Online Appendix for "Dynamic Information Design with Diminishing Sensitivity Over News"

Jetlir Duraj Kevin He

September 6, 2020

OA 1 Omitted Proofs from the Appendix

OA 1.1 Proof of Proposition 2

Proof. (1) Suppose μ is two-part linear with $\mu(x) = x$ for $x \geq 0$, $\mu(x) = \lambda x$ for $x < 0$, where $\lambda \geq 0$. Suppose $v(c_A) = 1$, $v(c_B) = 0$. In each period, $\mathbb{E}[\mu(\pi_t - \pi_{t-1})] = \mathbb{E}[(\pi_t - \pi_{t-1})^+ - \lambda(\pi_t - \pi_{t-1})^-]$. By the martingale property, $\mathbb{E}[(\pi_t - \pi_{t-1})^+] = \mathbb{E}[(\pi_t - \pi_{t-1})^-]$, so $\mathbb{E}[\mu(\pi_t - \pi_{t-1})] = \frac{1}{2}(1-\lambda)\mathbb{E}[|\pi_t - \pi_{t-1}|]$. This shows total expected news utility is $\mathbb{E}[\sum_{t=1}^T \mu(\pi_t - \pi_{t-1})] = \frac{1}{2}(1-\lambda)\mathbb{E}[\sum_{t=1}^T |\pi_t - \pi_{t-1}|]$. Note that $\mathbb{E}[\sum_{t=1}^T |\pi_t - \pi_{t-1}|]$ is strictly larger for gradual information than for one-shot information. If $\lambda > 1$, the agent strictly prefers one-shot information. If $0 \leq \lambda < 1$, the agent strictly prefers gradual information. If $\lambda = 1$, the agent is indifferent.

Now suppose $v(c_A) = 0$, $v(c_B) = 1$. By the same arguments, total expected news utility is $\mathbb{E}[\sum_{t=1}^T \mu(\rho_t - \rho_{t-1})] = \frac{1}{2}(1-\lambda)\mathbb{E}[\sum_{t=1}^T |\rho_t - \rho_{t-1}|]$. Note that $\mathbb{E}[\sum_{t=1}^T |\rho_t - \rho_{t-1}|]$ is strictly larger for gradual information than for one-shot information. So again, if $\lambda > 1$, the agent strictly prefers one-shot information. If $0 \leq \lambda < 1$, the agent strictly prefers gradual information. If $\lambda = 1$, the agent is indifferent.

(2) Anticipatory utility. If u is linear, then the agent is indifferent between gradual and one-shot information regardless of the sign of $v(c_A) - v(c_B)$. If u is strictly concave, then for $1 \leq t \leq T-1$, $\mathbb{E}[u(\pi_t)] < u(\pi_0)$ and $\mathbb{E}[u(\rho_t)] < u(\rho_0)$ by combining the martingale property and Jensen's inequality. So the agent strictly prefer to keep his prior beliefs until the last period and will therefore choose one-shot information, regardless of the sign of $v(c_A) - v(c_B)$.

(3) Suspense and surprise. Ely, Frankel, and Kamenica (2015) mention a "state-dependent" specification of their surprise and suspense utility functions. With two states, A and B, their specification uses weights $\alpha_A, \alpha_B > 0$ to differentially re-scale belief-based utilities for movements in the two different directions. Specifically, their re-scaled suspense utility is

$$\sum_{t=0}^{T-1} u\left(\mathbb{E}_t\left[\alpha_A \cdot (\pi_{t+1} - \pi_t)^2 + \alpha_B \cdot (\rho_{t+1} - \rho_t)^2\right]\right)$$

and their re-scaled surprise utility is

$$\mathbb{E}\left[\sum_{t=1}^{T} u\left(\alpha_A \cdot (\pi_{t+1} - \pi_t)^2 + \alpha_B \cdot (\rho_{t+1} - \rho_t)^2\right)\right].$$

We may consider agents with opposite preferences over states A and B as agents with different pairs of scaling weights (α_A, α_B). Specifically, say there are $\alpha^{\text{High}} > \alpha^{\text{Low}} > 0$. For an agent preferring A, $\alpha_A = \alpha^{\text{High}}, \alpha_B = \alpha^{\text{Low}}$. For an agent preferring B, $\alpha_A = \alpha^{\text{Low}}, \alpha_B = \alpha^{\text{High}}$. But note that we always have $\pi_{t+1} - \pi_t = -(\rho_{t+1} - \rho_t)$, so along every realized path of beliefs, $(\pi_{t+1} - \pi_t)^2 = (\rho_{t+1} - \rho_t)^2$. This means these two agents with the opposite scaling weights actually have identical objectives and therefore will have the same preference over gradual or one-shot information. □

OA 1.2 Proof of Proposition 3

Proof. We first justify by backwards induction that the value function is indeed given by $U_t^*(x) = (\text{cav}U_t(\cdot \mid x))(x)$, for all $x \in \Delta(\Theta)$ and all $t \leq T-1$, and that it is continuous in x.

If the receiver enters period $t = T-1$ with the belief $x \in \Delta(\Theta)$, the sender faces the following maximization problem.

$$[Q_{T-1}] \qquad \max_{\eta \in \Delta(\Delta(\Theta)), \mathbb{E}[\eta]=x} \int_{\Delta(\Theta)} U_{T-1}(p \mid x) d\eta(p).$$

This is because any sender strategy σ_{T-1} induces a Bayes plausible distribution of posterior beliefs, η with $\mathbb{E}[\eta] = x$, and conversely every such distribution can be generated by some sender strategy, as in Kamenica and Gentzkow (2011). It is well-known that the value of problem Q_{T-1} is $(\text{cav}U_{T-1}(\cdot \mid x))(x)$, justifying $U_{T-1}^*(x)$ as the value function for any $x \in \Delta(\Theta)$. The objective in Q_{T-1} is continuous in p (by assumption on N) and hence in η, and furthermore the constraint set $\{\eta \in \Delta(\Delta(\Theta)) : \mathbb{E}[\eta] = x\}$ is continuous in x. Therefore, $x \mapsto U_{T-1}^*(x)$ is continuous by Berge's Maximum Theorem.

Assume that we have shown that value function is continuous and given by $U_t^*(x)$ for all $t \geq S$. If the receiver enters period $t = S-1$ with belief x, then the sender's value must be:

$$[Q_t] \qquad \max_{\eta \in \Delta(\Delta(\Theta)), \mathbb{E}[\eta]=x} \int_{\Delta(\Theta)} N(p \mid x) + U_{t+1}^*(p) d\eta(p)$$

using the inductive hypothesis that $U_{t+1}^*(p)$ is the period $t+1$ value function. But $N(p \mid x) + U_{t+1}^*(p) = U_t(p \mid x)$ by definition, and it is continuous by the inductive hypothesis. So by the same arguments as in the base case, $U_{S-1}^*(x)$ is the time-$(S-1)$ value function and it is continuous, completing the inductive step.

In the first period, by Carathéodory's theorem, there exist weights $w^1, ..., w^K \geq 0$, beliefs $q^1, ..., q^K \in \Delta(\Theta)$, with $\sum_{k=1}^K w^k = 1$, $\sum_{k=1}^K w^k q^k = x$, such that $U_1^*(\pi_0) = \sum_{k=1}^K w^k U_1(q^k \mid \pi_0)$. Having now shown U_2^* is the period-2 value function, there must exist an optimal information structure where $\sigma_1(\cdot \mid \theta)$ induces beliefs q^k with probability w^k. This information structure induces one of the beliefs $q^1, ..., q^K$ in the second period. Repeating the same procedure for subsequent periods establishes the proposition. □

OA 1.3 Proof of Corollary 1

Proof. We verify Proposition 4's condition $\mu(1-v_0)-\mu(-v_0)+\mu^{'}(0^+)-\mu^{'}(1-v_0)+\mu(-1) > 0$, which is equivalent to $\mu^{'}(0^+) + \mu(1-\pi_0) - \mu(-\pi_0) > -\mu(-1) + \mu^{'}(1-\pi_0)$. We have that

$$\begin{aligned} LHS &= \alpha_p + \alpha_p(1-\pi_0) - \beta_p(1-\pi_0)^2 - [\beta_n \pi_0^2 - \alpha_n \pi_0] \\ RHS &= [-\beta_n + \alpha_n] + [\alpha_p - 2\beta_p(1-\pi_0)] \end{aligned}$$

By algebra, $LHS - RHS = (1-\pi_0)(\alpha_p - \alpha_n) + (1-\pi_0^2)(\beta_p + \beta_n)$. Given that $(\alpha_n - \alpha_p) \leq (\beta_p + \beta_n)$ and $1 - \pi_0^2 > 1 - \pi_0$ for $0 < \pi_0 < 1$,

$$LHS - RHS > -(1-\pi_0^2)(\beta_p + \beta_n) + (1-\pi_0^2)(\beta_p + \beta_n) = 0.$$

□

OA 1.4 Proof of Corollary 2

Proof. This follows from Proposition 4 because $\mu^{'}(0^+) = \infty$ for the power function. □

OA 1.5 Proof of Proposition 5

Proof. Suppose $\Theta = \{\theta_1, ..., \theta_K\}$ and assume without loss the states are associated with consumption levels $c_1 < ... < c_K$.

Let the message space be $M = \{m_1, ..., m_K, m_*\}$. In the first period,

- $\sigma_1(m_k \mid \theta_k) = 1$ for $1 \leq k \leq K-2$,
- $\sigma_1(m_* \mid \theta_{K-1}) = 1$,
- $\sigma_1(m_* \mid \theta_K) = \frac{\pi_0(\theta_{K-1})}{1-\pi_0(\theta_K)}$,
- $\sigma_1(m_K \mid \theta_K) = 1 - \sigma_1(m_* \mid \theta_K)$.

