Online material for: Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model

By Filip Matějka and Alisdair McKay[†]

C. Additional Proofs for Section II

Proof of Corollary 2. For $\mathcal{P}_i^0 > 0$, the first order condition on (14) with respect to $\mathcal{P}_i^0$, is

$$\lambda \int_{\mathbf{v}} \frac{e^{v_i/\lambda} - e^{v_N/\lambda}}{\sum_{j=1}^N \mathcal{P}_j^0 e^{v_j/\lambda}} G(d\mathbf{v}) = 0, \tag{C1}$$

where $\mathcal{P}_N^0$ denotes $1 - \sum_{i=1}^{N-1} \mathcal{P}_i^0$.

For $i \in \{1, \cdots, N-1\}$, we can write

$$\int_{\mathbf{v}} \frac{e^{v_i/\lambda}}{\sum_{i=j}^N \mathcal{P}_j^0 e^{v_j/\lambda}} G(d\mathbf{v}) = \int_{\mathbf{v}} \frac{e^{v_N/\lambda}}{\sum_{j=1}^N \mathcal{P}_j^0 e^{v_j/\lambda}} G(d\mathbf{v}) \equiv \mu.$$

Notice that $\mu = 1$ because

$$\begin{aligned}\sum_{i=1}^N \mathcal{P}_i^0 \mu &= \sum_{i=1}^N \mathcal{P}_i^0 \int_{\mathbf{v}} \frac{e^{v_i/\lambda}}{\sum_{j=1}^N \mathcal{P}_j^0 e^{v_j/\lambda}} G(d\mathbf{v}) \\ &= \int_{\mathbf{v}} \frac{\sum_{i=1}^N \mathcal{P}_i^0 e^{v_i/\lambda}}{\sum_{j=1}^N \mathcal{P}_j^0 e^{v_j/\lambda}} G(d\mathbf{v}) \\ &= \int_{\mathbf{v}} G(d\mathbf{v}) = 1\end{aligned}$$

so $\mu \sum_{i=1}^N \mathcal{P}_i^0 = 1$, but as $\left\{\mathcal{P}_i^0\right\}_{i=1}^N$ are probabilities we know $\sum_{i=1}^N \mathcal{P}_i^0 = 1$. Therefore, $\mu = 1$. Equation (C1) then becomes (17).

LEMMA 1: *The optimization problem in Corollary 1 always has a solution.*

PROOF:

Since (13) is a necessary condition for the maximum, then the collection $\{\mathcal{P}_i^0\}_{i=1}^N$ determines the whole solution. However, the objective is a continuous function of $\{\mathcal{P}_i^0\}_{i=1}^N$, since $\{\mathcal{P}_i(\mathbf{v})\}_{i=1}^N$ is also

[†] Matějka: CERGE-EI, A joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Academy of Sciences of the Czech Republic. Politických vězňů 7, Praha 1, 150 00, the Czech Republic, filip.matejka@cerge-ei.cz. McKay: Department of Economics, Boston University, 270 Bay State Road, Boston, Massachusetts, USA, amckay@bu.edu. This research was funded by GAČR P402/11/P236, by European Research Council under the European Community's Seventh Framework Programme FP7/2007-2013 grant agreement N. 263790, and by a grant RRC 11+001 from the CERGE-EI Foundation under a program of the Global Development Network. All opinions expressed are those of the authors and have not been endorsed by CERGE-EI or the GDN.

a continuous function of $\{\mathcal{P}_i^0\}_{i=1}^N$. Moreover, the admissible set for $\{\mathcal{P}_i^0\}_{i=1}^N$ is compact. Therefore, the maximum always exists.

Given a solution to Corollary 1 we can always construct a solution to Definition 1 as discussed in Footnote 13.

Uniqueness

Concerning uniqueness, there can be special cases where the decision maker is indifferent between processing more information in order to generate a higher $E[v_i]$ and processing less information and conserving on information costs. However, a rigid co-movement of payoffs is required for these cases to arise. Without this structure, if the decision maker were indifferent between two different strategies then their convex combination would be preferred as the entropy cost is convex in strategies, $\{\mathcal{P}_i(\mathbf{v})\}_{i=1}^N$, while $E[v_i]$ is linear.

Under the conditions stated below, there is a unique solution to Corollary 1 and therefore a unique $\mathcal{P}$ that is induced by the solution to Definition 1. However, there are many signaling strategies that lead to the same state-contingent choice probabilities.

LEMMA 2: ***Uniqueness:*** *If the random vectors $e^{v_j/\lambda}$ are linearly independent with unit scaling, i.e. if there does not exist a set $\{a_j\}_{j=2}^{N-1}$ such that $\sum_{j=2}^N a_j = 1$ and*

(C2)
$$e^{v_i/\lambda} = \sum_{j=2}^{N} a_j e^{v_j/\lambda} \qquad \textit{almost surely,}$$

then the agent's problem has a unique solution. Conversely, if the agent's problem has a unique solution, then the random vectors $e^{v_j/\lambda}$ for all j s.t. $\mathcal{P}_j^0 > 0$ are linearly independent with unit scaling.

PROOF:

Let us first study an interior optimum of (14), where the boundary constraint (15) is not binding. The first order conditions take the form of (17) for $i < N$ and denote $\mathcal{P}_N^0 = 1 - \sum_{k=1}^{N-1} \mathcal{P}_k^0$ to satisfy the constraint (16). The $(N-1)$ dimensional Hessian is

(C3)
$$H_{ij} = -\int_{\mathbf{v}} \Big(\frac{e^{v_i/\lambda} - e^{v_N/\lambda}}{\sum_{k=1}^N \mathcal{P}_k^0 e^{v_k/\lambda}} \Big) \Big(\frac{e^{v_j/\lambda} - e^{v_N/\lambda}}{\sum_{k=1}^N \mathcal{P}_k^0 e^{v_k/\lambda}} \Big) G(d\mathbf{v}) \qquad \forall i, j < N.$$

The hessian H is thus (-1) times a Gramian matrix, which is a matrix generated from inner products of random vectors $\frac{e^{v_i/\lambda} - e^{v_N/\lambda}}{\sum_{k=1}^N \mathcal{P}_k^0 e^{v_k/\lambda}}$. H is thus negative semi-definite at all interior points $\{\mathcal{P}_i^0\}_{i=1}^N$, not just at the optimum only. This implies that an interior optimum of the objective (14) is a unique maximum if and only if the Hessian is negative definite at the optimum. From (C3) we see that for $N = 2$ the Hessian is not negative-definite, i.e. the objective is not a strictly concave function of $\mathcal{P}_1^0$, only when $e^{v_1} = e^{v_2}$ almost surely, i.e. when the actions are identical. Moreover for $N > 2$, we can use the fact that Gramian matrices have a zero eigenvalue if and only if the generating vectors are linearly dependent, which means that there exist i and a set $\{a_j\}_{j=1,\neq i}^{N-1}$ such that

(C4)
$$\frac{e^{v_i/\lambda} - e^{v_N/\lambda}}{\sum_{k=1}^N \mathcal{P}_k^0 e^{v_k/\lambda}} = \sum_{j=1,\neq i}^{N-1} a_j \frac{e^{v_j/\lambda} - e^{v_N/\lambda}}{\sum_{k=1}^N \mathcal{P}_k^0 e^{v_k/\lambda}} \qquad \text{almost surely.}$$

