The Behavioral Implications of Rational Inattention with Shannon Entropy - Online Appendix

Not for Publication

August 6, 2013

1 Proofs of Lemmas

Lemma 1 Given $A \in \mathcal{F}$ and $G \in \mathcal{G}$, the set $\mathcal{E}_G(A) \subset \mathbb{R}^M$ defined by,

$$\mathcal{E}_G(A) \equiv \left\{ \begin{array}{c} (y, \mu_1, .., \mu_{M-1}) \in \mathbb{R} \times X | \\ \exists A \in \mathcal{F} \text{ and } (B, P, \gamma) \in \Lambda^{(\mu, A)} \text{ s.t. } y \leq \sum_{f \in B} P^f N_G^f(\gamma^f) \end{array} \right\},$$

is closed, convex, and bounded above in its first coordinate.

Proof. Given $A \in \mathcal{F}$ and $G \in \mathcal{G}$, consider $(y, \mu_1, .., \mu_{M-1})$, $(\tilde{y}, \tilde{\mu}_1, .., \tilde{\mu}_{M-1}) \in \mathcal{E}_G(A)$ together with finite sets $B, \tilde{B} \subset A$, probabilities on actions and associated posteriors, P^f, γ_m^f for $f \in B$ and $\tilde{P}^f, \tilde{\gamma}_m^f$ for $f \in \tilde{B}$, all $1 \leq m \leq M-1$ such that,

$$\begin{aligned} \mu_m &= \sum_{f \in B} P^f \gamma_m^f \text{ and } y \leq \sum_{f \in B} P^f N_G^f(\gamma^f); \\ \tilde{\mu}_m &= \sum_{f \in \tilde{B}}^{f} \tilde{P}^f \tilde{\gamma}_m^f \text{ and } \tilde{y} \leq \sum_{f \in \tilde{B}} \tilde{P}^f N_G^f(\tilde{\gamma}^f). \end{aligned}$$

Define $C = B \cup \tilde{B}$ and extend $P^f, \tilde{P}^f$ to this domain by setting them to zero on the unchosen acts.

Given $\lambda \in (0,1)$, define $R^f = \lambda P^f + (1-\lambda)\tilde{P}^f$ and

$$\eta_m^f = \frac{\lambda P^f \gamma_m^f + (1-\lambda)\tilde{P}^f \tilde{\gamma}_m^f}{\lambda P^f + (1-\lambda)\tilde{P}^f}.$$

It is immediate that $\eta^f \in \Gamma$ all $f \in C$ and that $\sum\limits_{f \in C} R^f \eta_m^f = \lambda \mu_m + (1-\lambda)\tilde{\mu}_m$ so that $(C, \eta, R) \in \Lambda^{\frac{\mu+\tilde{\mu}}{2}}$ Note that, for each $f \in C$

$$\begin{aligned}
& N_G^f\left(\eta^f\right) \\
= & \sum_{j=1}^{M} \eta_m^f U_m^f - G(\eta^f) \\
= & \frac{\lambda P^f}{\lambda P^f + (1-\lambda)\tilde{P}^f} \sum_{j=1}^{M} \gamma_m^f U_m^f + \frac{(1-\lambda)\tilde{P}^f}{\lambda P^f + (1-\lambda)\tilde{P}^f} \sum_{j=1}^{M} \bar{\gamma}_m^f U_m^f - G(\eta^f) \\
\geq & \frac{\lambda P^f}{\lambda P^f + (1-\lambda)\tilde{P}^f} N_G^f(\gamma^f) + \frac{(1-\lambda)\tilde{P}^f}{\lambda P^f + (1-\lambda)\tilde{P}^f} N_G^f(\tilde{\gamma}^f),
\end{aligned}$$

