Dynamic Information Design with
Diminishing Sensitivity Over News

Jetlir Duraj† Kevin He‡

First version: July 1, 2019
This version: September 6, 2020

Abstract

A Bayesian agent experiences gain-loss utility each period over changes in belief about future consumption (“news utility”), with diminishing sensitivity over the magnitude of news. We show the agent’s preference between an information structure that delivers news gradually and another that resolves all uncertainty at once depends on his consumption ranking of different states. One-shot resolution is better than gradual bad news, but it is not optimal among all information structures (under common functional forms). In a dynamic cheap-talk framework where a benevolent sender communicates the state over multiple periods, the babbling equilibrium is essentially unique without loss aversion. More loss-averse agents may enjoy higher news utility in equilibrium, contrary to the commitment case. We characterize the family of gradual good news equilibria that exist with high enough loss aversion, and find the sender conveys progressively larger pieces of good news. We discuss applications to media competition and game shows.

Keywords: diminishing sensitivity, news utility, dynamic information design, cheap talk, preference over skewness of information

*We thank Krishna Dasaratha, David Dillenberger, Ben Enke, Drew Fudenberg, Simone Galperti, Jerry Green, Faruk Gul, David Hagmann, Marina Halac, Johannes Hörner, David Laibson, Shengwu Li, Jonathan Libgober, Elliot Lipnowski, Erik Madsen, Pietro Ortoleva, Matthew Rabin, Collin Raymond, Joel Sobel, Charlie Sprenger, Tomasz Strzalecki, the MIT information design reading group, and our seminar participants for insightful comments. Kevin He thanks the California Institute of Technology for hospitality when some of the work on this paper was completed. All remaining errors are our own.

†University of Pittsburgh. Email: jetlirduraj@gmail.com
‡University of Pennsylvania. Email: hesichao@gmail.com
1 Introduction

People are sometimes willing to pay a cost to change how they receive news over time, even when the information in question does not help them make better decisions. Consider the following two scenarios:

**Scenario I.** Ann interviews for her dream job and is told that she will receive the decision by email next week. Ann knows that if the firm decides to reject her, she will receive the rejection email next Friday. But if the firm decides to hire her, she could hear back on any of the weekdays — in other words, no news is bad news. To avoid experiencing multiple instances of disappointment over the week in case she does not hear back for several days, Ann sets up an email filter to automatically redirect any emails from the firm into a holding tank, then releases all messages from the holding tank into her inbox at 5PM next Friday.

**Scenario II.** Once a week, Bob phones his relative who lives in another country. In today’s phone call, Bob learns that the relative has been quarantined due to possible exposure to a dangerous virus. If the relative has been infected, she will suddenly fall sick sometime next week. But if she remains healthy by their next weekly phone call, then she has definitely not been infected. Bob decides to change his routine and calls the relative daily over the next week, because he prefers to receive multiple pieces of reassuring news about her health.

In these scenarios, agents may be willing to exert costly effort to modify their informational environments because they experience *diminishingly sensitive* psychological reactions to good and bad news. They are elated by good news and disappointed by bad news in every period, and multiple congruent pieces of news received in different periods carry a greater total emotional impact than if the pieces were aggregated and the lump-sum news delivered in a single period. This kind of psychological consideration also influences how people convey news to others. When CEOs announce earnings forecasts to shareholders and when organization leaders update their teams about recent developments, they are surely mindful of their information’s emotional impact (in addition to its possible instrumental value). Finally, the psychological effects of news also play a prominent role in designing entertainment content like game shows, where the audience experiences positive and negative reactions over time to news and developments that have no bearing on their personal decision-making.

In this paper, we study the implications of diminishingly sensitive reactions to news for informational preference and dynamic communication. A person’s future consumption depends on an unknown state of the world. In each period, he observes some information about the state and experiences gain-loss utility over the *change* in his belief about said future consumption (“*news utility*”). How does this person prefer to learn about the state over time? If there is another agent who knows the state and who wants to maximize the
first person’s expected welfare, how will this informed agent communicate her information?