Since the equality needs to hold almost surely, we can get rid of the denominators, which are the same on both sides. By denoting $a_N = 1 - \sum_{j=1,\neq i}^{N-1} a_j$, this implies the following sufficient and necessary condition for non-uniqueness in the interior for $N \geq 1$: there exists a set $\{a_j\}_{j=1,\neq i}^{N}$ such that $\sum_{j=2}^{N} a_j = 1$, and

$$e^{v_i/\lambda} = \sum_{j=1,\neq i}^{N} a_j e^{v_j/\lambda} \qquad \text{almost surely.} \tag{C5}$$

Now, we extend the proof to non-interior solutions. In the optimization problem formulated in Corollary 1, the objective is a weakly concave function of $\{\mathcal{P}_i(\mathbf{v})\}_i$ on a convex set. Therefore, any convex linear combination of two solutions must be a solution, too. This implies that there always exists a solution with $\mathcal{P}_i^0 > 0$ for all $i \in S$, where S is a set of indices i for which there exists a solution with $\mathcal{P}_i^0 > 0$. For example, if there exist two distinct solutions such that $\mathcal{P}_1^0 > 0$ for one and $\mathcal{P}_2^0 > 0$ for the other, then there exists a solution with both $\mathcal{P}_1^0$ and $\mathcal{P}_2^0$ positive. Therefore, there always exists an interior solution on the subset S of actions that are ever selected in some solution. Moreover, we can create many such solutions by taking different convex combinations. However, if the conditions of the proposition are satisfied, then there can only be one interior solution and hence the multiple boundary solutions leads to a contradiction. For the actions that are not selected in any solution, the solution is unique with $\mathcal{P}_i^0 = 0$.

COROLLARY 1: *If any one of the following conditions is satisfied, then the solution is unique.*
(1) $N = 2$ and the payoffs of the two actions are not equal almost surely.
(2) The prior is symmetric and payoffs of the actions are not equal almost surely.

D. Proofs for Section III

Proof of Lemma 3. Let there exist an $\mathbf{x}$ such that $\mathcal{P}_i(\mathbf{x}) > 0$ w.l.o.g $i = 1$. We prove the statement by showing that action 1 is selected with positive probability for all vectors $\mathbf{y}$.

Let us first generate a vector $\mathbf{x^{11}}$ by copying x_1 to action 2: $\mathbf{x^{11}} = (x_1, x_1, x_3, \ldots, x_N)$. We find that $\mathcal{P}_1(\mathbf{x^{11}}) > 0$, which is due to Axiom 1 when it is applied to $\mathbf{x}$ and $\mathbf{x^{11}}$ with $i = 3$ and $j = 1$: the axiom implies that $\frac{\mathcal{P}_3(\mathbf{x})}{\mathcal{P}_1(\mathbf{x})} = \frac{\mathcal{P}_3(\mathbf{x^{11}})}{\mathcal{P}_1(\mathbf{x^{11}})}$. Now, Axiom 2 implies that as long as there is the same consequence for actions 1 and 2, then the probability of action 1 is positive independently of what the consequence is and what the consequences of the other actions are.

Finally, we show that $\mathcal{P}_1(\mathbf{y}) > 0$, where $\mathbf{y}$ is an arbitrary vector of consequences. This is due to the fact that $\mathcal{P}_1(\mathbf{y^{11}}) > 0$, which we just showed in the paragraph above, where $\mathbf{y^{11}} = (y_1, y_1, y_3, \ldots, y_N)$, and due to Axiom 1 when it is applied to $\mathbf{y^{11}}$, $\mathbf{y}$, $i = 3$ and $j = 1$. The Axiom implies that if $\mathcal{P}_1(\mathbf{y^{11}}) > 0$, then $\mathcal{P}_1(\mathbf{y}) > 0$ too, since $\mathbf{y}$ and $\mathbf{y}^{11}$ only differ in the consequences action 2.

To establish that if $\mathcal{P}_i(\mathbf{x}) = 0$ then $\mathcal{P}_i(\mathbf{y}) = 0$ for all $\mathbf{y}$ is straightforward: suppose $\mathcal{P}_i(\mathbf{y}) > 0$, then the argument above implies $\mathcal{P}_i(\mathbf{x}) > 0$.

Proof of Proposition 2. We refer to actions that are not positive actions, i.e. are never selected, as "zero actions." Assume w.l.o.g. that actions 1, 2, and N are positive. Fix a vector of consequences $\mathbf{x}$ and define

$$v(a) \equiv \log\left(\frac{\mathcal{P}_1(a, x_2, x_3, \cdots, x_N)}{\mathcal{P}_N(a, x_2, x_3, \cdots, x_N)}\right). \tag{D1}$$

So the payoff of consequence a is defined in terms of the probability that it is selected when it is inserted into the first position of a particular vector of consequences. Also define

$$\xi_k \equiv \frac{\mathcal{P}_k(\mathbf{x^k})}{\mathcal{P}_1(\mathbf{x^k})},$$

where $\mathbf{x^k}$ is defined as $(x_k, x_2, x_3, \cdots, x_N)$, which is $\mathbf{x}$ with the first element replaced by a second instance of the consequence in the k^{th} position. Notice that if k is a zero action, then $\xi_k = 0$. By Axiom 2, we have

$$\xi_k = \frac{\mathcal{P}_k(\mathbf{y^k})}{\mathcal{P}_1(\mathbf{y^k})},$$

where $\mathbf{y^k}$ is generated from an arbitrary vector of consequences $\mathbf{y}$ in the same manner that $\mathbf{x^k}$ was generated from $\mathbf{x}$.

Consider a vector of consequences $\mathbf{y}$ that shares the N^{th} entry with the generating vector $\mathbf{x}$, such that $y_N = x_N$. We will show

$$\text{(D2)} \qquad \mathcal{P}_i(\mathbf{y}) = \frac{\xi_i e^{v(y_i)}}{\sum_{j=1}^{N} \xi_j e^{v(y_j)}},$$

for all i. If i is a zero action, then (D2) holds trivially so we will suppose that i is a positive action. As the choice probabilities must sum to one, we proceed as follows

$$1 = \sum_j \mathcal{P}_j(\mathbf{y}) = \mathcal{P}_i(\mathbf{y}) \sum_j \frac{\mathcal{P}_j(\mathbf{y})}{\mathcal{P}_i(\mathbf{y})} = \mathcal{P}_i(\mathbf{y}) \sum_j \frac{\mathcal{P}_j(\mathbf{y})/\mathcal{P}_N(\mathbf{y})}{\mathcal{P}_i(\mathbf{y})/\mathcal{P}_N(\mathbf{y})}$$

$$\text{(D3)} \qquad \mathcal{P}_i(\mathbf{y}) = \frac{\mathcal{P}_i(\mathbf{y})/\mathcal{P}_N(\mathbf{y})}{\sum_j \mathcal{P}_j(\mathbf{y})/\mathcal{P}_N(\mathbf{y})}.$$

