, with $P(a_n|\omega_n)$ being a distribution over a_n conditioning on ω_n. Let $\pi_n(\cdot)$ denote a marginal distribution over state attributes ω_n. Then the efficient coding constraint is:

$$\sum_n \left[\max_{\pi_n} \left\{ \sum_{a_n} \sum_{\omega_n} \pi_n(\omega_n) P(a_n|\omega_n) \log \left[\frac{P(a_n|\omega_n)}{\sum_{\omega_n} \pi_n(\omega_n) P(a_n|\omega_n)} \right] \right\} \right] \leq C. \tag{17}$$

As in the preceding example, we can restrict to conditional action probabilities P for which the induced joint distribution defined by $P_\mu(a,\omega) \equiv \mu(\omega)P(a|\omega)$ satisfies NIAS (1), since wholesale action switches can only relax the efficient coding constraint (17). As before, we can associate each such (conditional) probability P through its induced joint distribution P_μ to a revealed experiment and action strategy (Q^P, q^P) satisfying expected utility maximization (6). Taking these associated revealed experiments Q^P as the feasible set of information structures yields the efficient coding model as an instance of our capacity-constrained model.

E Disaggregated Experiment Results

E.1 CCLN, by Difficulty Level

Incentive level					
Lower	Higher	NIS LHS	NIS RHS	NIS inequality fail?	p-value
1	2	0.58	0.61	Yes	0.02
1	4	0.58	0.61	Yes	0.03
1	8	0.58	0.63	Yes	< 0.01
1	16	0.58	0.63	Yes	< 0.01
1	32	0.58	0.59	Yes	0.31
2	4	1.22	1.22	Yes	0.48
2	8	1.22	1.25	Yes	0.20
2	16	1.22	1.26	Yes	0.19
2	32	1.22	1.18	No	0.78
4	8	2.45	2.50	Yes	0.22
4	16	2.45	2.52	Yes	0.18
4	32	2.45	2.37	No	0.77
8	16	5.01	5.03	Yes	0.45
8	32	5.01	4.73	No	0.91
16	32	10.06	9.46	No	0.92

Table 1: NIS inequalities in direction of increasing incentives (aggregate data from difficulty level 1 in CCLN)

DP1	DP2	LHS	RHS	NIS inequality fail?	p-value
1	2	0.60	0.64	Yes	< 0.01
1	4	0.60	0.65	Yes	< 0.01
1	8	0.60	0.64	Yes	< 0.01
1	16	0.60	0.68	Yes	< 0.01
1	32	0.60	0.66	Yes	< 0.01
2	4	1.29	1.31	Yes	0.28
2	8	1.29	1.29	Yes	0.49
2	16	1.29	1.35	Yes	0.05
2	32	1.29	1.32	Yes	0.22
4	8	2.61	2.57	No	0.71
4	16	2.61	2.71	Yes	0.14
4	32	2.61	2.65	Yes	0.36
8	16	5.14	5.41	Yes	0.04
8	32	5.14	5.29	Yes	0.24
16	32	10.82	10.58	No	0.71

Table 2: NIS inequalities in direction of increasing incentives (aggregate data from difficulty level 2 in CCLN)

DP1	DP2	LHS	RHS	NIS inequality fail?	p-value
1	2	0.62	0.65	Yes	0.01
1	4	0.62	0.64	Yes	0.09
1	8	0.62	0.65	Yes	0.08
1	16	0.62	0.68	Yes	< 0.01
1	32	0.62	0.69	Yes	< 0.01
2	4	1.30	1.28	No	0.76
2	8	1.30	1.29	No	0.63
2	16	1.30	1.37	Yes	0.07
2	32	1.30	1.37	Yes	0.10
4	8	2.56	2.58	Yes	0.42
4	16	2.56	2.73	Yes	0.02
4	32	2.56	2.74	Yes	0.04
8	16	5.16	5.46	Yes	0.06
8	32	5.16	5.48	Yes	0.05
16	32	10.93	10.96	Yes	0.47

Table 3: NIS inequalities in direction of increasing incentives (aggregate data from difficulty level 3 in CCLN)