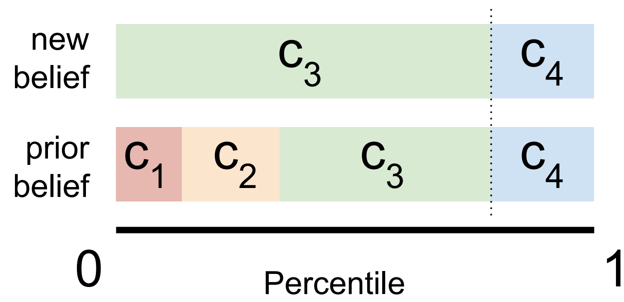

Figure OA.1: New belief about consumption after the muddled message m_* in an environment with 4 states, compared with the old belief given by the prior π_0.

So, message m_k perfectly reveals state θ_k, whereas m_* is a "muddled" message that implies the state is either θ_{K-1} or θ_K. By simple algebra, the probability that the receiver assigns to state θ_K after m_* is the same as the prior belief,

$$\mathbb{P}[\theta_K \mid m_*] = \frac{\pi_0(\theta_K) \cdot \sigma_1(m_* \mid \theta_K)}{\pi_0(\theta_K) \cdot \sigma_1(m_* \mid \theta_K) + \pi_0(\theta_{K-1}) \cdot 1} = \pi_0(\theta_K).$$

In the second period, the information structure perfectly reveals the true state regardless of the last message, $\sigma_2(m_k \mid \theta_k) = 1$ for all $1 \leq k \leq K$.

To compute the news utility of the muddled message m_*, note that at percentiles $p \in [0, \pi_0(\theta_1))$, the change in p-percentile consumption utility is $v(c_{K-1}) - v(c_1)$. Similarly, for $2 \leq k \leq K-2$, the change in consumption utility at percentile $p \in \left[\sum_{j=1}^{k-1} \pi_0(\theta_j), \pi_0(\theta_k) + \sum_{j=1}^{k-1} \pi_0(\theta_j)\right)$ is $v(c_{K-1}) - v(c_k)$. There are no changes at percentiles above $\sum_{j=1}^{K-2} \pi_0(\theta_j)$.

If $\theta = \theta_{K-1}$, total news utility from receiving m_* then m_{K-1} is

$$\underbrace{\left[\sum_{k=1}^{K-2} \pi_0(\theta_k) \cdot \mu(v(c_{K-1}) - v(c_k))\right]}_{\text{from } m_* \text{ in period 1}} + \underbrace{\pi_0(\theta_K) \cdot \mu(v(c_{K-1}) - v(c_K))}_{\text{from } m_{K-1} \text{ in period 2}}.$$

This is identical to the news utility from one-shot resolution in state θ_{K-1}. Similarly, the information structure just constructed gives the same news utility as one-shot resolution when the state is θ_k for $1 \leq k \leq K-2$, and when the state is θ_K and the receiver gets m_K in period 1.

When the receiver sees m_* in period 1 and m_K in period 2 in state θ_K, an event that happens with strictly positive probability since $\pi_0(\theta_{K-1}) < 1 - \pi_0(\theta_K)$ as $K \geq 3$, he gets strictly more news utility than from one-shot resolution.

If $\theta = \theta_K$, total news utility from receiving m_* then m_K is

$$\underbrace{\left[\sum_{k=1}^{K-2} \pi_0(\theta_k) \cdot \mu(v(c_{K-1}) - v(c_k))\right]}_{\text{from } m_* \text{ in period 1}} + \underbrace{\left[\sum_{k=1}^{K-1} \pi_0(\theta_k) \cdot \mu(v(c_K) - v(c_{K-1}))\right]}_{\text{from } m_K \text{ in period 2}},$$

while one-shot resolution gives $\sum_{k=1}^{K-1} \pi_0(\theta_k) \cdot \mu(v(c_K) - v(c_k))$. For each $1 \leq k \leq K-2$ (non-empty since $K \geq 3$),

$$\mu(v(c_K) - v(c_{K-1})) + \mu(v(c_{K-1}) - v(c_k)) > \mu(v(c_K) - v(c_k))$$

by sub-additivity in gains. This shows the constructed information structure gives strictly more news utility. □

OA 1.6 Proof of Corollary 3

We first state a sufficient condition for the sub-optimality of information structures with partial bad news with $T = 2$. Consider the chord connecting $(0, U_1(0 \mid \pi_0))$ and $(\pi_0, U_1(\pi_0 \mid \pi_0))$ and let $\ell(x)$ be its height at $x \in [0, \pi_0]$. Let $D(x) := \ell(x) - U_1(x \mid \pi_0)$.

Lemma OA.1. *For this chord to lie strictly above $U_1(p \mid \pi_0)$ for all $p \in (0, \pi_0)$, it suffices that $D^{'}(0) > 0$, $D^{'}(\pi_0) < 0$, and $D^{''}(p) = 0$ for at most one $p \in (0, \pi_0)$.*

Now we verify that the condition in Lemma OA.1 holds for the quadratic news utility, which in turn verifies the condition of Proposition 7 for $q = \pi_0$ and shows partial bad news information structures to be strictly suboptimal.

Proof. Clearly, $D(p)$ is a third-order polynomial, so $D^{''}(p)$ has at most one root.

For $p < \pi_0$, we have the derivative

$$\begin{aligned}\frac{d}{dp}U(p \mid \pi_0) =& 2\beta_n(p - \pi_0) + \alpha_n + \alpha_p(1-p) - \beta_p(1-p)^2 \\ &+ p(-\alpha_p + 2\beta_p(1-p)) - (\beta_n p^2 - \alpha_n p) + (1-p)(2\beta_n p - \alpha_n)\end{aligned}$$

The slope of the chord between 0 and π_0 is: $\alpha_p - \beta_p + (2\beta_p - \alpha_p + \alpha_n)\pi_0 - (\beta_p + \beta_n)\pi_0^2$. So, after straightforward algebra, $D^{'}(0) = (2(\beta_p + \beta_n) - (\alpha_p - \alpha_n))\pi_0 - (\beta_p + \beta_n)\pi_0^2$. Applying weak loss aversion with $z = 1$, $\alpha_p - \alpha_n \leq \beta_p - \beta_n$. This shows

$$\begin{aligned}D^{'}(0) &\geq (2(\beta_p + \beta_n) - (\beta_p - \beta_n))\pi_0 - (\beta_p + \beta_n)\pi_0^2 \\ &= (\beta_p + \beta_n)\pi_0(1 - \pi_0) + 2\beta_n\pi_0 > 0\end{aligned}$$

for $0 < \pi_0 < 1$.

We also derive $D^{'}(\pi_0) = (\alpha_p - 2\beta_p - 2\beta_n - \alpha_n)\pi_0 + (2\beta_p + 2\beta_n)\pi_0^2$. Note that this is a convex parabola in π_0, with a root at 0. Also, the parabola evaluated at 1 is equal to $\alpha_p - \alpha_n \leq 0$, where the inequality comes from the weak loss aversion with $z = 0$. This implies $D^{'}(\pi_0) < 0$ for $0 < \pi_0 < 1$. □

OA 1.7 Proof of Lemma OA.1

Proof. We need $D > 0$ in the region $(0, \pi_0)$. We know that $D(0) = D(\pi_0) = 0$. Given the conditions in the statement and the twice-differentiability of D in $(0, \pi_0)$ it follows that D'' changes sign only once. Moreover, it also follows that $D > 0$ in a right-neighborhood of $x = 0$ and a left-neighborhood of $x = \pi_0$. Suppose D has an interior minimum at $x_0 \in (0, \pi_0)$. Then it holds $D''(x_0) \geq 0$.

Suppose $D''(x) > 0$ for all small x. Then it follows $x_0 \leq p$, where we set $p = \pi_0$ if p doesn't exist. Because $D''(x) \geq 0$ for all $x \leq p$ we have that $D'(x) > 0$ for all $x \leq p$. In particular also $D(x) > 0$ for all such x due to the Fundamental Theorem of Calculus. Thus, the interior minimum is positive and so the claim about D in $(0, \pi)$ is proven in this case.

Suppose instead that $D''(x) < 0$ for all x near enough to 0. Then it follows that $x_0 \geq p$. In particular, for all $x > p$ we have $D''(x) > 0$. Since the derivative is strictly increasing for all $x \in (x_0, \pi_0)$ and $D'(\pi_0) < 0$ we have that $D'(x) < 0$ for all $x \in (x_0, \pi_0)$. In particular, from the Fundamental Theorem of Calculus, $D(\pi_0)$ is strictly below $D(x_0)$. Since $D(\pi_0) = 0$ we have again that $D(x_0) > 0$.

Given the boundary values of D and the signs of the derivatives at $0, \pi_0$ and that any interior minimum of D is strictly positive, we have covered all cases and so shown that $D > 0$ in $(0, \pi_0)$. □

OA 1.8 Proof of Lemma 1

Proof. **Part 1.** Fix a prior π_0 and a pair $(\bar{M}, \bar{\sigma})$ which induces an equilibrium as in Definition 3. We focus on the case that $|\bar{M}| > 2$ as the other cases are trivial.

Let $M = \{g, b\}$ and we will inductively define the sender's strategy σ_t on t so that (M, σ) is another equilibrium which delivers the same expected utility as $(\bar{M}, \bar{\sigma})$. In doing so we will successively define a sequence of subsets of histories, $H^t_{int} \subseteq M^t$ and $\bar{H}^t_{int} \subseteq \bar{M}^t$, which are length t histories associated with interior equilibrium beliefs about the state in the new and old equilibria, as well as a map ϕ that associates new histories to old ones.

Let $H^0_{int} = \bar{H}^0_{int} := \{\varnothing\}$, $\phi(\varnothing) = \varnothing$.