Now, by Axiom 1:

$$\frac{\mathcal{P}_k(\mathbf{y})}{\mathcal{P}_N(\mathbf{y})} = \frac{\mathcal{P}_k(\mathbf{y^k})}{\mathcal{P}_N(\mathbf{y^k})}$$

so (D3) becomes

$$\begin{aligned} \mathcal{P}_i(\mathbf{y}) &= \frac{\mathcal{P}_i(\mathbf{y^i})/\mathcal{P}_N(\mathbf{y^i})}{\sum_j \mathcal{P}_j(\mathbf{y^j})/\mathcal{P}_N(\mathbf{y^j})} \\ \text{(D4)} \qquad &= \frac{\xi_i \mathcal{P}_1(\mathbf{y^i})/\mathcal{P}_N(\mathbf{y^i})}{\sum_j \xi_j \mathcal{P}_1(\mathbf{y^j})/\mathcal{P}_N(\mathbf{y^j})}. \end{aligned}$$

For any k, as $y_N = x_N$ in the case we are considering, by Axiom 1 and definition of $v(\cdot)$ in (D1):

$$\begin{aligned} \frac{\mathcal{P}_1(\mathbf{y^k})}{\mathcal{P}_N(\mathbf{y^k})} &= \frac{\mathcal{P}_1(y_k, x_2, x_3, \cdots, x_N)}{\mathcal{P}_N(y_k, x_2, x_3, \cdots, x_N)} \\ &= e^{v(y_k)}. \end{aligned}$$

Therefore (D4) becomes (D2).

We will now consider an arbitrary $\mathbf{y}$ allowing for $y_N \neq x_N$ as well. We will show that (D2) still holds by using the axioms and some intermediate vectors to connect $\mathbf{y}$ to $\mathbf{x}$. Let $\mathbf{y^{wizj}}$ be the vector generated from $\mathbf{y}$ by replacing the first element of $\mathbf{y}$ with w_i and the last element of $\mathbf{y}$ with z_j for given vectors $\mathbf{w}$ and $\mathbf{z}$. For example:

$$\begin{aligned}
\mathbf{y}^{\mathbf{y1xN}} &= (y_1, y_2, y_3, \cdots, y_{N-1}, x_N) \\
\mathbf{y}^{\mathbf{x1xN}} &= (x_1, y_2, y_3, \cdots, y_{N-1}, x_N) \\
\mathbf{y}^{\mathbf{xNxN}} &= (x_N, y_2, y_3, \cdots, y_{N-1}, x_N) \\
\mathbf{y}^{\mathbf{yNyN}} &= (y_N, y_2, y_3, \cdots, y_{N-1}, y_N).
\end{aligned}$$

Consider $\mathbf{y}^{\mathbf{yNyN}}$. For any $i < N$, by Axiom 1:

$$\begin{aligned}
\mathcal{P}_i(\mathbf{y}^{\mathbf{yNyN}}) &= \mathcal{P}_1(\mathbf{y}^{\mathbf{yNyN}}) \frac{\mathcal{P}_i(\mathbf{y}^{\mathbf{yNxN}})}{\mathcal{P}_1(\mathbf{y}^{\mathbf{yNxN}})} \\
&= \mathcal{P}_1(\mathbf{y}^{\mathbf{yNyN}}) \frac{\xi_i e^{v(y_i)}}{\xi_1 e^{v(y_N)}},
\end{aligned} \tag{D5}$$

where the second equality follows from the fact that (D2) holds for $\mathbf{y} = \mathbf{y}^{\mathbf{yNxN}}$ as its N^{th} entry is x_N and we have already established that (D2) holds for vectors $\mathbf{y}$ for which $y_n = x_N$. For $i = N$ we have, by Axiom 2:

$$\begin{aligned}
\mathcal{P}_N(\mathbf{y}^{\mathbf{yNyN}}) &= \mathcal{P}_1(\mathbf{y}^{\mathbf{yNyN}}) \frac{\mathcal{P}_N(\mathbf{y}^{\mathbf{xNxN}})}{\mathcal{P}_1(\mathbf{y}^{\mathbf{xNxN}})} \\
&= \mathcal{P}_1(\mathbf{y}^{\mathbf{yNyN}}) \frac{\xi_N e^{v(x_N)}}{\xi_1 e^{v(x_N)}} = \mathcal{P}_1(\mathbf{y}^{\mathbf{yNyN}}) \frac{\xi_N}{\xi_1}.
\end{aligned} \tag{D6}$$

Combining (D5) and (D6),

$$\frac{\mathcal{P}_i(\mathbf{y}^{\mathbf{yNyN}})}{\mathcal{P}_N(\mathbf{y}^{\mathbf{yNyN}})} = \frac{\xi_i e^{v(y_i)}}{\xi_N e^{v(y_N)}} \tag{D7}$$

for all i. As the probabilities sum to one, we arrive at

$$\mathcal{P}_N(\mathbf{y}^{\mathbf{yNyN}}) = \frac{\xi_N e^{v(y_N)}}{\sum_j \xi_j e^{v(y_j)}} \tag{D8}$$

and (D2) for $\mathbf{y} = \mathbf{y}^{\mathbf{yNyN}}$ follows from (D7) and (D8).

Finally, we turn our attention to the arbitrary $\mathbf{y}$. For any $j < N$, we use Axiom 1 to write

$$\frac{\mathcal{P}_j(\mathbf{y})}{\mathcal{P}_2(\mathbf{y})} = \frac{\mathcal{P}_j(\mathbf{y}^{\mathbf{y1xN}})}{\mathcal{P}_2(\mathbf{y}^{\mathbf{y1xN}})} = \frac{\xi_j e^{v(y_j)}}{\xi_2 e^{v(y_2)}}, \tag{D9}$$

where the second equality follows from the fact that (D2) has already been established for $\mathbf{y} = \mathbf{y}^{\mathbf{y1xN}}$. For $j = N$, by Axiom 1 we can write

$$\frac{\mathcal{P}_N(\mathbf{y})}{\mathcal{P}_2(\mathbf{y})} = \frac{\mathcal{P}_N(\mathbf{y}^{\mathbf{yNyN}})}{\mathcal{P}_2(\mathbf{y}^{\mathbf{yNyN}})} = \frac{\xi_N e^{v(y_N)}}{\xi_2 e^{v(y_2)}}, \tag{D10}$$

where the second equality follows from the fact that (D2) has already been established for $\mathbf{y} =$

$\mathbf{y}^{\mathbf{yNyN}}$. Using $\sum_j \mathcal{P}_j(\mathbf{y}) = 1$ we arrive at

$$\mathcal{P}_2(\mathbf{y}) = \frac{\xi_2 e^{v(y_2)}}{\sum_j \xi_j e^{v(y_j)}}$$

and then (D2) follows from (D9) and (D10). To complete the proof, we apply the normalization $\mathcal{P}_i^0 = \xi_i / \sum_j \xi_j$ for all i.

E. Proofs for Section IV

E1. Monotonicity

Proof of Proposition 3. The agent's objective function, (14), can be rewritten to include the constraint $\sum_{i=1}^{N} \mathcal{P}_i^0 = 1$

$$\int_{\mathbf{v}} \lambda \log \left[\sum_{i=1}^{N-1} \mathcal{P}_i^0 e^{v_i/\lambda} + \left(1 - \sum_{i=1}^{N-1} \mathcal{P}_i^0\right) e^{v_N/\lambda} \right] G(d\mathbf{v}).$$

Written in this way, the agent is maximizing over $\{\mathcal{P}_i^0\}_{i=1}^{N-1}$ subject to (15). Let us first assume that the constraint (15) is not binding and later on we show the statement holds in general.