Our main innovation is the focus on diminishing sensitivity — a classical but surprisingly under-studied assumption. Diminishing sensitivity in reference dependence traces back to Kahneman and Tversky (1979)’s original formulation of prospect theory. Based on Weber’s law and experimental findings about human perception, these authors envisioned a gain-loss utility based on deviations from a reference point, where larger deviations carry smaller marginal effects. Diminishing sensitivity is mentioned in much of the subsequent work on reference-dependent preferences, including Kőszege and Rabin (2009), who first introduced a model of news utility. In almost all cases, however, researchers then specialize for simplicity to a two-part linear gain-loss utility function that allows for loss aversion but precludes diminishing sensitivity. Four decades since Kahneman and Tversky (1979)’s publication, O’Donoghue and Sprenger (2018)’s review of the ensuing literature summarizes the situation:

“Most applications of reference-dependent preferences focus entirely on loss aversion, and ignore the possibility of diminishing sensitivity [...] The literature still needs to develop a better sense of when diminishing sensitivity is important.”

We show that diminishing sensitivity leads to novel and testable predictions in the domain of information design. The first prediction concerns the choice between gradual information and one-shot information. As Kőszege and Rabin (2009) point out, the two-part linear news-utility model predicts that people prefer resolving all uncertainty in one period (“one-shot resolution”) over any other dynamic information structure. At the same time, some other theories (e.g., Ely, Frankel, and Kamenica (2015)’s suspense and surprise utility) predict that one-shot resolution is the worst possible information structure. Unlike these theories that predict an agent will either always choose or always avoid one-shot information, news utility with diminishing sensitivity predicts the same person can make different choices in different situations — in particular, it depends on his consumption ranking over the states.

In a world with two possible states (A and B) associated with two different consumption prizes, suppose state A realizes if and only if a sequence of intermediate events all take place successfully over time. We show that when the agent prefers the consumption prize in state A, he will choose to observe the intermediate events resolve in real-time (gradual information). But when he prefers the consumption prize in state B, he will choose to only learn the final state (one-shot information). At the population level, this result shows that an underlying diversity in consumption preferences within a society can create a diversity in informational preferences, and suggests a mechanism for media competition. The result also rationalizes a “sudden death” format often found in game shows, where the contestant must overcome every challenge in a sequence to win the grand prize (as opposed to the grand prize being contingent on beating at least one of several challenges.)
The above relationship between consumption preference and informational preference arises because diminishing sensitivity generates a preference over the direction of news skewness. We show that information structures where good news arrives all at once but bad news arrives gradually in small pieces — such as waiting for the job offer in Ann’s scenario — are strictly worse than one-shot resolution. On the other hand, we also show that some information structures with the opposite skewness — good news arrives gradually but bad news all at once, such as Bob’s information about his relative’s health when he makes daily phone calls — are strictly better than one-shot resolution for a class of news-utility functions with diminishing sensitivity. This class includes the commonly used power-function specification. It also includes a tractable quadratic specification, whenever diminishing sensitivity is strong enough relative to the degree of loss aversion.

Another novel consequence of diminishing sensitivity is a credibility problem for an informed benevolent sender who knows the state but lacks commitment power. The information structures that strictly improve on one-shot resolution may not be implementable in the equilibrium of a cheap-talk game. In the bad state, the sender may strictly prefer to lie and convey a positive message intended for the good state. This temptation exists despite the fact that the sender is far-sighted and maximizes the receiver’s total news utility over time. The intuition is that when the sender knows the state is bad, she also knows the receiver will experience disappointment in the future. Diminishing sensitivity implies the receiver’s marginal utility of unwarranted partial good news today is larger than his marginal disutility of heightened future disappointment. This perverse incentive to provide false hope in the bad state may prevent any meaningful communication at all. Indeed, we show that if the receiver has diminishing sensitivity but no loss aversion (or has low loss aversion), then every equilibrium is payoff-equivalent to the babbling equilibrium, which implements one-shot resolution as the receiver’s belief stays constant until the state is exogenously revealed in the final period. But, high enough loss aversion can restore the equilibrium credibility of good-news messages by increasing the future disappointment cost of inducing false hope in the bad state. As a consequence, receivers with higher loss aversion may enjoy higher equilibrium payoffs, which does not happen when the sender has commitment power. Diminishing sensitivity thus drives a wedge between the commitment solution and the equilibrium outcome, whereas the two coincide without it.