By the convexity of G

Thus we have that

$$\begin{aligned}
\sum_{f \in C} R^f N_G^f\left(\eta^f\right) &= \sum_{f \in C} \left(\lambda P^f + (1-\lambda)\tilde{P}^f\right) N_G^f\left(\eta^f\right) \\
&\geq \lambda \sum_{f \in B} P^f N_G^f(\gamma^f) + (1-\lambda) \sum_{f \in \tilde{B}} \tilde{P}^f N_G^f(\tilde{\gamma}^f) = \lambda y + (1-\lambda)\bar{y},
\end{aligned}$$

confirming that $\lambda\left(y, \mu_1, .., \mu_{M-1}\right) + (1-\lambda)(\tilde{y}, \tilde{\mu}_1, .., \tilde{\mu}_{M-1}) \in \mathcal{E}_G(A)$

To establish closedness, consider a sequence $(y(n), \mu(n)) \in \mathcal{E}_G(A)$ converging to (y^L, μ^L) (to simplify notation we use the full prior as the second argument since μ_M is anyway implied) and corresponding triples $(B(n), P(n), \gamma(n)) \in \Lambda^{(\mu(n), A)}$, so that $\mu(n) = \sum_{f \in B(n)} P^f(n)\gamma^f(n)$ and $y(n) \leq \sum_{f \in B(n)} P^f(n) N_G^f(\gamma^f(n))$.

We show now that there is no loss of generality in assuming $|B(n)| \leq M + 1$. Suppose initially that $|B(n)| > M$. By Charateodory's theorem, since $\{\gamma^f(n) \in \Gamma | f \in B(n)\}$ contain $\mu(n)$ in its convex hull, there exists $B_1(n) \subset B(n)$ with $|B_1(n)| \leq M + 1$ for which there exists a strictly positive probability weights $P_1^f(n) > 0$ on $f \in B_1(n)$ such that $\mu = \sum_{f \in B^1(n)} P_1^f(n)\gamma^f(n)$. If expected net utility is no lower,

$$y(n) \leq \sum_{f \in B(n)} P^f(n) N_G^f(\gamma^f(n)) \leq \sum_{f \in B_1(n)} P_1^f(n) N_G^f(\gamma^f(n)),$$

we are done. If not, identify the smallest scalar $\alpha_1 \in (0, 1)$ such that,

$$\alpha_1 P_1^f(n) = P^f(n),$$

some $f \in B_1(n)$. That such a scalar exists follows from the fact that

$$\sum_{f \in B_1(n)} P_1^f(n) = \sum_{f \in B(n)} P^f(n) = 1,$$

with all components in both sums strictly positive and with $|B(n)| > |B_1(n)|$.

We now define a second set of probability weights $P_2^f(n)$,

$$P_2^f(n) = \frac{P^f(n) - \alpha_1 P_1^f(n)}{1 - \alpha^1}.$$

for $f \in B_1(n)$. Correspondingly, we define,

$$B_2(n) = \{f \in B(n) | P_2^f(n) > 0\},$$

noting that $|B_2(n)| \leq |B(n)| - 1$. By construction $\mu = \sum_{f \in B(n)} P_2^f(n)\gamma^f(n)$. Moreover,

$$\sum_{f \in B_2(n)} P_2^f(n) N_G^f(\gamma^f(n)) = \sum_{f \in B(n)} \left[\frac{P^f(n) - \alpha_1 P_1^f(n)}{1 - \alpha^1} \right] N_G^f(\gamma^f(n)) > \sum_{f \in B(n)} P^f(n) N_G^f(\gamma^f(n)).$$

Iteration from this point establishes that indeed we can identify a set $\tilde{B}(n) \subset B(n)$ with $\left|\tilde{B}(n)\right| \leq M+1$ and $\tilde{P}(n) > 0$ on $f \in \tilde{B}(n)$ such that $\mu = \sum_{f \in \tilde{B}(n)} P_1^f(n)\gamma^f(n)$ and,

$$\sum_{f \in \tilde{B}(n)} P_1^f(n) N_G^f(\gamma^f(n)) \geq y(n).$$

Given this, there is no loss of generality in assuming that $|B(n)| \leq M+1$ in our original sequence.