The first order condition with respect to $\mathcal{P}_1^0$ is

$$\lambda \int_{\mathbf{v}} \frac{e^{v_1/\lambda} - e^{v_N/\lambda}}{\sum_{j=1}^{N} \mathcal{P}_j^0 e^{v_j/\lambda}} G(d\mathbf{v}) = 0, \tag{E1}$$

where $\mathcal{P}_N^0$ denotes $1 - \sum_{i=1}^{N-1} \mathcal{P}_i^0$.

$\hat{G}(\cdot)$ is generated from $G(\cdot)$ by increasing the payoffs of action 1 and this change can be implemented using a function $f(\mathbf{v}) \geq 0$, where $\int f(\mathbf{v}) G(d\mathbf{v}) > 0$, which describes the increase in v_1 in various states. Let v be transformed such that $e^{\hat{v}_1/\lambda} = e^{v_1/\lambda}(1 + f(\mathbf{v}))$ and with $\hat{v}_j = v_j$ for all $\mathbf{v}$ and $j = 2, \cdots, N$. Under the new prior, $\hat{G}(\cdot)$, the left-hand side of (E1) becomes

$$\Delta_1 \equiv \lambda \int_{\mathbf{v}} \frac{e^{v_1/\lambda}(1 + f(\mathbf{v})) - e^{v_N/\lambda}}{\sum_{j=1}^{N} \mathcal{P}_j^0 e^{v_j/\lambda} + \mathcal{P}_1^0 e^{v_1/\lambda} f(\mathbf{v})} G(d\mathbf{v}). \tag{E2}$$

Notice that $[\Delta_1, \cdots, \Delta_{N-1}]$ is the gradient of the agent's objective function under the new prior evaluated at the optimal strategy under the original prior. We now consider a marginal improvement in the direction of $f(\mathbf{v})$. In particular, consider an improvement of $\varepsilon f(\mathbf{v})$ for some $\varepsilon > 0$. Equation (E2) becomes

$$\Delta_1 = \lambda \int_{\mathbf{v}} \frac{e^{v_1/\lambda}(1 + \varepsilon f(\mathbf{v})) - e^{v_N/\lambda}}{\sum_{j=1}^{N} \mathcal{P}_j^0 e^{v_j/\lambda} + \mathcal{P}_1^0 e^{v_1/\lambda} \varepsilon f(\mathbf{v})} G(d\mathbf{v}). \tag{E3}$$

Differentiating with respect to ε at $\varepsilon = 0$ leads to

$$\text{(E4)} \qquad \frac{\partial \Delta_1}{\partial \varepsilon}\bigg|_{\varepsilon=0} = \lambda \int_{\mathbf{v}} \frac{e^{v_1/\lambda} f(\mathbf{v}) \sum_{j=1}^{N} \mathcal{P}_j^0 e^{v_j/\lambda} - \left(e^{v_1/\lambda} - e^{v_N/\lambda}\right) \mathcal{P}_1^0 e^{v_1/\lambda} f(\mathbf{v})}{\left(\sum_{j=1}^{N} \mathcal{P}_j^0 e^{v_j/\lambda}\right)^2} G(d\mathbf{v})$$

$$\text{(E5)} \qquad = \lambda \int_{\mathbf{v}} e^{v_1/\lambda} f(\mathbf{v}) \frac{\sum_{j=2}^{N} \mathcal{P}_j^0 e^{v_j/\lambda} + e^{v_N/\lambda} \mathcal{P}_1^0}{\left(\sum_{j=1}^{N} \mathcal{P}_j^0 e^{v_j/\lambda}\right)^2} G(d\mathbf{v}) > 0.$$

This establishes that at the original optimum, $\left\{\mathcal{P}_i^0\right\}_{i=1}^{N-1}$, the impact of a marginal improvement in action 1 is to increase the gradient of the new objective function with respect to the probability of the first action. Therefore the agent will increase $\mathcal{P}_1^0$ in response to the marginal improvement. Notice that this holds for a marginal change derived from any total $f(\mathbf{v})$. Therefore, if the addition of the total $f(\mathbf{v})$ were to decrease $\mathcal{P}_1^0$, then by regularity there would have to be a marginal change of the prior along $\nu f(\mathbf{v})$, where $\nu[0,1]$, such that $\mathcal{P}_1^0$ decreases due to this marginal change too. However, we showed that the marginal change never decreases $\mathcal{P}_1^0$.

We conclude the proof by addressing the cases when the (15) can be binding. We already know that monotonicity holds everywhere in the interior, therefore the only remaining case that could violate the monotonicity is if $\mathcal{P}_1^0 = 1$, while $\hat{\mathcal{P}}_1^0 = 0$. In other words, if after an increase of payoff in some states the action switches from being selected with probability one to never being selected. However, this is not possible since the expected payoff of action 1 increases. If $\mathcal{P}_1^0 = 1$, the agent processes no information and thus the expected utility equals the expectation of v_1. After the transformation of the payoffs of action 1, under $\hat{G}$, each strategy that ignores action 1 delivers the same utility as before the transformation, but the expected utility from selecting the action 1 with probability one is higher than before. Therefore, $\hat{\mathcal{P}}_1^0 = 1$. Strict monotonicity holds in the interior and weak monotonicity on the boundaries.

E2. Duplicates

Proof of Proposition 4. Let us consider a problem with $N+1$ actions, where the actions N and $N+1$ are duplicates. Let $\{\hat{\mathcal{P}}_i^0(\mathbf{u})\}_{i=1}^{N+1}$ be the unconditional probabilities in the solution to this problem. Since u_N and u_{N+1} are almost surely equal, then we can substitute u_N for u_{N+1} in the first order condition (13) to arrive at:

$$\text{(E6)} \qquad \hat{\mathcal{P}}_i(\mathbf{u}) = \frac{\hat{\mathcal{P}}_i^0 e^{u_i/\lambda}}{\sum_{j=1}^{N-1} \hat{\mathcal{P}}_j^0 e^{u_j/\lambda} + (\hat{\mathcal{P}}_N^0 + \hat{\mathcal{P}}_{N+1}^0) e^{u_N/\lambda}} \qquad \text{almost surely}, \forall i < N$$

$$\text{(E7)} \qquad \hat{\mathcal{P}}_N(\mathbf{u}) + \hat{\mathcal{P}}_{N+1}^0(\mathbf{u}) = \frac{(\hat{\mathcal{P}}_N^0 + \hat{\mathcal{P}}_{N+1}^0) e^{u_i/\lambda}}{\sum_{j=1}^{N-1} \hat{\mathcal{P}}_j^0 e^{u_j/\lambda} + (\hat{\mathcal{P}}_N^0 + \hat{\mathcal{P}}_{N+1}^0) e^{u_N/\lambda}} \qquad \text{almost surely}$$

Therefore, the right hand sides do not change when only $\hat{\mathcal{P}}_N^0$ and $\hat{\mathcal{P}}_{N+1}^0$ change if their sum stays constant. Inspecting (E6)-(E7), we see that any such strategy produces the same expected value as the original one. Moreover, the amount of processed information is also the same for both strategies. To show this we use (13) to rewrite (9) as:[1]

(E8)
$$\kappa = \int \sum_{i=1}^{N+1} \hat{\mathcal{P}}_i(\mathbf{u}) \log \frac{\hat{\mathcal{P}}_i(\mathbf{u})}{\hat{\mathcal{P}}_i^0} G(d\mathbf{u}) = \int \sum_{i=1}^{N+1} \hat{\mathcal{P}}_i(\mathbf{u}) \log \frac{e^{u_i/\lambda}}{\sum_{j=1}^{N-1} \hat{\mathcal{P}}_j^0 e^{u_j/\lambda} + (\hat{\mathcal{P}}_N^0 + \hat{\mathcal{P}}_{N+1}^0) e^{u_N/\lambda}} G(d\mathbf{u}).$$

Therefore, the achieved objective in (10) is the same for any such strategy as for the original strategy, and all of them solve the decision maker's problem.