With enough loss aversion, there exist non-babbling equilibria featuring gradual good news. We characterize the entire family of such equilibria and study how quickly the receiver learns the state. For a class of news-utility functions that include the square-root and quadratic specifications mentioned before, the sender always conveys progressively larger pieces of good news over time, so the receiver’s equilibrium belief grows at an increasing
rate in the good state. The idea is that in equilibrium, the sender must be made indifferent between giving false hope and telling the truth in the bad state, and diminishing sensitivity implies that sustaining said indifference requires a greater amount of false hope when the receiver’s current belief is more optimistic. This conclusion also puts a uniform bound on the number of periods of informative communication across all time horizons and all equilibria.

The rest of the paper is organized as follows. Section 2 defines the timing of events and introduces a model of news utility. This model is then embedded into three different environments in the next three sections. Section 3 considers agents choosing between gradual information and one-shot information about the state, and how their choices depend on their consumption rankings of the states. Section 4 studies a sender-receiver framework, where a benevolent informed sender with commitment power communicates the state to a receiver who experiences news utility. Section 5 drops the commitment assumption and focuses on the credibility problems in the resulting cheap-talk game. Section 6 discusses related literature and contrasts our results with the predictions of other models of preference over non-instrumental information. It also relates our theoretical results to experiments on preference over information. Section 7 concludes. As an extension, Online Appendix OA 2 looks at a variant of the model without a deterministic horizon. Proofs of main results appear in the Appendix and proofs of remaining results appear in the Online Appendix.

2 Model

2.1 Timing of Events

We consider a discrete-time model with periods 0, 1, 2, ..., T, where T ≥ 2. There is a finite state space Θ with |Θ| = K ≥ 2. There is an agent who experiences news utility over consumption (to be explained below). In state θ, the agent receives a consumption prize cθ ∈ R in period T, deriving from it consumption utility v(cθ) where v is strictly increasing. There is no consumption in other periods, and we assume that cθ′ ≠ cθ″ when θ′ ≠ θ″. We may normalize without loss minθ∈Θ[v(cθ)] = 0, maxθ∈Θ[v(cθ)] = 1.

The agent starts with a prior belief π0 ∈ Δ(Θ) about the state, where π0(θ) > 0 for all θ ∈ Θ. In every period t = 1, ..., T, the agent observes some information and forms the Bayesian posterior belief πt ∈ Δ(Θ) about the state. The information is non-instrumental in that no actions taken in these interim periods affect the state or the consumption utility in period T.
2.2 News Utility

The agent derives utility based on changes in his belief about the final period’s consumption. Specifically, he has a continuous news-utility function \( N : \Delta(\Theta) \times \Delta(\Theta) \to \mathbb{R} \), mapping his pair of new and old beliefs about the state into a real-valued felicity.\(^1\) He receives utility \( N(\pi_t \mid \pi_{t-1}) \) at the end of period \( 1 \leq t \leq T \). Utility flow is undiscounted and the agent has the same \( N \) in all periods,\(^2\) so his total payoff is \( \sum_{t=1}^{T} N(\pi_t \mid \pi_{t-1}) + v(c) \). We assume for every \( \pi \in \Delta(\Theta) \), both \( N(\cdot \mid \pi) \) and \( N(\pi \mid \cdot) \) are continuously differentiable except possibly at \( \pi \).