Once we have defined σ_{t-1}, $H_{int}^{t-1}, \bar{H}_{int}^{t-1}$ and $\phi : H_{int}^{t-1} \to \bar{H}_{int}^{t-1}$, we then define σ_t. If $h^{t-1} \notin H_{int}^{t-1}$, then simply let $\sigma_t(h^{t-1}, \theta)(g) = 0.5$ for both $\theta \in \{G, B\}$. For each $h^{t-1} \in H_{int}^{t-1}$, by the definition of $\bar{H}_{int}^{t-1}$, the equilibrium belief π_{t-1} associated with $\phi(h^{t-1})$ in the old equilibrium satisfies $0 < \pi_{t-1} < 1$. Let $\Phi_G(h^{t-1})$ and $\Phi_B(h^{t-1})$ represent the sets of posterior beliefs that the sender induces with positive probability in the good and bad states following public history $\phi(h^{t-1}) \in \bar{H}_{int}^{t-1}$ in $(\bar{M}, \bar{\sigma})$.

We must have $\Phi_G(h^{t-1}) \backslash \Phi_B(h^{t-1}) \subseteq \{1\}$ and $\Phi_B(h^{t-1}) \backslash \Phi_G(h^{t-1}) \subseteq \{0\}$, since any message unique to either state is conclusive news of the state. We construct $\sigma_t(h^{t-1}, \theta)$ based on the following four cases.

Case 1: $1 \in \Phi_G(h^{t-1})$ and $0 \in \Phi_B(h^{t-1})$. Let $\sigma_t(h^{t-1}, G)$ assign probability 1 to g and let $\sigma_t(h^{t-1}, B)$ assign probability 1 to b.

Case 2: $1 \in \Phi_G(h^{t-1})$ but $0 \notin \Phi_B(h^{t-1})$. By Bayesian plausibility, there exists some smallest $q^* \in (0, \pi_{t-1})$ with $q^* \in \Phi_G(h^{t-1}) \cap \Phi_B(h^{t-1})$, induced by some message $\bar{m}_b \in \bar{M}$ sent with positive probabilities in both states. Also, some message $\bar{m}_g \in \bar{M}$ sent with positive probability in state G induces belief 1. Let $\sigma_t(h^{t-1}, B)(b) = 1$ and let $\sigma_t(\varnothing, G)(b) = x$ where $x \in (0, 1)$ solves $\frac{\pi_{t-1}x}{\pi_{t-1}x + (1-\pi_{t-1})} = q^*$.

Case 3: $1 \notin \Phi_G(h^{t-1})$ but $0 \in \Phi_B(h^{t-1})$. By Bayesian plausibility, there exists some largest $q^* \in (\pi_{t-1}, 1)$ with $q^* \in \Phi_G(h^{t-1}) \cap \Phi_B(h^{t-1})$. Let $\sigma_t(h^{t-1}, G)(g) = 1$ and let $\sigma_t(h^{t-1}, B)(g) = x$ where $x \in (0, 1)$ solves $\frac{\pi_{t-1}}{\pi_{t-1} + (1-\pi_{t-1})x} = q^*$.

Case 4: $1 \notin \Phi_G(h^{t-1})$ and $0 \notin \Phi_B(h^{t-1})$. By Bayesian plausibility, $\Phi_G(h^{t-1}) = \Phi_B(h^{t-1})$, and there exist some largest $q_L \leq \pi_{t-1}$ and smallest $q_H \geq \pi_{t-1}$ in this common set of posterior beliefs, and further there exist $x, y \in (0, 1)$ so that $\frac{\pi_{t-1}x}{\pi_{t-1}x + (1-\pi_{t-1})y} = q_H$ and $\frac{\pi_{t-1}(1-x)}{\pi_{t-1}(1-x) + (1-\pi_{t-1})(1-y)} = q_L$. Let $\sigma(h^{t-1}, G)(g) = x$ and $\sigma(h^{t-1}, B)(g) = y$.

Having constructed σ_t, let H_{int}^t be those on-path period t histories with interior equilibrium beliefs, that is $h^t = (h^{t-1}, m) \in H_{int}^t$ if and only if $h^{t-1} \in H_{int}^{t-1}$ and $\sigma(h^{t-1}, \theta)(m) > 0$ for both $\theta \in \{G, B\}$. A property of the construction of σ_t is that if $h^{t-1} \in H_{int}^{t-1}$, then both (h^{t-1}, g) and (h^{t-1}, b) are on-path. That is, off-path histories can only be continuations of histories with degenerate beliefs in $\{0, 1\}$.

Let $\bar{H}_{int}^t$ be on-path period t histories with interior equilibrium beliefs in $(\bar{M}, \bar{\sigma})$. By the definition of σ_t, there exists $\bar{m} \in \bar{M}$ so that h^t induces the same equilibrium belief in the new equilibrium as the history $(\phi(h^{t-1}), \bar{m}) \in \bar{H}_{int}^t$ in the old equilibrium, and we define $\phi(h^t) := (\phi(h^{t-1}), \bar{m})$.

The receiver's expected payoff in both the B and G states are the same as in the old equilibrium. To see this, note that by our construction, the receiver's expected payoff in state B is the same as if we took a deterministic selection of messages $m_1, m_2, ...$ in the old equilibrium with the property that $\sigma_1(\varnothing, B)(m_1) > 0$ and, for $t \geq 2$, $\sigma_t(m_1, ..., m_{t-1}, \theta)(m_t) >$

0. Then, we had the sender play message m_t in period t. Since this sequence of messages is played with positive probability in state B of the old equilibrium, it must yield the expected payoff under B — if it yields higher or lower payoffs, then we can construct a deviation that improves the receiver's ex-ante expected payoffs in the old equilibrium. A similar argument holds for state G.

It remains to check that (M, σ) is an equilibrium by ruling out one-shot deviations. We argued before that all off-path histories must follow an on-path history with equilibrium belief in 0 or 1. There are no profitable deviations at off-path histories or at on-path histories with degenerate beliefs, because the receiver does not update beliefs after such histories regardless of the sender's play.

So consider an on-path history with a non-degenerate belief, i.e. a member $h^t \in H^t_{int}$. A one-shot deviation following h^t corresponds to a deviation following $\phi(h^t)$ in $(\bar{M}, \bar{\sigma})$, and must not be strictly profitable.

Part 2. We now turn to the second claim. If $T \leq T'$, then for any equilibrium with horizon T, we may construct an equilibrium of horizon T' which sends messages in the same way in periods $1, ..., T-1$, but babbles starting in period T. This equilibrium has the same expected payoff as the old one. □

Note that the first claim of Lemma 1 also holds for the infinite horizon model of Online Appendix OA 2.3. Nothing in the argument relies on T being finite. This is because the proof argument relies on the one-shot deviation property which holds for equilibria in both finite and infinite horizon models. Thus, in particular, in the proof of Proposition OA.2 we can also focus on a binary signal space.

OA 1.9 Proof of Lemma 2

Proof. Due to sub-additivity,

$$\mu(p) < \mu(p - \pi) + \mu(\pi). \tag{2}$$

Note that symmetry implies $\mu(-p) = -\mu(p)$ and that $\mu(-\pi) = -\mu(\pi)$. Rearranged (2) is precisely $N(0; \pi) < N(p; \pi)$. □

OA 1.10 Proof of Lemma A.1

Proof. We have $\frac{\partial N_G(p;\pi)}{\partial p} = \mu'(p-\pi) - \mu'(1-p)$. For $0 \leq p < \pi$ and under greater sensitivity to losses, $\mu'(p-\pi) \geq \mu'(\pi - p)$. Since $\mu''(x) < 0$ for $x > 0$, $\mu'(\pi - p) > \mu'(1-p)$. This shows $\frac{\partial N_G(p;\pi)}{\partial p} > 0$ for $p \in [0, \pi)$.

The symmetry results follow from simple algebra and do not require any assumptions.

Note that $\frac{\partial^2 N_G(p;\pi)}{\partial p^2} = \mu''(p-\pi) + \mu''(1-p) < 0$ for any $p \in [\pi, 1]$, due to diminishing sensitivity. Combined with the required symmetry, this means $\frac{\partial N_G(p;\pi)}{\partial p}$ crosses 0 at most once on $[\pi, 1]$, so for each $p_1 \in [\pi, 1]$, we can find at most one p_2 so that $N_G(p_1; \pi) = N_G(p_2; \pi)$. In particular, this implies at every intermediate $p_1 \in (\pi, 1)$, we get $N_G(p_1; \pi) > N_G(\pi; \pi)$ since we already have $N_G(1; \pi) = N_G(\pi; \pi)$. This shows $N_G(\cdot; \pi)$ is strictly larger on $[\pi, 1]$ than on $[0, \pi)$.

A similar argument, using $\mu''(x) > 0$ for $x < 0$, establishes that for each $p_1 \in [0, \pi]$, we can find at most one p_2 so that $N_B(p_1; \pi) = N_B(p_2; \pi)$. □

OA 1.11 Proof of Lemma A.2

Proof. Suppose $|P_G| = 1$.

If $P_G = \{\pi\}$, then any equilibrium message not inducing π must induce 0. By the Bayes' rule, the sender cannot induce belief 0 with positive probability in the bad state, so $P_B = \{\pi\}$ as well.

If $P_G = \{1\}$, then any equilibrium message not inducing 1 must induce 0. Furthermore, the sender cannot send equilibrium messages inducing belief 1 with positive probability in the bad state, else the equilibrium belief associated with these messages should be strictly less than 1. Thus $P_B = \{0\}$.

If $P_G = \{p_1\}$ for some $0 \leq p_1 < \pi$, then any equilibrium message not inducing p_1 must induce 0. This is a contradiction since the posterior beliefs do not average out to π.