Finally, even the corresponding strategy with $\hat{\mathcal{P}}_{N+1}^0 = 0$ is a solution. Moreover, this implies that the remaining $\{\hat{\mathcal{P}}_i^0\}_{i=1}^N$ is the solution to the problem without the duplicate action $N+1$, which completes the proof.

E3. Similar actions

Proof of Proposition 5. We proceed similarly as in the proof of Proposition 3 by showing that $\Delta_2 \equiv \frac{\partial E[U]}{\partial \mathcal{P}_1^0} + \frac{\partial E[U]}{\partial \mathcal{P}_2^0}$ decreases at all points $\{P_i^0\}_{i=1}^N$ after a marginal change of prior in the direction of interest. Notice that Δ_2 is a scalar product of the gradient of $E[U]$ and the vector $(1, 1, 0, \ldots, 0)$. We thus show that at each point, the gradient passes through each plane of constant $\mathcal{P}_1^0 + \mathcal{P}_2^0$ more in the direction of the negative change of $\mathcal{P}_1^0 + \mathcal{P}_2^0$ than before the change of the prior.

The analog to equation (E2), after relocating $\epsilon\Pi$ probability from state 1 to 3 and from state 2 to 4, the sum of the left hand sides of the first order conditions for $i = 1$ and $i = 2$, becomes:

$$\begin{aligned} \Delta_2 =& \lambda \int_{\mathbf{v}} \frac{e^{v_1/\lambda} + e^{v_2/\lambda} - e^{v_N/\lambda}}{\sum_{j=1}^N \mathcal{P}_j^0 e^{v_j/\lambda}} G(d\mathbf{v}) \\ &+ \lambda\Pi\Big(\frac{(1-\epsilon)(e^{H/\lambda} + e^{L/\lambda} - e^{v_N/\lambda})}{\mathcal{P}_1^0 e^{H/\lambda} + \mathcal{P}_2^0 e^{L/\lambda} + a} + \frac{(1-\epsilon)(e^{H/\lambda} + e^{L/\lambda} - e^{v_N/\lambda})}{\mathcal{P}_1^0 e^{L/\lambda} + \mathcal{P}_2^0 e^{H/\lambda} + a} \\ &+ \frac{\epsilon(2e^{H/\lambda} - e^{v_N/\lambda})}{\mathcal{P}_1^0 e^{H/\lambda} + \mathcal{P}_2^0 e^{H/\lambda} + a} + \frac{\epsilon(2e^{L/\lambda} - e^{v_N/\lambda})}{\mathcal{P}_1^0 e^{L/\lambda} + \mathcal{P}_2^0 e^{L/\lambda} + a}\Big), \end{aligned} \tag{E9}$$

where $a = \sum_{j=3}^N \mathcal{P}_j^0 e^{v_j/\lambda}$ is constant across the states 1-4, since for $j > 2$, v_j is constant there.

The analog to equation (E4) when we differentiate $\Delta_2 = \frac{\partial E[U]}{\partial \mathcal{P}_1^0} + \frac{\partial E[U]}{\partial \mathcal{P}_2^0}$ with respect to ϵ is:

$$\begin{aligned} \frac{\partial \Delta_2}{\partial \varepsilon}\Big|_{\varepsilon=0} =& \lambda\Big(-\frac{e^{H/\lambda} + e^{L/\lambda} - e^{v_N/\lambda}}{\mathcal{P}_1^0 e^{H/\lambda} + \mathcal{P}_2^0 e^{L/\lambda} + a} - \frac{e^{H/\lambda} + e^{L/\lambda} - e^{v_N/\lambda}}{\mathcal{P}_1^0 e^{L/\lambda} + \mathcal{P}_2^0 e^{H/\lambda} + a} \\ &+ \frac{2e^{H/\lambda} - e^{v_N/\lambda}}{\mathcal{P}_1^0 e^{H/\lambda} + \mathcal{P}_2^0 e^{H/\lambda} + a} + \frac{2e^{L/\lambda} - e^{v_N/\lambda}}{\mathcal{P}_1^0 e^{L/\lambda} + \mathcal{P}_2^0 e^{L/\lambda} + a}\Big). \end{aligned} \tag{E10}$$

Multiplying the right hand side by the positive denominators, the resulting expression can be

[1]Here we use the fact that the mutual information between random variables X and Y can be expressed as $E_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right]$. See Cover and Thomas (2006, p. 20).

re-arranged to

$$-\lambda(e^{H/\lambda} - e^{L/\lambda})^2$$
$$\Big[a^2(\mathcal{P}_1^0 + \mathcal{P}_2^0) + e^{H/\lambda}e^{L/\lambda}(\mathcal{P}_1^0 - \mathcal{P}_2^0)^2(\mathcal{P}_1^0 + \mathcal{P}_2^0) + a(e^{H/\lambda} + e^{L/\lambda})((\mathcal{P}_1^0)^2 + (\mathcal{P}_2^0)^2)$$
$$+ e^{v_N/\lambda}\mathcal{P}_1^0\mathcal{P}_2^0(2a + e^{H/\lambda}\mathcal{P}_1^0 + e^{L/\lambda}\mathcal{P}_1^0 + e^{H/\lambda}\mathcal{P}_2^0 + e^{L/\lambda}\mathcal{P}_2^0)\Big]$$

which is negative, and thus $\frac{\partial \Delta_2}{\partial \varepsilon}\Big|_{\varepsilon=0}$ is negative, too. After the marginal relocation of probabilities that makes the payoffs of actions 1 and 2 co-move more closely, the optimal $\mathcal{P}_1^0 + \mathcal{P}_2^0$ decreases. The treatment of the boundary cases is analogous to that in the proof of Proposition 3.

F. Derivations for examples

F1. Auxillary example

This is perhaps the simplest example of how rational inattention can be applied to a discrete choice situation. We present it here principally because this analysis forms the basis of our proof of Proposition 6 in Appendix F.F3, but also because it provides some additional insight into the workings of the model.

Suppose there are two actions, one of which has a known payoff while the payoff of the other takes one of two values. One interpretation is that the known payoff is an outside option or reservation value.

PROBLEM 1: *There are two states and two actions with payoffs:*

	state 1	*state 2*
action 1	0	1
action 2	R	R
prior prob.	g_0	$1 - g_0$

with $R \in (0, 1)$.