With this, we can focus on a subsequence (we continue to index by n for notational simplicity) with all sets $B(n)$ of the same cardinality $K \leq M$. In each set $B(n)$ we index the acts in (arbitrary) order by $f(k,n) \in$ for $1 \leq k \leq K$, and correspondingly label that associated posteriors and act probabilities as $\gamma^k(n), P^k(n)$. Given the compactness of Γ, we can further select subsequences to ensure that there is a full set of limit posteriors and limit probabilities $\bar{\gamma}^k$ and $\bar{P}^k$, for $1 \leq k \leq K$,

$$\lim_{n \to \infty} \gamma^k(n) = \bar{\gamma}^k; \lim_{n \to \infty} P^k(n) = \bar{P}^k$$

For all acts $f \in A$, we can compute the net utility at all limit posteriors,

$$N_G^f(\bar{\gamma}^k) = \sum_{m=1}^{M} U_m^f \bar{\gamma}_m^k - G(\bar{\gamma}^k)$$

Since $\{U_m^f \in \mathbb{R}^M | f \in A\}$ is bounded above then so is $N_G^f(\bar{\gamma}^k)$ (with respect to $f \in A$). . Since $\{U_m^f \in \mathbb{R}^M | f \in A\}$ is closed, the upper bound is achieved. Hence we can find acts $\bar{f}(k) \in A$ that maximize the above net utilities,

$$N(\bar{f}(k), k) \geq N(f, k),$$

all $f \in A$.

We now define $\bar{B} = \cup_{k=1}^{K} \bar{f}(k)$. Note that, by construction

$$\sum_{k=1}^{K} \bar{P}^k \bar{\gamma}^k = \mu^L,$$

so that $(\bar{B}, \bar{\gamma}, \bar{P}) \in \Lambda^{(\mu^L, A)}$. Note also that, for each for all n,

$$\sum_{k=1}^{K} \bar{P}^k N^{\bar{f}(k)}(\bar{\gamma}^k) \geq \sum_{k=1}^{K} \bar{P}^k N_G^{f(k,n)}(\bar{\gamma}^k).$$

In light of continuity of all functions N_G^f, taking the limit on the RHS as $n \to \infty$ yields,

$$\sum_{k=1}^{K} \bar{P}^k N^{\bar{f}(k)}(\bar{\gamma}^k) \geq \lim_{n\to\infty} \sum_{k=1}^{K} P^k(n) N_G^f(\gamma^f(n)) \geq y^L, \tag{1}$$

This completes the proof that $(y^L, \mu^L) \in \mathcal{E}_G(A)$, hence that $\mathcal{E}_G(A)$ is closed. Boundedness above of the first coordinate follows from the fact that $\{U_m^f \in \mathbb{R}^M | f \in A\}$ is bounded above for all $A \in \mathcal{F}$. ■

Lemma 2 Strategy $(B, P, \gamma) \in \Lambda^{(\mu, A)}$ is rationally inattentive for $G \in \mathcal{G}$ if and only if it there exists λ_m for $1 \leq m \leq M-1$ such that **property SH** holds:

$$N_G^g(\gamma) - \sum_{m=1}^{M-1} \lambda_m \gamma_m \leq N_G^f(\gamma^f) - \sum_{m=1}^{M-1} \lambda_m \gamma_m^f;$$

all $f \in B$, $g \in A$ and $\gamma \in \Gamma$.

Proof. Necessity: Given $(B, P, \gamma) \in \hat{\Lambda}_G^{(\mu, A)}$, $\left(\sum_{f \in B} P^f N_G^f(\gamma^f), \mu_1, .., \mu_{M-1} \right)$ is an upper boundary point boundary of $\mathcal{E}_G(A)$. Lemma 1 establishes that such sets are always closed, convex, and bounded above in the first coordinate. This implies existence of a supporting hyperplane defined by normal vector

$(1, -\lambda_1, ..., -\lambda_{M-1})$ such that, for all $(y_0, y_1, .., y_{M-1}) \in \mathcal{E}_G(A)$,

$$y_0 - \sum_{m=1}^{M-1} \lambda_m y_m \leq \sum_{f \in B} P^f N_G^f(\gamma^f) - \sum_{m=1}^{M-1} \lambda_m \mu_m = \sum_{f \in B} P^f [N_G^f(\gamma^f) - \lambda_m \gamma_m^f]. \quad (2)$$