This leaves the case of $P_G = \{p_1\}$ for some $\pi < p_1 < 1$. Any equilibrium message not inducing p_1 must induce 0. Furthermore, the sender must induce the belief p_1 in the bad state with positive probability, else we would have $p_1 = 1$. At the same time, the sender must also induce belief 0 with positive probability in the bad state, else we violate Bayes' rule. So $P_B = \{0, p_1\}$.

Now suppose $|P_G| = 2$.

In the good state, the sender must be indifferent between two beliefs p_1, p_2 both induced with positive probability. By Lemma A.1, $N_G(p; \pi)$ is strictly increasing on $[0, \pi]$ and strictly higher on $[\pi, 1]$ than on $[0, \pi)$, while for each $p_1 \in [\pi, 1]$, there exists exactly one point $p_2 \in [\pi, 1]$ so that $N_G(p_1; \pi) = N_G(p_2; \pi)$. This means we must have $p_1 \in [\pi, \frac{1+\pi}{2}]$, $p_2 = 1 - p_1 + \pi$.

If $P_G = \{\pi, 1\}$, any equilibrium message not inducing π or 1 must induce 0. Also, $1 \notin P_B$, because any message sent with positive probability in the bad state cannot induce belief 1. We cannot have $P_B = \{0\}$, because then the message inducing belief π actually induces 1. We cannot have $P_B = \{\pi\}$ for then we violate Bayes' rule. This leaves only $P_B = \{0, \pi\}$.

If $P_G = \{p_1, p_2\}$ for some $p_1 \in (\pi, \frac{1+\pi}{2})$, then any equilibrium message not inducing p_1

or p_2 must induce 0. Also, $p_1, p_2 \in P_B$, else messages inducing these beliefs give conclusive evidence of the good state. By Bayes' rule, we must have $P_B = \{0, p_1, p_2\}$.

It is impossible that $|P_G| \geq 3$, since, by Lemma A.1, $N_G(p;\pi)$ is strictly increasing on $[0,\pi]$ and strictly higher on $[\pi, 1]$ than on $[0,\pi)$, while for each $p_1 \in [\pi, 1]$, there exists exactly one point $p_2 \in [\pi, 1]$ so that $N_G(p_1;\pi) = N_G(p_2;\pi)$. So the sender cannot be indifferent between 3 or more different posterior beliefs of the receiver in the good state. □

OA 1.12 Proof of Corollary 4

Proof. First, μ exhibits greater sensitivity to losses, because $\mu(-x) = -\lambda\mu(x)$ for all $x > 0$ and we have $\lambda \geq 1$.

To apply Proposition 10, we only need to verify that $\min_{z\in[0,1-\pi_0]} \frac{\mu'(z)}{\mu'(-(\pi_0+z))} > 1$. For the λ-scaled μ, $\min_{z\in[0,1-\pi_0]} \frac{\mu'(z)}{\mu'(-(\pi_0+z))} = \frac{1}{\lambda} \cdot \min_{z\in[0,1-\pi_0]} \frac{\tilde{\mu}'_{pos}(z)}{\tilde{\mu}'_{pos}(\pi_0+z)}$. The assumption that $\min_{z\in[0,1-\pi_0]} \frac{\tilde{\mu}'_{pos}(z)}{\tilde{\mu}'_{pos}(\pi_0+z)} > \lambda$ gives the desired conclusion. □

OA 1.13 Proof of Corollary 5

Proof. We apply Proposition 12 to the case of quadratic news utility. Recall the relevant indifference equation in the good state.

$$\mu(-q_t) = \mu(q_{t+1} - q_t) + \mu(-q_{t+1}). \tag{3}$$

Plugging in the quadratic specification and algebraic transformations lead to

$$0 = (\alpha_p - \alpha_n)(q_{t+1} - q_t) - \beta_p(q_{t+1} - q_t) + \beta_n(q_{t+1} - q_t)(q_{t+1} + q_t)$$

Define $r = q_{t+1} - q_t$. Then this relation can be written as

$$(\beta_p - \beta_n)r^2 + (\alpha_n - \alpha_p - 2\beta_n q_t)r = 0,$$

i.e. r is a zero of a second order polynomial. For P^* to be non-empty we need this root r to be in $(0, 1-q_t)$. In particular the peak/trough $\bar{r}$ of the parabola defined by the second order polynomial should satisfy $\bar{r} \in (0, \frac{1-q_t}{2})$. Given that $\bar{r} = \frac{2\beta_n q_t - (\alpha_n - \alpha_p)}{2(\beta_p - \beta_n)}$ for the case that $\beta_p \neq \beta_n$, we get the equivalent condition on the primitives $0 < \frac{2\beta_n q_t - (\alpha_n - \alpha_p)}{2(\beta_p - \beta_n)} < \frac{1-q_t}{2}$. The root r itself is given by $r = \frac{2\beta_n q_t - (\alpha_n - \alpha_p)}{\beta_p - \beta_n}$, which leads to the recursion

$$q_{t+1} = q_t \frac{\beta_p + \beta_n}{\beta_p - \beta_n} - \frac{\alpha_n - \alpha_p}{\beta_p - \beta_n}. \tag{4}$$

This leads to the formula for $P^*(\pi)$ in part 1).

Case 1: When $\beta_p < \beta_n$ the coefficient in front of q_t is negative so that the recursion in Equation (4) leads to

$$q_{t+1} - q_t = q_t \frac{2\beta_n}{\beta_p - \beta_n} - \frac{\alpha_n - \alpha_p}{\beta_p - \beta_n} < 0.$$

This also shows that for the case that $\beta_p < \beta_n$, a GGN equilibrium with 1 or more intermediate beliefs only exists when the prior is low enough: namely $\pi_0 < \frac{\alpha_n - \alpha_p}{2\beta_n} =: q^*$.

Case 2: When $\beta_p > \beta_n$ the slope in Equation (4) is above 1 so that for all priors π_0 large enough we get an increasing sequence q_t which satisfies Equation (3). It is also easy to see from Equation (4) that

$$(q_{t+2} - q_{t+1}) - (q_{t+1} - q_t) = \left(\frac{\beta_p + \beta_n}{\beta_p - \beta_n} - 1\right) > 0,$$

proving the statement in the text after the corollary.

That an equilibrium can exist where partial good news are released for more than two periods, is shown by the example in the main text following the statement of the Corollary (see Figure 5). □

OA 1.14 Proof of Corollary 6

Proof. We verify the sufficient condition in Proposition 13. We get $\frac{\partial}{\partial p} N_B(p;\pi) = \frac{\alpha}{(p-\pi)^{1-\alpha}} - \frac{\lambda\alpha}{p^{1-\alpha}}$, so $\frac{\partial}{\partial p} N_B(p;\pi)|_{p=\pi} = \infty$.

To show that $|P^*(\pi)| \leq 1$, it suffices to show that $\frac{\partial}{\partial p} N_B(p;\pi) = 0$ for at most one $p > \pi$. For the derivative to be zero, we need $(\frac{p}{p-\pi})^{1-\alpha} = \lambda$. As the LHS is decreasing for $p > \pi$, it can have at most one solution. □

OA 2 A Random-Horizon Model

In this section, we study a version of our information design problem without a deterministic horizon. Each period, with probability $1-\delta \in (0,1]$, the true state of the world is exogenously revealed to the receiver and the game ends. Until then, the informed sender communicates with the receiver each period as in the model from Section 2. We verify that our results from the finite-horizon setting extend analogously into this random-horizon environment.

OA 2.1 The Environment

Consider an environment where the consumption event takes place far in the future, but the sender is no longer the receiver's only source of information in the interim. Instead, a third party perfectly discloses the state to the receiver with some probability each period. For instance, the sender may be the chair of a central bank who has decided on the bank's monetary policy for next year and wishes to communicate this information over time, while the third party is an employee of the bank who also knows the planned policy. With some probability each period, the employee goes to the press and leaks the future policy decision.

Time is discrete with $t = 0, 1, 2, \ldots$ The sender commits to an information structure (M, σ) at time 0. The information structure consists of a finite message space M and a sequence of message strategies $(\sigma_t)_{t=1}^{\infty}$ where each $\sigma_t(\cdot \mid h^{t-1}, \theta) \in \Delta(M)$ specifies how the sender will mix over messages in period t as a function of the public history h^{t-1} so far and the true state θ.

The sender learns the state at the beginning of period 1 and sends a message according to σ_1. At the start of each period $t = 2, 3, 4, \ldots$, there is probability $(1 - \delta) \in (0, 1]$ that the receiver exogenously and perfectly learns the state θ. If so, the game effectively ends because no further communication from the sender can change the receiver's belief. If not, then the sender sends the next message according to σ_t. The randomization over exogenous learning is i.i.d. across periods, so the time of state revelation (i.e., the horizon of the game) is a geometric random variable.

OA 2.2 The Value Function with Commitment

Let $V_\delta : [0, 1] \to \mathbb{R}$ be the value function of the problem with continuation probability δ — that is, $V_\delta(p)$ is the highest possible total expected news utility up to the period of state revelation, when the receiver holds belief p in the current period and state revelation does not happen this period. The value function satisfies the recursion $V_\delta(p) = \tilde{V}_\delta(p \mid p)$, where

$$\tilde{V}_\delta(\cdot \mid p) := \text{cav}_q[\mu(q - p) + \delta V_\delta(q) + (1 - \delta)(q \cdot \mu(1 - q) + (1 - q) \cdot \mu(-q))].$$

Ely (2017) studies an infinite-horizon information design problem whose value function also involves concavification. Unlike in Ely (2017), the current belief enters the objective function for our news-utility problem.

Our first result shows this recursion has a unique solution which increases in δ for any fixed $p \in [0, 1]$.