To solve the problem, we must find $\{\mathcal{P}_i^0\}_{i=1}^2$. We show below that the solution is:

(F1)
$$\mathcal{P}_1^0 = \max\left(0, \min\left(1, -\frac{e^{\frac{R}{\lambda}}\left(-e^{\frac{1}{\lambda}} + e^{\frac{R}{\lambda}} - g_0 + g_0 e^{\frac{1}{\lambda}}\right)}{\left(e^{\frac{1}{\lambda}} - e^{\frac{R}{\lambda}}\right)\left(-1 + e^{\frac{R}{\lambda}}\right)}\right)\right)$$
$$\mathcal{P}_2^0 = 1 - \mathcal{P}_1^0.$$

For a given set of parameters, the unconditional probability $\mathcal{P}_1^0$ as a function of R is shown in Figure $F1$. For R close to 0 or to 1, the decision maker decides not to process information and selects one of the actions with certainty. In the middle range however, the decision maker does process information and the selection of action 1 is less and less probable as the reservation value, R, increases, since action 2 is more and more appealing. For $g_0 = 1/2$ and $R = 1/2$, solutions take the form of the multinomial logit, i.e. $\mathcal{P}_1^0 = \mathcal{P}_2^0 = 1/2$. If the decision maker observed the values, he would choose action 1 with the probability $(1 - g_0) = 1/2$ for any reservation value R. However, the rationally inattentive agent chooses action 1 with higher probability when R is low.

Figure F2 again shows the dependance on R, but this time it presents the probability of selecting the first action *conditional* on the realized value $v_1 = 1$, it is $\mathcal{P}_1(1, R)$. Since $R < 1$, it would be optimal to always select the action 1 when its value is 1. The decision maker obviously does not

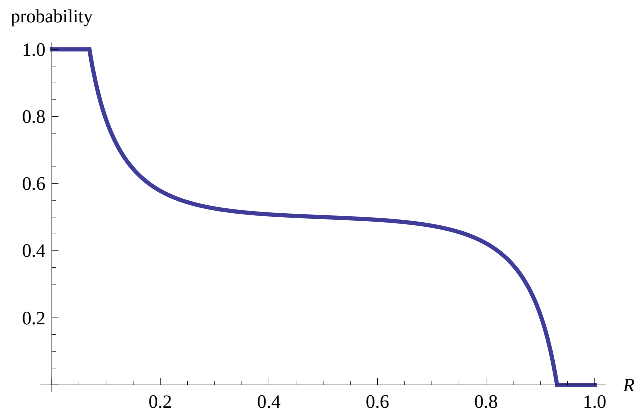

FIGURE F1. $\mathcal{P}_1^0$ AS A FUNCTION OF R AND $\lambda = 0.1, g_0 = 0.5$.

choose to do that because he is not sure what the realized value is. When R is high, the decision maker processes less information and selects a low $\mathcal{P}_1^0$. As a result, $\mathcal{P}_1(1, R)$ is low.

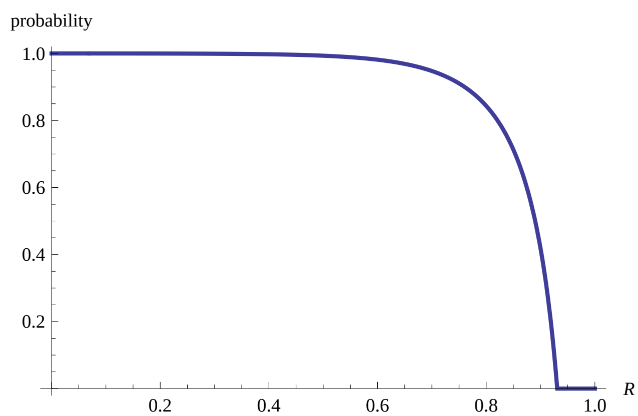

FIGURE F2. $\mathcal{P}_1(1, R)$ AS A FUNCTION OF R AND $\lambda = 0.1, g_0 = 0.5$.

In general, one would expect that as R increases, the decision maker would be more willing to reject action 1 and receive the certain value R. Indeed, differentiating the non-constant part of (F1) one finds that the function is non-increasing in R. Similarly, the unconditional probability of selecting action 1 falls as g_0 rises, as it is more likely to have a low payoff. Moreover, we see from equation (F1) that, for $R \in (0, 1)$, $\mathcal{P}_1^0$ equals 1 for g_0 in some neighborhood of 0 and it equals 0 for g_0 close to 1.[2] For these parameters, the decision maker chooses not to process information.

The following Proposition summarizes the immediate implications of equation (F1). Moreover, the findings hold for any values of the uncertain payoff $\{a, b\}$ such that $R \in (a, b)$.

PROPOSITION 1: *Solutions to Problem 1 have the following properties:*

1) *The unconditional probability of action 1, $\mathcal{P}_1^0$, is a non-increasing function of g_0 and the payoff of the other action, R.*

[2]The non-constant argument on the right-hand side of (F1) is continuous and decreasing in g_0, and it is greater than 1 at $g_0 = 0$ and negative at $g_0 = 1$.

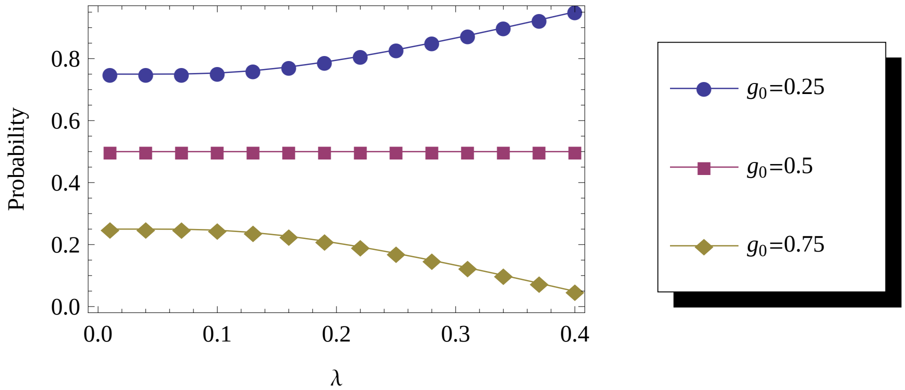

FIGURE F3. $\mathcal{P}_1^0$ AS A FUNCTION OF λ EVALUATED AT VARIOUS VALUES OF g_0 AND $R = 0.5$.

2) For all $R \in (0,1)$ and $\lambda > 0$, there exist g_m and g_M in $(0,1)$ such that if $g_0 \leq g_m$, the decision maker does not process any information and selects action 1 with probability one. Similarly, if $g_0 \geq g_M$, the decision maker processes no information and selects action 2 with probability one.

Figure F3 plots $\mathcal{P}_1^0$ as a function of the information cost λ for three values of the prior, g_0. When $\lambda = 0$, $\mathcal{P}_1^0$ is just equal to $1 - g_0$ because the decision maker will have perfect knowledge of the value of action 1 and choose it when it has a high value, which occurs with probability $1 - g_0$. As λ increases, $\mathcal{P}_1^0$ fans out away from $1 - g_0$ because the decision maker no longer possesses perfect knowledge about the payoff of action 1 and eventually just selects the action with the higher expected payoff according to the prior.