We show now property SH is satisfied for such a normal vector. Substitution of $(N_G^f(\gamma^f), \gamma_1^f, .., \gamma_{M-1}^f) \in \mathcal{E}_G(A)$ on the LHS for $f \in B$ yields,

$$N_G^f(\gamma^f) - \sum_{m=1}^{M-1} \lambda_m \gamma_m^f \leq \sum_{f \in B} P^f [N_G^f(\gamma^f) - \lambda_m \gamma_m^f].$$

This implies that these inequalities are in fact equations for all $f \in B$, since this is the only way to prevent one of the sums on the RHS from being strictly higher than their weighted average on the LHS. This implies that $N_G^f(\gamma^f) - \sum_{m=1}^{M-1} \lambda_m \gamma_m^f$ can be plugged in to the right hand side of equation 2, which in turn establishes that, given $f, g \in B$, and $\gamma \in \Gamma$

$$N_G^g(\gamma^g) - \sum_{m=1}^{M-1} \lambda_m \gamma_m^g = N_G^f(\gamma^f) - \sum_{m=1}^{M-1} \lambda_m \gamma_m^f,$$

as necessary for property SH. Again, equation 2 tells us that all $f \in B$, γ^f solves,

$$\max_{\gamma \in \Gamma} N_G^f(\gamma) - \sum_{m=1}^{M-1} \lambda_m \gamma_m,$$

as again required for SH. The final aspect of condition SH to confirm is that, given $f \in B$, $g \in A \backslash B$ and $\gamma \in \Gamma$

$$N_G^g(\gamma) - \sum_{m=1}^{M-1} \lambda_m \gamma_m \leq N_G^f(\gamma^f) - \sum_{m=1}^{M-1} \lambda_m \gamma_m^f,$$

This is again immediate from 2 since $(N^g(\gamma), \gamma_1, .., \gamma_{M-1}) \in \mathcal{E}_G(A)$.

Sufficiency: If property SH holds, it directly implies existence of a normal vector $(1, -\lambda_1, ..., -\lambda_{M-1})$ such that, given $(y_0, y_1, .., y_{M-1}) \in \mathcal{E}_G(A)$,

$$y_0 - \sum_{m=1}^{M-1} \lambda_m y_m \leq N_G^f(\gamma^f) - \lambda_m \gamma_m^f,$$

any $f \in B$. Applying lemma 1, this implies that all points $(N_G^f(\gamma^f), \gamma_1^f, .., \gamma_{M-1}^f) \in \mathcal{E}_G(A)$ are in the upper boundary of $\mathcal{E}_G(A)$. Hence this applies also to any convex combination of them such as that defined by $(\sum_{f\in B} P^f N_G^f(\gamma^f), \mu_1, .., \mu_{M-1}) \in \mathcal{E}_G(A)$, completing the proof. ■

Lemma 3 Given $G \in \mathcal{G}$ that is differentiable on Γ^I, the interior of Γ, strategy $(B, P, \gamma) \in \Lambda^{(\mu,A)}$ with $\gamma_m^f \in (0,1)$ satisfies $(B, P, \gamma) \in \hat{\Lambda}_G^{(\mu,A)}$ if and only if it satisfies CT, ED, and UB:

1. **Common Tangent for Chosen Acts (CT)**: Given $f, g \in B$,

$$N_G^f(\gamma^f) - \sum_{m=1}^{M-1} \left[\frac{\partial N_G^f(\gamma^f)}{\partial \gamma_m}\right] \gamma_m^f = N_G^g(\gamma^g) - \sum_{m=1}^{M-1} \left[\frac{\partial N_G^g(\gamma^g)}{\partial \gamma_m}\right] \gamma_m^g.$$