DP1	DP2	LHS	RHS	NIS inequality fail?	p-value
1	2	0.72	0.73	Yes	0.22
1	4	0.72	0.74	Yes	0.04
1	8	0.72	0.73	Yes	0.24
1	16	0.72	0.79	Yes	< 0.01
1	32	0.72	0.80	Yes	< 0.01
2	4	1.46	1.49	Yes	0.15
2	8	1.46	1.46	Yes	0.47
2	16	1.46	1.57	Yes	< 0.01
2	32	1.46	1.59	Yes	< 0.01
4	8	2.98	2.92	No	0.78
4	16	2.98	3.14	Yes	0.02
4	32	2.98	3.19	Yes	0.01
8	16	5.84	6.28	Yes	< 0.01
8	32	5.84	6.37	Yes	< 0.01
16	32	12.57	12.74	Yes	0.32

Table 4: NIS inequalities in direction of increasing incentives (aggregate data from difficulty level 6 in CCLN)

E.2 DN20 Dots Task, by Incentive Level

DP 1 (incentive quartile)	DP 2 (incentive quartile)	LHS	RHS	NIS inequality fail?	p-value
1st	2nd	6.03	6.94	Yes	< 0.01
1st	3rd	6.03	7.82	Yes	< 0.01
1st	4th	6.03	9.01	Yes	< 0.01
2nd	3rd	20.28	22.86	Yes	< 0.01
2nd	4th	20.28	26.34	Yes	< 0.01
3rd	4th	37.89	43.67	Yes	< 0.01

Table 5: Aggregate NIS test results of the Dot task of DN20 (only $10 sessions)

DP 1 (incentive quartile)	DP 2 (incentive quartile)	LHS	RHS	NIS inequality fail?	p-value
1st	2nd	6.33	7.16	Yes	< 0.01
1st	3rd	6.33	8.74	Yes	< 0.01
1st	4th	6.33	9.61	Yes	< 0.01
2nd	3rd	20.92	25.54	Yes	< 0.01
2nd	4th	20.92	28.10	Yes	< 0.01
3rd	4th	42.34	46.59	Yes	< 0.01

Table 6: Aggregate NIS test results of the Dot task of DN20 (only $20 sessions)

E.3 DN20 Angles Task, by Incentive Level

DP 1 (incentive quartile)	DP 2 (incentive quartile)	LHS	RHS	NIS inequality fail?	p-value
1st	2nd	5.64	5.60	No	0.55
1st	3rd	5.64	5.91	Yes	0.18
1st	4th	5.64	5.71	Yes	0.41
2nd	3rd	16.38	17.27	Yes	0.13
2nd	4th	16.38	16.68	Yes	0.35
3rd	4th	28.63	27.66	No	0.75

Table 7: Aggregate NIS test results of the Angle task of DN20 ($10 sessions only)

DP 1 (incentive quartile)	DP 2 (incentive quartile)	LHS	RHS	NIS inequality fail?	p-value
1st	2nd	6.02	5.80	No	0.80
1st	3rd	6.02	5.94	No	0.61
1st	4th	6.02	5.98	No	0.57
2nd	3rd	16.94	17.38	Yes	0.27
2nd	4th	16.94	17.48	Yes	0.25
3rd	4th	28.81	28.98	Yes	0.45

Table 8: Aggregate NIS test results of the Angle task of DN20 ($20 sessions only)

References

Caplin, Andrew and Mark Dean (2015). “Revealed preference, rational inattention, and costly information acquisition”. In: *American Economic Review* 105.7, pp. 2183–2203.

Caplin, Andrew and Daniel Martin (2015). “A testable theory of imperfect perception”. In: *The Economic Journal* 125.582, pp. 184–202.

Caplin, Andrew, Daniel Martin, and Philip Marx (2025a). “Modeling Machine Learning: A Cognitive Economic Approach”. In: *Journal of Economic Theory* 224, p. 105970. DOI: 10.1016/j.jet.2025.105970.

— (2025b). “Rationalizable Learning”. In: *Economic Theory* 80.1, pp. 171–202. DOI: 10.1007/s00199-024-01627-z.

Frydman, Cary and Lawrence J. Jin (2022). “Efficient coding and risky choice”. In: *The Quarterly Journal of Economics* 137.1, pp. 161–213.

Kamenica, Emir and Matthew Gentzkow (2011). “Bayesian persuasion”. In: *American Economic Review* 101.6, pp. 2590–2615.

Khaw, Mel Win, Ziang Li, and Michael Woodford (2021). “Cognitive imprecision and small-stakes risk aversion”. In: *The review of economic studies* 88.4, pp. 1979–2013.

Sims, Christopher A (2003). “Implications of Rational Inattention”. In: *Journal of Monetary Economics* 50.3, pp. 665–690.

Woodford, Michael (2012). “Prospect theory as efficient perceptual distortion”. In: *American Economic Review* 102.3, pp. 41–46.