Solving for the choice probabilities in Problem 1. To solve the problem, we must find $\mathcal{P}_1^0$, while $\mathcal{P}_2^0 = 1 - \mathcal{P}_1^0$. These probabilities must satisfy the normalization condition, equation (17):

$$\text{(F2)} \qquad 1 = \frac{g_0}{\mathcal{P}_1^0 + \mathcal{P}_2^0 e^{\frac{R}{\lambda}}} + \frac{(1-g_0)e^{\frac{1}{\lambda}}}{\mathcal{P}_1^0 e^{\frac{1}{\lambda}} + \mathcal{P}_2^0 e^{\frac{R}{\lambda}}} \qquad \text{if } \mathcal{P}_1^0 > 0,$$

$$\text{(F3)} \qquad 1 = \frac{g_0 e^{\frac{R}{\lambda}}}{\mathcal{P}_1^0 + \mathcal{P}_2^0 e^{\frac{R}{\lambda}}} + \frac{(1-g_0)e^{\frac{R}{\lambda}}}{\mathcal{P}_1^0 e^{\frac{1}{\lambda}} + \mathcal{P}_2^0 e^{\frac{R}{\lambda}}} \qquad \text{if } \mathcal{P}_2^0 > 0.$$

There are three solutions to this system,

$$\text{(F4)} \qquad \begin{aligned} \mathcal{P}_1^0 &\in \left\{ 0, 1, -\frac{e^{\frac{R}{\lambda}}\left(-e^{\frac{1}{\lambda}} + e^{\frac{R}{\lambda}} - g_0 + g_0 e^{\frac{1}{\lambda}}\right)}{\left(e^{\frac{1}{\lambda}} - e^{\frac{R}{\lambda}}\right)\left(-1 + e^{\frac{R}{\lambda}}\right)} \right\} \\ \mathcal{P}_2^0 &= 1 - \mathcal{P}_1^0. \end{aligned}$$

Now, we make an argument using the solution's uniqueness to deduce the true solution to the decision maker's problem. The first solution to the system, $\mathcal{P}_1^0 = 0$, corresponds to the case when the decision maker chooses action 2 without processing any information. The realized utility is then R with certainty. The second solution, $\mathcal{P}_1^0 = 1$, results in the *a priori* selection of action 1 so the expected utility equals $(1 - g_0)$. The third solution describes the case when the decision maker chooses to process a positive amount of information.

In Problem 1, there are just two actions and they do not have the same payoffs with probability one. Therefore, Corollary 1 establishes that the solution to the decision maker's optimization problem must be unique.

Since the expected utility is a continuous function of $\mathcal{P}_1^0, R, \lambda$ and g_0, then the optimal $\mathcal{P}_1^0$ must be a continuous function of the parameters. Otherwise, there would be at least two solutions at the point of discontinuity of $\mathcal{P}_1^0$. We also know that, when no information is processed, action 1 generates higher expected utility than action 2 for $(1-g_0) > R$, and vice versa. So for some configurations of parameters $\mathcal{P}_1^0 = 0$ is the solution and for some configurations of parameters $\mathcal{P}_1^0 = 1$ is the solution. Therefore, the solution to the decision maker's problem has to include the non-constant branch, the third solution. To summarize this, the only possible solution to the decision maker's optimization problem is

$$\mathcal{P}_1^0 = \max\left(0, \min\left(1, -\frac{e^{\frac{R}{\lambda}}\left(-e^{\frac{1}{\lambda}} + e^{\frac{R}{\lambda}} - g_0 + g_0 e^{\frac{1}{\lambda}}\right)}{\left(e^{\frac{1}{\lambda}} - e^{\frac{R}{\lambda}}\right)\left(-1 + e^{\frac{R}{\lambda}}\right)}\right)\right). \tag{F5}$$

F2. Problem 3

To find the solution to Problem 3 we must solve for $\left\{\mathcal{P}_r^0, \mathcal{P}_b^0, \mathcal{P}_t^0\right\}$. The normalization condition $\mathcal{P}_r^0 = \int_{\mathbf{v}} \mathcal{P}_r(\mathbf{v}) G(d\mathbf{v})$ yields:

$$\begin{aligned} 1 = \quad & \frac{\frac{1}{4}(1+\rho)}{\mathcal{P}_r^0 + \mathcal{P}_b^0 + (1-\mathcal{P}_r^0 - \mathcal{P}_b^0)e^{1/2\lambda}} + \frac{\frac{1}{4}(1-\rho)\,e^{1/\lambda}}{\mathcal{P}_r^0 e^{1/\lambda} + \mathcal{P}_b^0 + (1-\mathcal{P}_r^0 - \mathcal{P}_b^0)e^{1/2\lambda}} \\ + \quad & \frac{\frac{1}{4}(1-\rho)}{\mathcal{P}_r^0 + \mathcal{P}_b^0 e^{1/\lambda} + (1-\mathcal{P}_r^0 - \mathcal{P}_b^0)e^{1/2\lambda}} + \frac{\frac{1}{4}(1+\rho)\,e^{1/\lambda}}{\mathcal{P}_r^0 e^{1/\lambda} + \mathcal{P}_b^0 e^{1/\lambda} + (1-\mathcal{P}_r^0 - \mathcal{P}_b^0)e^{1/2\lambda}} \end{aligned} \tag{F6}$$

Due to the symmetry between the buses, we know $\mathcal{P}_r^0 = \mathcal{P}_b^0$. This makes the problem one equation with one unknown, $\mathcal{P}_r^0$. The problem can be solved analytically using the same arguments as in Appendix F.F1. The resulting analytical expression is:

$$\mathcal{P}_r^0 = \max\left(0, \min\left(0.5, \frac{\left(\begin{array}{l} e^{\frac{1}{2\lambda}} - 8e^{\frac{1}{\lambda}} + 14e^{\frac{3}{2\lambda}} - 8e^{2/\lambda} + e^{\frac{5}{2\lambda}} \\ +\frac{1}{2}e^{\frac{1}{2\lambda}}(1-\rho) - e^{\frac{3}{2\lambda}}(1-\rho) + \frac{1}{2}e^{\frac{5}{2\lambda}}(1-\rho) \\ +e^{\frac{1}{2\lambda}}\left(-1+e^{\frac{1}{\lambda}}\right)x \end{array}\right)}{2\left(4e^{\frac{1}{2\lambda}} - 16e^{\frac{1}{\lambda}} + 24e^{\frac{3}{2\lambda}} - 16e^{2/\lambda} + 4e^{\frac{5}{2\lambda}}\right)}\right)\right),$$

where

$$x = \sqrt{\begin{array}{l} 2 - 2e^{\frac{1}{\lambda}} + e^{2/\lambda} - 8e^{\frac{1}{2\lambda}}(1-\rho) + 14e^{\frac{1}{\lambda}}(1-\rho) \\ -8e^{\frac{3}{2\lambda}}(1-\rho) + e^{2/\lambda}(1-\rho) + \frac{1}{4}(1-\rho)^2 \\ -\frac{1}{2}e^{\frac{1}{\lambda}}(1-\rho)^2 + \frac{1}{4}e^{2/\lambda}(1-\rho)^2 - \rho \end{array}}.$$

F3. Inconsistency with a random utility model

This appendix establishes that the behavior of the rationally inattentive agent is not consistent with a random utility model. The argument is based on the counterexample described in Section IV.C. Let Problem A refer to the choice among actions 1 and 2 and Problem B refer to the choice among all three actions. For simplicity, $\mathcal{P}_i(s)$ denotes the probability of selecting action i conditional on the state s, and $g(s)$ is the prior probability of state s.