2. **Equal Derivative for Chosen Acts (ED)** : Given $f, g \in B$,

$$\frac{\partial N_G^f(\gamma^f)}{\partial \gamma_m} = \frac{\partial N_G^g(\gamma^g)}{\partial \gamma_m}.$$

3. **Unchosen Act Bound (UB) :** Given $f \in B$ and $g \in B \backslash A$,

$$N_G^g(\gamma^g) - \sum_{m=1}^{M-1} \left[\frac{\partial N_G^f(\gamma^f)}{\partial \gamma_m}\right] \gamma_m^g \leq N_G^f(\gamma^f) - \sum_{m=1}^{M-1} \left[\frac{\partial N_G^f(\gamma^f)}{\partial \gamma_m}\right] \gamma_m^f,$$

   where $\gamma^g \in \Gamma$ maximizes on $N_G^g(\gamma) - \sum_{m=1}^{M-1} \left[\frac{\partial N_G^f(\gamma^f)}{\partial \gamma_m}\right] \gamma_m$ on $\gamma \in \Gamma$.

Proof. In light of lemma 2, the first part requires us to show that, when $G \in \mathcal{G}$ is differentiable, property SH is satisfied for $(B, P, \gamma) \in \Lambda^{(\mu,A)}$ with $\gamma_m^f \in (0,1)$

if and only if (B, P, γ) satisfies conditions ED, CT, and UB. That these three conditions are sufficient for property SP is immediate using $\lambda_m = \frac{\partial N_G^f(\gamma^f)}{\partial \gamma_m}$ for any $f \in B$ and applying UB. That they are necessary for SP to be satisfied in cases with $\gamma_m^f \in (0,1)$ and with G differentiable derives from the fact that SP certainly requires that, for each $f \in B$, γ^f solves,

$$\max_{\gamma \in \Gamma} N_G^f(\gamma) - \sum_{m=1}^{M-1} \lambda_m \gamma_m.$$

Given that $\gamma_m^f \in (0,1)$ and that $G \in \mathcal{G}$ is differentiable, solving this problem requires $\lambda_m = \frac{\partial N_G^f(\gamma^f)}{\partial \gamma_m}$. Given this, SP directly implies CT, ED, and UB as illustrated in the proof of lemma 1. ■

2 Proofs of Corollaries

All corollaries apply more generally than to the Shannon model. However for consistency with the text they are stated only for this case. The appropriate generalization to separable cost functions is in each case clear.

Corollary 1 (Locally Invariant Posteriors - LIP): If $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,A)}$ and $(C, Q, \eta) \in \Lambda^{(\rho,A)}$ with $C \subset B$ satisfies $\eta^f = \gamma^f$ all $f \in C$, then $(C, Q, \eta) \in \hat{\Lambda}^{(\rho,A)}$.

Proof. Note by the necessity aspect of lemma 2 that if $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,A)}$ then condition SH is satisfied. Neither the prior $\mu \in \Gamma$ nor the probability map $P : B \to \mathbb{R}$ feature in condition SP, while deletion of acts can only weaken the check. Hence if $(C, Q, \eta) \in \Lambda^{(\rho,A)}$ with $C \subset B$ satisfies $\eta^f = \gamma^f$ all $f \in C$, condition SH remains valid and the sufficiency aspect of lemma 2 implies it is optimal. ■

Corollary 2 (Envelope Condition): Given $(\mu, A) \in \Gamma \times \mathcal{F}$ such that $\mu_m > 0$, the value function $V^A : \Gamma \longrightarrow \mathbb{R}$ is differentiable at μ and has continuous partial derivatives,

$$\frac{\partial V^A(\mu)}{\partial \mu_m} = \frac{\partial N^f}{\partial \gamma_m}(\hat{\gamma}^f),$$

where $f \in \hat{B}$ some $(\hat{B}, \hat{P}, \hat{\gamma}) \in \hat{\Lambda}^{(\mu,A)}$.