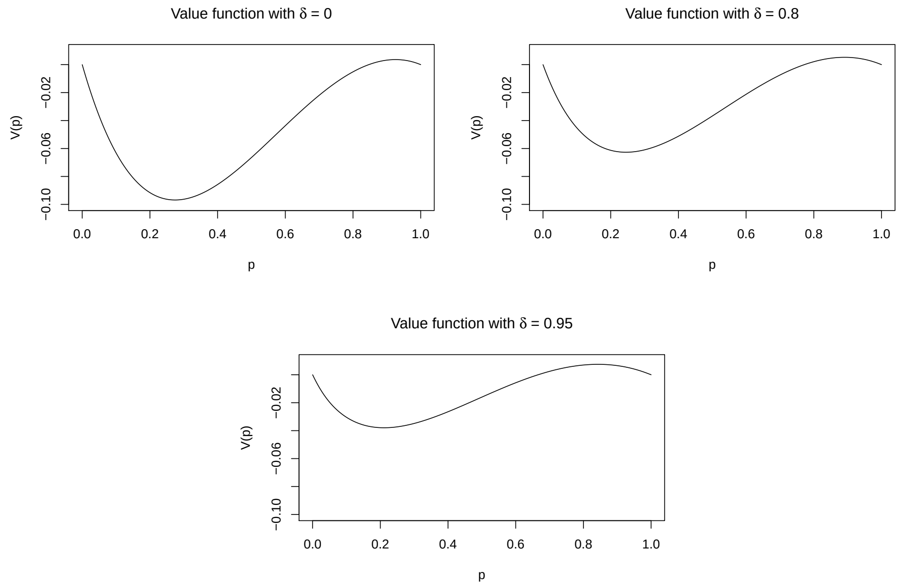

Figure OA.2: The value function for $\delta = 0, 0.8, 0.95$. Consistent with Proposition OA.1, the value function is pointwise higher for higher δ.

Proposition OA.1. *For every $\delta \in [0, 1)$, the value function V_δ exists and is unique. Furthermore, $V_\delta(p)$ is increasing in δ for every $p \in [0, 1]$.*

Figure OA.2 illustrates this result by plotting $V_\delta(p)$ for the quadratic news utility with $\alpha_p = 2$, $\alpha_n = 2.1$, $\beta_p = 1$, and $\beta_n = 0.2$ for three different values of δ : 0, 0.8, and 0.95. (In fact, the monotonicity of the value function in δ also holds when there are more than two states.)

The monotonicity of V_δ in δ says that when the sender is benevolent and has commitment power, third-party leaks are harmful for the receiver's expected welfare. This result can be explained intuitively as follows. Just as with increasing T in the finite-horizon model, increasing δ expands the set of implementable belief paths. The idea behind implementing a payoff from a shorter horizon / lower δ is that the sender switches to babbling forever after certain histories. This switching happens at a deterministic calendar time in the finite-horizon setting but at a random time in the random-horizon setup, mimicking the random arrival of the state revelation period.

OA 2.3 Gradual Good News Equilibria Without Commitment

Now we turn to equilibria of the random-horizon cheap talk game when the sender lacks commitment power. Analogously to the case of finite horizon, a *strict gradual good news* equilibrium (strict GGN) features a deterministic sequence of increasing posteriors $q^{(0)} <$

$q^{(1)} < \ldots$ such that $q^{(0)} = \pi_0$ is the receiver's prior before the game starts and $q^{(t)}$ is his belief in period t, provided state revelation has not occurred. An analog of Proposition 12 continues to hold.

Proposition OA.2. *Let $P^*(\pi) \subseteq (\pi, 1]$ be those beliefs p satisfying $N_B(p;\pi) = N_B(0;\pi)$. Suppose μ exhibits diminishing sensitivity and loss aversion. There exists a gradual good news equilibrium with a (possibly infinite) sequence of intermediate beliefs $q^{(1)} < q^{(2)} < ...$ if and only if $q^{(j)} \in P^*(q^{(j-1)})$ for every $j = 1, 2, ...,$ where $q^{(0)} := \pi_0$.*

The P^* set is the same in the finite- and random-horizon environments. Corollary 6 then implies that even in the random-horizon environment where the game could continue for arbitrarily many periods, intermediate beliefs grow at an increasing rate in GGN equilibria for quadratic and square-roots μ, and there exists a finite bound on the number of periods of informative communication that applies for all $\delta \in [0, 1)$.

OA 2.4 Proofs

OA 2.4.1 Proof of Proposition OA.1

Proof. Consider the following operator ϕ on the space of continuous functions on $[0, 1]$. For $V : [0, 1] \to \mathbb{R}$, define $\phi(V)(p) := \tilde{V}(p \mid p)$, where

$$\tilde{V}(\cdot \mid p) := \text{cav}_q[\mu(q - p) + \delta V(q) + (1 - \delta)(q \cdot \mu(1 - q) + (1 - q) \cdot \mu(-q))].$$

We show that ϕ satisfies the Blackwell conditions and so is a contraction mapping.

Suppose that $V_2 \geq V_1$ pointwise. Then for any $p, q \in [0, 1]$,

$$\mu(q-p)+\delta V_2(q)+(1-\delta)(q\mu(1-q)+(1-q)\mu(-q)) \geq (q-p)+\delta V_1(q)+((1-\delta)(q\mu(1-q)+(1-q)\mu(-q))$$

therefore $\tilde{V}_2(\cdot \mid p) \geq \tilde{V}_1(\cdot \mid p)$ pointwise as well. In particular, $\tilde{V}_2(p \mid p) \geq \tilde{V}_1(p \mid p)$, that is $\phi(V_2)(p) \geq \phi(V_1)(p)$.

Also, let $k > 0$ be given and let $V_2 = V_1 + k$ pointwise. It is easy to see that $\tilde{V}_2(\cdot \mid p) = \tilde{V}_1(\cdot \mid p) + \delta k$ for every p, because the argument to the concavification operator will be pointwise higher by δk. So in particular, $\phi(V_2)(p) = \phi(V_1)(p) + \delta k$. By the Blackwell conditions, the operator ϕ is a contraction mapping on the metric space of continuous functions on $[0, 1]$ with the supremum norm. Thus, the value function exists and is also unique.

To show pointwise monotonicity in δ, suppose $0 \leq \delta < \delta' < 1$. First, $V_\delta(0) = V_\delta(1) = 0$ for any $\delta \in [0, 1)$. Now consider an environment where full revelation happens at the end of each period with probability $1-\delta$, and fix a prior $p \in (0, 1)$. There exists some binary information structure with message space $M = \{0, 1\}$, public histories $H^t = (M)^t$ for $t = 0, 1, ...,$ and sender strategies $(\sigma_t)_{t=0}^{\infty}$ with $\sigma_t : H^t \times \Theta \to \Delta(M)$, such that (M, σ) induces expected news utility of $V_\delta(p)$ when starting at prior p.

We now construct a new information structure, $(\bar{M}, \bar{\sigma})$ to achieve expected news utility $V_\delta(p)$ when starting at prior p in an environment where full revelation happens at the end of each period with probability $1-\delta^{'}$, with $\delta^{'} > \delta$. Let $\bar{M} = \{0, 1, \varnothing\}$. The idea is that when full revelation has not happened, there is a $1 - \frac{\delta}{\delta'}$ probability each period that the sender enters into a babbling regime forever. When the sender enters the babbling regime at the start of period $t + 1$, the receiver's expected utility going forward is the same as if full revelation happened at the start of $t + 1$.

To implement this idea, after any history $h^t \in H^t$ not containing $\varnothing$, let

$$\bar{\sigma}_{t+1}(h^t; \theta) = \begin{cases} \varnothing & \text{w/p } 1 - \frac{\delta}{\delta'} \\ 1 & \text{w/p } \frac{\delta}{\delta'} \cdot \sigma_{t+1}(h^t; \theta)(1) \\ 0 & \text{w/p } \frac{\delta}{\delta'} \cdot \sigma_{t+1}(h^t; \theta)0) \end{cases} .$$

That is, conditional on not entering the babbling regime, $\bar{\sigma}$ behaves in the same way as σ. But, after any history $h^t \in H^t$ containing at least one $\varnothing$, $\bar{\sigma}_{t+1}(h^t; \theta) = \varnothing$ with probability 1. Once the sender enters the babbling regime, she babbles forever (until full revelation exogenously arrives at some random date). We need to verify that payoff from this strategy is indeed $V_\delta(p)$. Fix a history h^t not containing $\varnothing$ and a state θ, and suppose $p^*(h^t) = q$. Under $\bar{\sigma}_{t+1}$, with probability of $(1 - \delta') + \delta'(1 - \frac{\delta}{\delta'}) = 1 - \delta$ the receiver gets the expected babbling payoff $q\mu(1 - q) + (1 - q)\mu(-q)$ in the period of state revelation. Analogously, under σ_{t+1}, there is probability $1 - \delta$ that state revelation happens in period $t + 1$ and the receiver gets $q\mu(1 - q) + (1 - q)\mu(-q)$ in expectation. With probability $\delta' \frac{\delta}{\delta'} = \delta$, the receiver facing $\bar{\sigma}$ gets the payoff induced by $\sigma_{t+1}(h^t; \theta)$ in period $t + 1$ and the same distribution of continuation histories as under σ. The same argument applies to all these continuation histories, so $\bar{\sigma}$ must induce the same expected payoff as σ when starting at $(h^t; \theta)$. □

OA 2.4.2 Proof of Proposition OA.2

Proof. We show first sufficiency. Consider the following strategy profile. In period t where the public history so far h^{t-1} does not contain any b, let $\sigma(h^{t-1}; G)(g) = 1$, $\sigma(h^{t-1}; B)(g) = x$ where $x \in (0, 1)$ satisfies $\frac{p_{t-1}}{p_{t-1}+(1-p_{t-1})x} = p_t$. But if public history contains at least one b, then $\sigma(h^{t-1}; G)(b) = 1$ and $\sigma(h^{t-1}; B)(b) = 1$. In terms of beliefs, suppose h^t is so that every message so far has been g. Such histories are on-path and get assigned the Bayesian posterior belief. If h^t contains at least one b, then belief is 0. It is easy to verify that these beliefs are derived from Bayes' rule whenever possible.