For many of our results, we study a mean-based news-utility model. Kószegi and Rabin (2009) discuss this model, but mostly focus on another model that makes percentile-by-percentile comparisons between old and new beliefs. We use the mean-based model to derive the implications of diminishing sensitivity in the simplest setup. The agent applies a gain-loss utility function, \( \mu : [-1, 1] \to \mathbb{R} \), to changes in expected period-\( T \) consumption utility. That is, \( N(\pi_t \mid \pi_{t-1}) = \mu ([\sum_{\theta \in \Theta} \pi_t(\theta)v(c_\theta)] - [\sum_{\theta \in \Theta} \pi_{t-1}(\theta)v(c_\theta)]) \). Throughout we assume \( \mu \) is continuous, strictly increasing, twice differentiable except possibly at 0, and \( \mu(0) = 0 \). We maintain further assumptions on \( \mu \) to reflect diminishing sensitivity and loss aversion.

**Definition 1.** Say \( \mu \) satisfies diminishing sensitivity if \( \mu''(x) < 0 \) and \( \mu''(-x) > 0 \) for all \( x > 0 \). Say \( \mu \) satisfies (weak) loss aversion if \( -\mu(-x) \geq \mu(x) \) for all \( x > 0 \). There is strict loss aversion if \( -\mu(-x) > \mu(x) \) for all \( x > 0 \).

We now discuss two important functional forms of \( \mu \). In Online Appendix OA 3.2.2, we compare the optimal information structures for this model and for Kószegi and Rabin (2009)’s percentile-based model, a class of news-utility functions that do not admit mean-based representations.

### 2.2.1 Quadratic News Utility

The quadratic news-utility function \( \mu : [-1, 1] \to \mathbb{R} \) is given by

\[
\mu(x) = \begin{cases} 
\alpha_p x - \beta_p x^2 & x \geq 0 \\
\alpha_n x + \beta_n x^2 & x < 0 
\end{cases}
\]

---

\(^1\) Since different states lead to different levels of consumption, beliefs over states induce beliefs over consumption.

\(^2\) Our preference satisfies Segal (1990)’s time neutrality axiom. We abstract away from preferences for early or late resolution of uncertainty.
with $\alpha_p, \beta_p, \alpha_n, \beta_n > 0$. So we have

$$
\mu'(x) = \begin{cases} 
\alpha_p - 2\beta_p x & x > 0 \\
\alpha_n + 2\beta_n x & x < 0
\end{cases}, \quad
\mu''(x) = \begin{cases} 
-2\beta_p & x > 0 \\
2\beta_n & x < 0
\end{cases}.
$$

The parameters $\alpha_p, \alpha_n$ control the extent of loss aversion near 0, while $\beta_p, \beta_n$ determine the amount of curvature — i.e., the second derivative of $\mu$. The maintained general assumptions on $\mu$ imply the following parametric restrictions.

1. **Monotonicity**: $\alpha_p > 2\beta_p$ and $\alpha_n > 2\beta_n$. These inequalities hold if and only if $\mu$ is strictly increasing.

2. **Loss aversion**: $\alpha_n - \alpha_p \geq (\beta_n - \beta_p)z$ for all $z \in [0, 1]$. This condition is equivalent to loss aversion from Definition 1 for this class of news-utility functions.

A family of quadratic news-utility functions that satisfy these two restrictions can be constructed by choosing any $\alpha > 2\beta > 0$ and $\lambda \geq 1$, then set $\alpha_p = \alpha$, $\alpha_n = \lambda \alpha$, $\beta_p = \beta$, $\beta_n = \lambda \beta$. Figure 1 plots some of these news-utility functions for different values of $\alpha, \beta,$ and $\lambda$.

The quadratic news utilities are simple enough to be tractable but rich enough to exhibit both diminishing sensitivity and loss aversion. We show later in Section 4.5 that we can explicitly characterize the optimal information structure for these utility functions.