Proof. Given $(\mu, A) \in \Gamma \times \mathcal{F}$, note that $\left(\mu_1, .., \mu_{M-1}\right)$ is in the interior of $X = \{(\mu_1, .., \mu_{M-1}) \in \mathbb{R}_+^{M-1} | \sum_{m=1}^{M-1} \mu_m \leq 1\}$. By lemma 1 an optimal policy exists. Consider a corresponding optimal strategy $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,A)}$ that therefore achieves the value,

$$V^A(\mu) = \sum_{f \in B} P^f N^f(\gamma^f).$$

Define a composite act $\hat{h} \in F$ with state dependent payoffs,

$$U_m^{\hat{h}} = \sum P^f U_m^f.$$

Define the net payoff function to $N^{\hat{h}} : \Gamma \to \mathbb{R}$ in standard fashion, and apply the envelope theorem of Benveniste and Scheinkman [1979] to functions $V^A, N^{\hat{h}} : \Gamma \to \mathbb{R}$, noting that both are concave, that $V(\mu) \geq N^{\hat{h}}(\mu)$ on the interior of X, and that $V(\mu) = N^{\hat{h}}(\mu)$, and that $N^{\hat{h}}(\mu)$ is differentiable on the interior of X. With this we conclude that V^A is differentiable at μ and that,

$$\frac{\partial V^A}{\partial \mu_m}(\mu) = \frac{\partial N^{\hat{h}}}{\partial \gamma_m}(\mu) = \sum_{f \in B} P^f \frac{\partial N^f}{\partial \gamma_m}(\gamma^f).$$

By lemma 3, the ED conditions are satisfied,

$$f, g \in B \Longrightarrow \frac{\partial N^f}{\partial \gamma_m}(\gamma^f) = \frac{\partial N^g}{\partial \gamma_m}(\gamma^g),$$

completing the proof in light of $\sum_{f \in B} P^f = 1$. ■

Corollary 3 (States Bound Acts - SBA): Given $(\mu, A) \in \Gamma \times \mathcal{F}$, there exists a rationally inattentive strategy with $|B| \leq M$.

Proof. Consider $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,A)}$ such that $|B| > M$. By the necessity aspect of lemma 2, condition SH is satisfied. This condition remains valid for any subset of acts $\tilde{B} \subset B$ with $\tilde{\gamma}^f = \gamma^f$ on $f \in \tilde{B}$. By the sufficiency aspect of lemma 2, $\left(\tilde{B}, \tilde{P}, \tilde{\gamma}\right) \in \hat{\Lambda}^{(\mu,A)}$ provided only that μ in the convex hull of the family of vectors $\left\{\gamma_m^f\right\}_{f \in \tilde{B}}$. Charateodory's theorem implies that we can reduce the cardinality of B to M while retaining μ in this convex hull, completing the proof. ■

Corollary 4 (Unique Posteriors): If $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,A)}$ and $(B, Q, \rho) \in \hat{\Lambda}^{(\mu,C)}$, then $\gamma(f) = \rho(f)$ all $f \in B$.

Proof. Note first that if $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,A)}$ and $B \subset C \subset A$, then $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,C)}$. To see this, note from the necessity aspect of lemma 2 that if $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,A)}$, then condition SH is satisfied. Since $B \subset C \subset A$, $(B, P, \gamma) \in \Lambda^{(\mu,C)}$ and condition SH is still satisfied. Hence $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,C)}$ follows in light of the sufficiency aspect of lemma 2. We conclude that since $(B, P, \gamma) \in \hat{\Lambda}^{(\mu,A)}$ and $(B, Q, \rho) \in \hat{\Lambda}^{(\mu,C)}$, then $(B, P, \gamma), (B, Q, \rho) \in \hat{\Lambda}^{(\mu,B)}$.