We verify that the sender has no incentive to deviate. Consider period t with history h^{t-1} that does not contain any b. The receiver's current belief is p_{t-1} by construction.

In state B, we first calculate the sender's equilibrium payoff after sending g. For any realization of the exogenous revelation date, the receiver's total news utility in the good state along the equilibrium path is given by $\sum_{j=1}^{J} \mu(p_{t-1+j} - p_{t-2+j}) + \mu(-p_{t-1+J})$ for some integer $J \geq 1$. Since $p_{t-1+J} \in P^*(p_{t-2+J})$, we have $N_B(p_{t-1+J}; p_{t-2+J}) = N_B(0; p_{t-2+J})$, that is to say $\mu(p_{t-1+J} - p_{t-2+J}) + \mu(-p_{t-1+J}) = \mu(-p_{t-2+J})$. We may therefore rewrite the receiver's total news utility as $\sum_{j=1}^{J-1} \mu(p_{t-1+j} - p_{t-2+j}) + \mu(-p_{t-2+J})$. But by repeating this argument, we conclude that the receiver's total news utility is just $\mu(-p_{t-1})$. Since this result holds regardless of J, the sender's expected total utility from sending g today is $\mu(-p_{t-1})$, which is the same as the news utility from sending b today. Thus, sender is indifferent between g and b and has no profitable deviation.

In state G, the sender gets at least $\mu(1 - p_{t-1})$ from following the equilibrium strategy. This is because for any realization of the exogenous revelation date, the receiver's total news utility in the good state along the equilibrium path is given by $\sum_{j=1}^{J} \mu(p_{t-1+j} - p_{t-2+j}) + \mu(1 - p_{t-1+J})$ for some integer $J \geq 1$. By sub-additivity in gains, this sum is strictly larger than $\mu(1 - p_{t-1})$. If the sender deviates to sending b today, then the receiver updates belief to 0 today and belief remains there until the exogenous revelation, when belief updates to 1. So this deviation has the total news utility $\mu(-p_{t-1}) + \mu(1)$. We have

$$\begin{aligned} \mu(1) &< \mu(1 - p_{t-1}) + \mu(p_{t-1}) \\ &\leq \mu(1 - p_{t-1}) - \mu(-p_{t-1}), \end{aligned}$$

where the first inequality comes from sub-additivity in gains, and the second from weak loss aversion. This shows $\mu(-p_{t-1}) + \mu(1) < \mu(1 - p_{t-1})$, so the deviation is strictly worse than sending the equilibrium message.

Finally, at a history containing at least one b, the receiver's belief is the same at all continuation histories. So the sender has no deviation incentives since no deviations affect future beliefs.

We now show necessity. Suppose that we have a (possibly infinite) gradual good news equilibrium given by the sequence $p_0 < p_1 < \cdots < p_t < \ldots$. By Bayesian plausibility and because we are focusing on two-message equilibria the sender must be sending the messages $\{0, p_t\}$ in period t if the state is bad. The sender must thus be indifferent between these two posteriors in the bad state. Formally, $N_B(0; p_t) = N_B(p_{t+1}; p_t)$ for all $t \geq 0$, as long as there is no babbling. Written equivalently in the language of P^*: $p_{t+1} \in P^*(p_t)$ for all $t \geq 0$, as long as there's no babbling, where here $p_0 = \pi_0$. □

OA 3 Additional Results about News Utility with Diminishing Sensitivity

OA 3.1 Preference for Dominated Consumption Lotteries

So far, we have taken the prior distribution over states $\pi_0 \in \Delta(\Theta)$ as exogenously given. Fixing an information structure, a news-utility agent may strictly prefer a dominated distribution over states. This distinguishes our news-utility preference from other preferences, such as recursive preferences and Gul, Natenzon, and Pesendorfer (2019)'s risk consumption preference.

We now give an example. Suppose $T = 2$ and there are two states, $\Theta = \{G, B\}$. Normalize consumption utility to be $v(c_G) = 1$, $v(c_B) = 0$. Let the news utility function be $\mu(z) = \sqrt{z}$ for $z \geq 0$, $\mu(z) = -\lambda\sqrt{-z}$ for $z < 0$, where $\lambda \geq 1$. At time $t = 0$, the agent holds a prior belief π_0 with $\pi_0(G) = p \in [0, 1]$. At time $t = 1$, the agent learns the state perfectly, so π_1 is degenerate with probability 1. Consumption takes place at time $t = 2$. For any λ, the agent strictly prefers state G for sure ($\pi_0(G) = 1$) over state B for sure ($\pi_0(G) = 0$), as both environments provide zero news utility. But, the agent may strictly prefer state B for sure over an interior probability of the good state, $\pi_0(G) = p$. In fact, this happens when $p + p\sqrt{1-p} - \lambda(1-p)\sqrt{p} < 0$, which says $\lambda > \frac{\sqrt{p}(1+\sqrt{1-p})}{1-p}$. A sufficiently loss-averse agent may strictly prefer no chance of winning a consumption lottery than a low chance of winning.

OA 3.2 Residual Consumption Uncertainty

OA 3.2.1 A Model of Residual Consumption Uncertainty

In the main text, we studied a model where the sender has perfect information about the receiver's final-period consumption level.

Now suppose the sender's information is imperfect. In state θ, the receiver will consume a random amount c in period $T + 1$, drawn as $c \sim F_\theta$, deriving from it consumption utility $v(c)$. As before, v is a strictly increasing consumption-utility function. We interpret the state θ as the sender's private information about the receiver's future consumption, while the distribution F_θ captures the receiver's residual consumption uncertainty conditional on what the sender knows. The case where F_θ is degenerate for every $\theta \in \Theta$ nests the baseline model.

Assume that $\mathbb{E}_{c\in F_{\theta'}}[v(c)] \neq \mathbb{E}_{c\in F_{\theta''}}[v(c)]$ when $\theta' \neq \theta''$. We may without loss normalize $\min_{\theta\in\Theta} \mathbb{E}_{c\in F_\theta}[v(c)] = 0$, $\max_{\theta\in\Theta} \mathbb{E}_{c\in F_\theta}[v(c)] = 1$.

The mean-based news-utility function $N(\pi_t \mid \pi_{t-1})$ in this environment is the same as in the environment where the receiver always gets consumption utility $\mathbb{E}_{c\sim F_\theta}[v(c)]$ in state θ. This is because given a pair of beliefs $F_{\text{old}}, F_{\text{new}} \in \Delta(\Theta)$ about the state, the receiver derives news utility $N(F_{\text{new}} \mid F_{\text{old}})$ based on the difference in *expected* consumption utilities, $\mu(\mathbb{E}_{c\sim F_{\text{new}}}[v(c)] - \mathbb{E}_{c\sim F_{\text{old}}}[v(c)])$. So, all of the results in the paper concerning mean-based news utility immediately extend. The two results in the paper that are not specific to mean-based news utility, Propositions 3 and 4, apply to *any* functions $N(\pi_t \mid \pi_{t-1})$ satisfying the continuous differentiability condition stated in Section 2, without requiring any relationship between N and consumptions in different states.

We now define N using Kőszegi and Rabin (2009)'s percentile-based news-utility model with a power-function gain-loss utility, in an environment with residual consumption uncertainty. We apply Proposition 4 to the resulting N and show that one-shot resolution is strictly sub-optimal. This result applies for any $K \geq 2$.

Corollary OA.1. *Consider the percentile-based model with* $\mu(x) = \begin{cases} x^\alpha & x \geq 0 \\ -\lambda(-x)^\alpha & x < 0 \end{cases}$ *for* $0 < \alpha < 1$, $\lambda \geq 1$. *Suppose there are two states* $\theta_G, \theta_B \in \Theta$ *with distributions of consumption utilities* $v(F_{\theta_B}) = \text{Unif}[0, L]$, $v(F_{\theta_G}) = J + v(F_{\theta_B})$ *for some* $L, J > 0$. *One-shot resolution is strictly suboptimal for any finite* T.

Proof. We show that $\lim_{\epsilon\to 0} \frac{N(1_G|(1-\epsilon)1_G\oplus\epsilon 1_B)}{\epsilon} = \infty$ under this set of conditions. The argument behind Proposition 4 then implies some information structure involving perfect revelation of states other than θ_G, θ_B, one-shot bad news, partial good news for the two states θ_G, θ_B is strictly better than one-shot resolution.

For $r \in [0, 1]$, write F_r for the distribution of consumption utilities under the belief $r1_G \oplus (1-r)1_B$.

Note we must have $\int_0^1 c_{F_1}(q) - c_{F_{1-\epsilon}}(q)dq = J\epsilon$, and that $c_{F_1}(q) - c_{F_{1-\epsilon}}(q) \geq 0$ for all q.

Let $q^* = \min(\epsilon \cdot J/L, \epsilon)$. It is the quantile at which $c_{F_{1-\epsilon}}(q^*) = J$.

For all $q \geq q^*$, $c_{F_1}(q) - c_{F_{1-\epsilon}}(q) \leq \epsilon L$.

Case 1: $J \geq L$, so $q^* = \epsilon$.

$$\begin{aligned}\int_0^{q^*} c_{F_1}(q) - c_{F_{1-\epsilon}}(q)dq &= \int_0^\epsilon J - q \cdot \frac{1}{\epsilon} \cdot ((1-\epsilon)L)dq \\ &= J\epsilon - \frac{1}{2}\epsilon(1-\epsilon)L.\end{aligned}$$

This implies $\int_{q^*}^1 c_{F_1}(q) - c_{F_{1-\epsilon}}(q)dq = \frac{1}{2}\epsilon(1-\epsilon)L$.