LEMMA 3: *For all $\epsilon > 0$ there exists Y s.t. the decision maker's strategy in Problem B satisfies*

$$\mathcal{P}_3(1) > 1 - \epsilon, \qquad \mathcal{P}_3(2) < \epsilon.$$

Proof: For $Y > 1$, an increase of $\mathcal{P}_3(1)$ (decrease of $\mathcal{P}_3(2)$) and the corresponding relocation of the choice probabilities from (to) other actions increases the agent's expected payoff. The resulting marginal increase of the expected payoff is larger than $(Y - 1)\min(g(1), g(2))$. Selecting Y allows us to make the marginal increase arbitrarily large and therefore the marginal value of information arbitrarily large.

On the other hand, with λ being finite, the marginal change in the cost of information is also finite as long as the varied conditional probabilities are bounded away from zero. See equation (9), the derivative of entropy with respect to $\mathcal{P}_i(s)$ is finite at all $\mathcal{P}_i(s) > 0$. Therefore, for any ϵ there exists high enough Y such that it is optimal to relocate probabilities from actions 1 and 2 unless $\mathcal{P}_3(1) > 1 - \epsilon$, and to actions 1 and 2 unless $\mathcal{P}_3(2) < \epsilon$.

Proof of proposition 6. We will show that there exist $g(1) \in (0, 1)$ and $Y > 0$ such that action 1 has zero probability of being selected in Problem A, while the probability is positive in both states in Problem B. Let us start with Problem A. According to Proposition 1, there exists a sufficiently high $g(1) \in (0, 1)$, call it g_M, such that the decision maker processes no information and $\mathcal{P}_1(1) = \mathcal{P}_1(2) = 0$. We will show that for $g(1) = g_M$ there exists a high enough Y, such the choice probabilities of action 1 are positive in Problem B.

Let $\mathcal{P} = \{\mathcal{P}_i(s)\}_{i=1,s=1}^{3,2}$ be the solution to Problem B. We now show that the optimal choice probabilities of actions 1 and 2, $\{\mathcal{P}_i(s)\}_{i=1,s=1}^{2,2}$, solve a version of Problem A with modified prior probabilities. The objective function for Problem B is

$$\begin{aligned}\max_{\{\mathcal{P}_i(s)\}_{i=1,s=1}^{3,2}} &\sum_{i=1}^{3}\sum_{s=1}^{2} v_i(s)\mathcal{P}_i(s)g(s) \\ &- \lambda\left[-\sum_{s=1}^{2} g(s)\log g(s) + \sum_{i=1}^{3}\sum_{s=1}^{2}\mathcal{P}_i(s)g(s)\log\frac{\mathcal{P}_i(s)g(s)}{\sum_{s'}\mathcal{P}_i(s')g(s')}\right],\end{aligned} \tag{F7}$$

where we have written the information cost as $H(s) - E[H(s|i)]$.[3] If $\mathcal{P}_3(1)$ and $\mathcal{P}_3(2)$ are the conditional probabilities of the solution to Problem B, the remaining conditional probabilities solve the following maximization problem.

$$\max_{\{\mathcal{P}_i(s)\}_{i=1,s=1}^{2,2}} \sum_{i=1}^{2}\sum_{s=1}^{2} v_i(s)\mathcal{P}_i(s)g(s) - \lambda\left[\sum_{i=1}^{2}\sum_{s=1}^{2}\mathcal{P}_i(s)g(s)\log\frac{\mathcal{P}_i(s)g(s)}{\sum_{s'}\mathcal{P}_i(s')g(s')}\right], \tag{F8}$$

[3]Recall that $H(Y|X) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(y|x)$ (Cover and Thomas, 2006, p. 17).

subject to $\mathcal{P}_1(s) + \mathcal{P}_2(s) = 1 - \mathcal{P}_3(s)$, $\forall s$. Equation (F8) is generated from (F7) by omitting the terms independent of $\{\mathcal{P}_i(s)\}_{i=1,s=1}^{2,2}$. Now, we make the following substitution of variables.

$$\mathcal{R}_i(s) = \mathcal{P}_i(s)/\Big(1 - \mathcal{P}_3(s)\Big) \tag{F9}$$

$$\hat{g}(s) = Kg(s)\Big(1 - \mathcal{P}_3(s)\Big) \tag{F10}$$

$$1/K = \sum_{s=1}^{2} g(s)\Big(1 - \mathcal{P}_3(s)\Big). \tag{F11}$$

where K, which is given by (F11), is the normalization constant that makes the new prior, $\hat{g}(s)$, sum up to 1.

The maximization problem (F8) now takes the form:

$$\max_{\{\mathcal{R}_i(s)\}_{i=1,s=1}^{1,2}} \sum_{i=1}^{2}\sum_{s=1}^{2} v_i(s)\mathcal{R}_i(s)\hat{g}(s) - \lambda \sum_{i=1}^{2}\sum_{s=1}^{2} \mathcal{R}_i(s)\hat{g}(s) \log \frac{\mathcal{R}_i(s)\hat{g}(s)}{\sum_{s'} \mathcal{R}_i(s')\hat{g}(s')}, \tag{F12}$$

subject to

$$\mathcal{R}_1(s) + \mathcal{R}_2(s) = 1 \quad \forall s. \tag{F13}$$

The objective function of this problem is equivalent to (F8) up to a factor of K, which is a positive constant. The optimization problem (F12) subject to (F13) is equivalent to Problem A with the prior modified to from $g(s)$ to $\hat{g}(s)$, let us call it Problem C.[4]

According to Proposition 1, there exists $\hat{g}_m \in (0,1)$ such that the decision maker always selects action 1 in Problem C for all $\hat{g}(1) \leq \hat{g}_m$. From equations (F10) and (F11) we see that for any $\hat{g}_m > 0$ and $g(1)$, $g(2) \in (0,1)$ there exists $\epsilon > 0$ such that if $\mathcal{P}_3(1) > 1 - \epsilon$ and $\mathcal{P}_3(2) < \epsilon$, then $\hat{g}(1) < \hat{g}_m$.[5] Moreover, Lemma 3 states that for any such $\epsilon > 0$ there exists Y such that $\mathcal{P}_3(1) > 1 - \epsilon$ and $\mathcal{P}_3(2) < \epsilon$. Therefore there is a Y such that in Problem C, action 1 is selected with positive probability in both states, which also implies it is selected with positive probabilities in Problem B, see equation (F9).

*

REFERENCES

Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of Information Theory.* Hoboken, NJ:Wiley.

[4]To see the equivalence to Problem A, observe that this objective function has the same form as (**??**) except for a) the constant corresponding to $H(s)$ and b) we only sum over $i = 1, 2$.

[5]$\hat{g}(1) = \frac{g(1)(1-\mathcal{P}_3(1))}{\sum_s g(s)(1-\mathcal{P}_3(s))} < \frac{g(1)(1-\mathcal{P}_3(1))}{g(2)(1-\mathcal{P}_3(2))} < \frac{g(1)\epsilon}{g(2)(1-\epsilon)}$.