Given $f \in B$ define $R^f = \frac{P^f + Q^f}{2}$ and $\eta^f \in \Gamma$ by,

$$\eta_m^f = \frac{P^f \gamma_m^f + Q^f \rho_m^f}{P^f + Q^f}.$$

By construction, $(B, \eta, R) \in \Lambda^{(\mu,B)}$. If $\gamma(f) \neq \rho(f)$ some $f \in B$, we can apply the strict version of Jensen's inequality as in lemma 1 to establish the contradiction that net utility must be strictly higher at (B, η, R) than at either (B, P, γ) and (B, Q, ρ),

$$\sum_{f \in B} R^f N^f(\eta^f) > \lambda \sum_{f \in B} P^f N^f(\gamma^f) + (1 - \lambda) \sum_{f \in B} Q^f N^f(\rho^f).$$

■

3 Comparison with KKT Conditions

Following Matejka and McKay [2011], consider the constrained optimization problem of maximizing expected prize utility less Shannon attention costs, subject to constraints associated with rational expectations, with act-specific posteriors adding to unity, and with probabilities being non-negative. Let $\pi \in \mathbb{R}^M$ be the multipliers on the rational expectations constraints and $\eta : A \to \mathbb{R}$ the multipliers on posteriors. With act set A countable, the associated Lagrangean is,

$$\mathcal{L} = \sum_{f \in A} P^f \sum_{m=1}^{M} \gamma_m^f (U_m^f - \kappa \ln \gamma_m^f) - \sum_{m=1}^{M} \pi_m (\sum_{f \in A} P^f \gamma_m^f - \mu_m) - \eta^f \, (\sum_{m=1}^{M} \gamma_m^f - 1).$$

Treating this using standard KKTcondition, a necessary condition for (B, P, γ) to be rationally inattentive is that there exists $\hat{\pi} \in \mathbb{R}^M$, $\hat{\eta} : A \to \mathbb{R}$, and posteriors $\gamma_m^f \in \Gamma$ for $f \in A/B$ such that conditions KKT 1, KKT 2, and KKT 3 are satisfied:

KKT1: For $f \in B$,

$$P^f \left[U_m^f - \kappa \ln \gamma_m^f - \kappa - \hat{\pi}_m \right] - \hat{\eta}^f \; = 0 \text{ for } 1 \leq m \leq M.$$

KKT2: For $f \in B$, if

$$\begin{aligned} P^f &\in (0,1) \Longrightarrow \sum_{m=1}^{M} \gamma_m^f \left(U_m^f - \hat{\pi}_m - \kappa \ln \gamma_m^f \right) = 0; \\ P^f &= 1 \Longrightarrow \sum_{m=1}^{M} \gamma_m^f \left(U_m^f - \hat{\pi}_m - \kappa \ln \gamma_m^f \right) \geq 0. \end{aligned}$$

KKT3: For $f \in A/B$,

$$\sum_{m=1}^{M} \gamma_m^f \left(U_m^f - \hat{\pi}_m - \kappa \ln \gamma_m^f \right) \leq 0.$$

The reason that these KKT conditions do not characterize rationally inattentive policies is that the objective function is not concave in the choice variables, involving as it does product terms as between beliefs and posteriors. As a result, one can find non-optimal solutions. To illustrate, consider the case in the text with two acts, f and g, with $\kappa = 1$, and with $U_1^f = U_2^g = \ln(1+e)$ and $U_2^f = U_1^g = 0$. Note that an attention strategy can be fully specified by P^f and $\gamma_1^{f,g} \in [0,1] \geq 0$. Now consider the equal prior $\mu = 0.5$ and note that the following strategy is feasible and, together with the specified multipliers, satisfies all KKT necessary conditions:

$$(P^f, \gamma_1^f, \gamma_1^g) = (1, 0.5, 0.5); \hat{\pi}_1 = \ln(1+e) + \ln 2; \hat{\pi}_2 = \ln 2, \hat{\eta}^1 = \hat{\eta}^2 = 0.$$

Yet $(P^f, \gamma_1^1, \gamma_1^2)$ is not optimal, since net utility to the feasible triple $(0.5, \frac{1+e}{2+e}, \frac{1}{2+e})$ is strictly higher,

$$N(0.5, \frac{1+e}{2+e}, \frac{1}{2+e}) = \left(\frac{1+e}{2+e} \right) \ln(1+e) - \ln 0.5 > \frac{\ln(2+e)}{2} - \ln 0.5 = N(1, 0.5, 0.5).$$