The worst case is when the difference is ϵL on some q-interval, and 0 elsewhere. For small $\epsilon < 0$ so that $\epsilon L < 1$,

$$\int_{q^*}^{1} (c_{F_1}(q) - c_{F_{1-\epsilon}}(q))^\alpha dq \geq (\epsilon L)^\alpha \cdot \frac{(1/2) \cdot \epsilon(1-\epsilon)L}{\epsilon L}$$
$$= \frac{1}{2}(\epsilon L)^\alpha (1-\epsilon).$$

Therefore, for small $\epsilon > 0$, $\frac{N(1_G|(1-\epsilon)1_G \oplus \epsilon 1_B)}{\epsilon} = \frac{1}{2}\frac{1}{\epsilon^{1-\alpha}} L^\alpha (1-\epsilon)$, which diverges to ∞ as $\epsilon \to 0$.

Case 2: $J < L$, so $q^* = \epsilon J/L$.

$$\int_0^{\epsilon J/L} c_{F_1}(q) - c_{F_{1-\epsilon}}(q) dq = \int_0^{\epsilon J/L} J - q \cdot \frac{1}{\epsilon J/L}(J - \frac{J}{L}\epsilon \cdot L) dq$$
$$= \frac{1}{2}\frac{J^2}{L}\epsilon + \frac{1}{2}\frac{J^2}{L^2}\epsilon^2 L$$
$$< \frac{1}{2}J\epsilon + \frac{1}{2}L\epsilon^2$$

using $J < L$. This then implies $\int_{q^*}^{1} c_{F_1}(q) - c_{F_{1-\epsilon}}(q) dq > \frac{1}{2}J\epsilon - \frac{1}{2}L\epsilon^2$.

So, again using the worst-case of the difference being ϵL on some q-interval, and 0 elsewhere,

$$\frac{N(1_G \mid (1-\epsilon)1_G \oplus \epsilon 1_B)}{\epsilon} > \frac{1}{\epsilon}(\epsilon L)^\alpha \cdot \frac{\frac{1}{2}J\epsilon - \frac{1}{2}L\epsilon^2}{\epsilon L}$$
$$= \frac{1}{\epsilon^{1-\alpha}} L^\alpha \cdot \left(\frac{1}{2}J/L - \frac{1}{2}\epsilon\right).$$

As $\epsilon \to 0$, RHS converges to ∞. □

OA 3.2.2 A Calibration Comparing Percentile-Based News Utility and Mean-Based News Utility

Since Proposition 3's procedure for computing the optimal information structure applies to general N, including both the percentile-based and the mean-based news-utility functions in an environment with residual consumption uncertainty, we can compare the solutions to the sender's problem for these two models.

Consider two states of the world, $\Theta = \{G, B\}$. For some $\sigma > 0$, suppose consumption is distributed normally conditional on θ with $F_G = \mathcal{N}(1, \sigma^2)$, $F_B = \mathcal{N}(0, \sigma^2)$, consumption utility is $v(x) = x$, and gain-loss utility (over consumption) is $\mu(x) = \sqrt{x}$ for $x \geq 0$, $\mu(x) = -1.5\sqrt{-x}$ for $x < 0$. We calculated the optimal information structure for the mean-based model in an analogous environment, as reported in Figure 2.

With the percentile-based model, an agent who believes $\mathbb{P}[\theta = G] = \pi$ has a belief over final consumption given by a mixture normal distribution, $\pi F_G \oplus (1-\pi)F_B$, illustrated in Figure OA.3.

We plot in Figure OA.4 the optimal information structures for $T = 5$, $\sigma = 1$. The optimal information structures for $\sigma = 0.1, 1, 10$ all involve gradual good news, one-shot bad news. Table OA.1 lists the optimal disclosure of good news over time. Not only are the shapes of the concavification problems qualitatively similar to those of the mean-based model, but the resulting optimal information structures also bear striking quantitative similarities.

	$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$
percentile-based, $\sigma = 0.1$	0.50	0.55	0.61	0.69	0.80	1.00
percentile-based, $\sigma = 1$	0.50	0.55	0.62	0.71	0.83	1.00
percentile-based, $\sigma = 10$	0.50	0.56	0.63	0.72	0.84	1.00
mean-based, any σ	0.500	0.556	0.626	0.715	0.834	1.000

Table OA.1: Optimal disclosure of good news. The optimal information structure under a square-root gain-loss function with $\lambda = 1.5$ takes the form of gradual good news, one-shot bad news both in the mean-based model and the percentile-based model for $T = 5$, $\sigma = 0.1, 1, 10$. The table shows belief movements conditional on the good state in different periods.

From Table OA.1, it appears that percentile-based and mean-based models deliver more similar results for larger σ^2. We provide an analytic result consistent with the idea that these two models generate similar amounts of news utility when the state-dependent consumption utility distributions have large variances.

Proposition OA.3. *Suppose $\Theta = \{B, G\}$ and the distributions of consumption utilities in states B and G are Unif$[0, L]$ and Unif$[d, L+d]$ respectively, for $L, d > 0$. Let $N^{perc}(p_2 \mid p_1)$ be the news utility associated with changing belief in $\theta = G$ from p_1 to p_2 in a percentile-based news-utility model with a continuous gain-loss utility μ. Then,*

$$\lim_{L\to\infty} \left(\sup_{0 \le p_1, p_2 \le 1} |N^{perc}(p_2 \mid p_1) - \mu[(p_2 - p_1)d]| \right) = 0.$$

In a uniform environment, if there is enough unresolved consumption risk even conditional on the state θ, then the difference between percentile-based news utility and mean-based news utility goes to zero uniformly across all possible belief changes.[12]

Proof. Let $F_p(x)$ be the distribution function of the mixed distribution $p \cdot \text{Unif}[d, L+d] \oplus (1-p) \cdot \text{Unif}[0, L]$, and $F_p^{-1}(q)$ its quantile function for $q \in [0, 1]$. By a simple calculation,

[12]Lemma 3 in the Online Appendix of Kőszegi and Rabin (2009) states a similar result, but for a different order of limits.

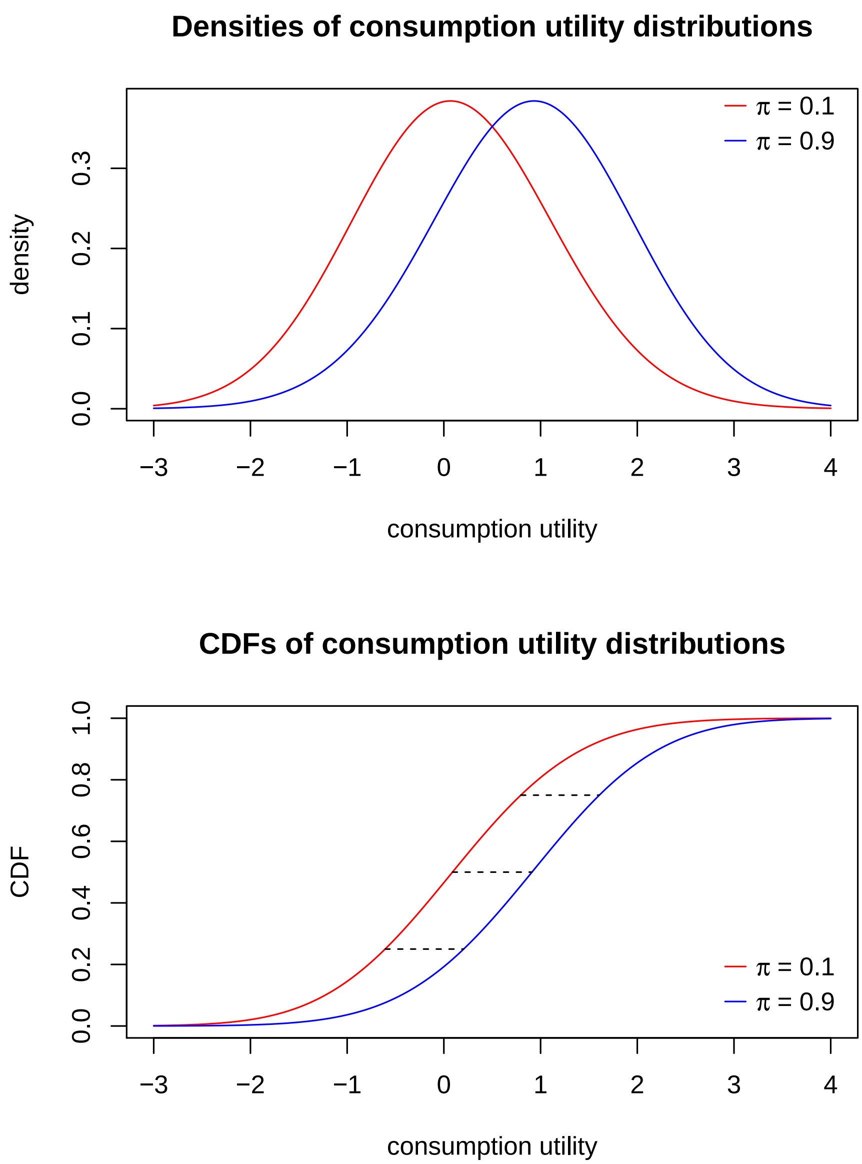

Figure OA.3: The densities and CDFs of final consumption utility distributions under two beliefs about $\mathbb{P}[\theta = G]$, $\pi = 0.1$ and $\pi = 0.9$. The dashed black lines in the CDFs plot show the differences in consumption utilities at the 25th percentile, 50th percentile, and 75th percentile levels between these two beliefs. The news utility associated with updating belief from $\pi = 0.1$ to $\pi = 0.9$ in the percentile-based model is calculated by applying a gain-loss function μ to all these differences in consumption utilities at various quantiles, then integrating over all quantiles levels in $[0, 1]$.

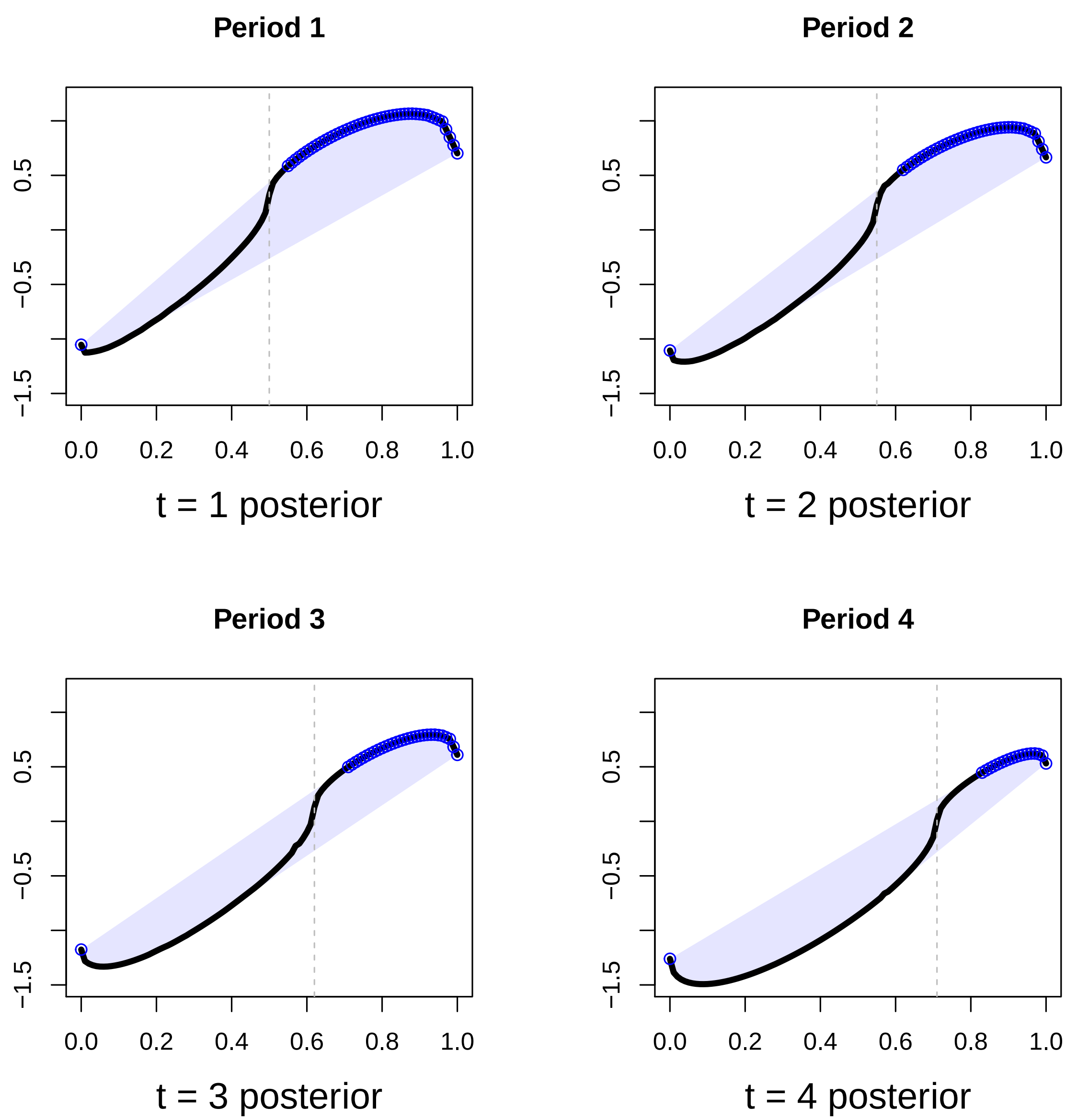

Figure OA.4: The concavifications giving the optimal information structure with horizon $T = 5$, gain-loss function $\mu(x) = \begin{cases} \sqrt{x} & \text{for } x \geq 0 \\ -1.5\sqrt{-x} & \text{for } x < 0 \end{cases}$, prior $\pi_0 = 0.5$, using Kőszegi and Rabin (2009)'s percentile-based model in a Gaussian environment with $\sigma = 1$. The y-axis in each graph shows the sum of news utility this period and the value function of entering next period with a certain belief.

$F_p^{-1}(d/L) = d + pd$ and $F_p^{-1}(1-d/L) = L + pd - d$. At the same time, for $d/L \leq q \leq 1-d/L$ where $q = d/L + y$, we have $F_p^{-1}(q) = d + pd + yL$.

This shows that over the intermediate quantile values between d/L and $1 - d/L$,

$$\int_{d/L}^{1-d/L} \mu \left[F_{p_2}^{-1}(q) - F_{p_1}^{-1}(q)\right] dq = \int_{d/L}^{1-d/L} \mu \left[(p_2 - p_1)d\right] dq = (1 - 2d/L) \cdot \mu[(p_2 - p_1)d].$$

For the lower part of the quantile integral $[0, d/L]$, using the fact that $F_p^{-1}(d/L) = d + pd$, we have the uniform bound $0 \leq F_p^{-1}(q) \leq 2d$ for all $p \in [0,1]$ and $q \leq d/L$. So,

$$\left| \int_0^{d/L} \mu \left[F_{p_2}^{-1}(q) - F_{p_1}^{-1}(q)\right] dq \right| \leq \frac{d}{L} \cdot \max_{x \in [-2d, 2d]} |\mu(x)|.$$

By an analogous argument,

$$\left| \int_{1-d/L}^{1} \mu \left[F_{p_2}^{-1}(q) - F_{p_1}^{-1}(q)\right] dq \right| \leq \frac{d}{L} \cdot \max_{x \in [-2d, 2d]} |\mu(x)|.$$

So for any $0 \leq p_1, p_2 \leq 1$,

$$|N^{\text{perc}}(p_2 \mid p_1) - \mu[(p_2 - p_1)d]| \leq \frac{2d}{L} \max_{x \in [d,d]} |\mu(x)| + \frac{2d}{L} \max_{x \in [-2d, 2d]} |\mu(x)|,$$

an expression not depending on p_1, p_2. The max terms are seen to be finite by applying extreme value theorem to the continuous μ, so the RHS tends to 0 as $L \to \infty$. $\square$

OA 4 Relation to Other Models

OA 4.1 Optimal Information Structure for Anticipatory Utility

We show that if the receiver has anticipatory utility and gets $A\left(\sum \pi_t(\theta) \cdot v(c_\theta)\right)$ when she ends period t with posterior belief $\pi_t \in \Delta(\Theta)$, then a sender with commitment power has an optimal information structure that only discloses information in period $t = 1$.

Consider any information structure (M, σ). Find the period t^* with the highest ex-ante anticipatory utility, i.e., $t^* \in \underset{1 \leq t \leq T-1}{\arg\max} \mathbb{E}_{(M,\sigma)}\left[A\left(\sum \pi_t(\theta) \cdot v(c_\theta)\right)\right]$. Consider another information structure that generates the (feasible) distribution of beliefs π_{t^*} in period 1, then reveals no additional information in periods $2, ..., T-1$. This new information structure gives weakly higher expected anticipatory utility than (M, σ) in every period. Therefore there exists an optimal information structure that only discloses information in $t = 1$.

OA 4.2 Risk Consumption Preferences

Gul, Natenzon, and Pesendorfer (2019) study a model of preference over random evolving lotteries and propose a class of risk consumption preferences. Translated into our setting, an agent with risk consumption preference values an information structure (M, σ) according to utility function

$$\mathbb{E}_{(M,\sigma)}\left[\int v(u_2(\pi_t))d\eta\right].$$

Here $u_2 : \Delta(\Theta) \to \mathbb{R}$ is affine and v is strictly increasing. The term $v(u_2(\pi_t))$ is viewed as a function from the time periods $\{0, 1, ..., T-1\}$ into the reals and $d\eta$ denotes the Choquet integral with respect to a capacity η on $\{0, 1, ..., T-1\}$.

To show that our model of mean-based news utility is not nested under the class of risk consumption preferences, we show that risk consumption preferences cannot exhibit the preference patterns from Online Appendix OA 3.1: that is, strictly preferring winning a lottery for sure to not winning it for sure, but also strictly preferring not winning for sure to winning with some interior probability $p \in (0, 1)$ in the $T = 2$ setup.

By an abuse of notation, the belief assigning probability q to state G will simply be denoted q. The first part of the preference gives $v(u_2(1)) > v(u_2(0))$, since Choquet integral of a constant function returns the same constant. When the prior winning probability is $p \in (0, 1)$, the Choquet integrand is either $f_G : \{0, 1\} \to \mathbb{R}$ with $f_G(0) = v(u_2(p))$ and $f_G(1) = v(u_2(1))$, or $f_B : \{0, 1\} \to \mathbb{R}$ with $f_B(0) = v(u_2(p))$ and $f_B(0) = v(u_2(0))$. The two integrands correspond to belief paths where the agent wins or loses the lottery. Since v is strictly increasing, u_2 is affine, and $v(u_2(1)) > v(u_2(0))$, we have $v(u_2(p)) > v(u_2(0))$. Thus both f_G and f_B dominate the constant function $v(u_2(0))$ in every period. By monotonicity of the Choquet integral, the agent must prefer p probability of winning the lottery to no chance of winning it.