### 2.2.2 Power-Function News Utility

The power-function news-utility $\mu : [-1, 1] \to \mathbb{R}$ is given by
\[ \mu(x) = \begin{cases} 
  x^\alpha & x \geq 0 \\
  -\lambda |x|^\beta & x < 0 
\end{cases} \]

with \( 0 < \alpha, \beta < 1 \) and \( \lambda \geq 1 \). Parameters \( \alpha, \beta \) determine the degree of diminishing sensitivity to good news and bad news, while \( \lambda \) controls the extent of loss aversion. This class of functions nests the square-root case when \( \alpha = \beta = 0.5 \) and is the only class of gain-loss functions to appear in Tversky and Kahneman (1992).

3 Choosing Between Gradual Information and One-Shot Information

Consider an environment where a sequence of signal realizations gradually determine a binary state. We show that agents with opposite consumption preferences over the two states can exhibit opposite preferences when choosing between observing the signals as they arrive or only learning the final state.

There are two states of the world, \( A \) and \( B \). In each period \( t = 1, 2, ..., T \), a binary signal \( X_t \) realizes, where \( \mathbb{P}[X_t = 1] = q_t \) with \( 0 < q_t < 1 \). Each \( X_t \) is independent of the other ones. The signals determine the state. If \( X_t = 1 \) for all \( t \), then the state is \( A \). Otherwise, when \( X_t = 0 \) for at least one \( t \), the state is \( B \). At time 0, the agent chooses between observing the realizations of the signals \( (X_t)_{t=1}^T \) in real time (gradual information), or only learning the state of the world at the end of period \( T \) (one-shot information).

As an example, imagine a televised debate between two political candidates \( A \) and \( B \) where \( A \) loses as soon as she makes a "gaffe" during the debate.\(^3\) If \( A \) does not make any gaffes, then \( A \) wins. In this example, \( \{X_t = 1\} \) corresponds to the event of that candidate \( A \) does not make a gaffe during the \( t \)-th minute of the debate. States \( A \) and \( B \) correspond to candidates \( A \) and \( B \) winning the debate. An individual chooses between watching the debate live (i.e., observing the stochastic process \( (X_t) \) in real time) or only reading the outcome of the debate the following morning (i.e., getting one-shot information about the state).

The individual could be someone who benefits from candidate \( A \) winning the debate (that is, \( v(c_A) = 1, v(c_B) = 0 \)), or someone who benefits from candidate \( B \) winning the debate (that is, \( v(c_A) = 0, v(c_B) = 1 \)). The agent forms the Bayesian belief \( \pi_t \in [0,1] \) about the probability of state \( A \) at the end of each period \( t \), starting with the correct Bayesian prior \( \pi_0 \). For notational convenience, we also write \( \rho_t = 1 - \pi_t \) as the belief in state \( B \) at the end.

\(^3\)Augenblick and Rabin (2019) use a similar example of political gaffes to illustrate Bayesian belief movements.
of $t$, with the prior $\rho_0 = 1 - \pi_0$. If the agent prefers state $A$, he gets news utility $\mu(\pi_t - \pi_{t-1})$ at the end of period $t$. If the agent prefers state $B$, then he gets news utility $\mu(\rho_t - \rho_{t-1})$

Under diminishing sensitivity, someone rooting for state $B$ prefers to only learn the final state to avoid piecemeal bad new, while someone hoping for state $A$ wants to watch the events unfold in real time to “celebrate the small victories,” provided they are not too loss averse. The next proposition formalizes this intuition. To quantify the amount of loss aversion, we consider the parametric class of $\lambda$-scaled news-utility functions. We fix some $\tilde{\mu}_{pos} : [0, 1] \to \mathbb{R}_+$, strictly increasing and strictly concave with $\tilde{\mu}_{pos}(0) = 0$, and consider the family of $\mu$’s given by $\mu_\lambda(x) = \tilde{\mu}_{pos}(x)$, $\mu_\lambda(-x) = -\lambda \tilde{\mu}_{pos}(x)$ for $x > 0$ as we vary $\lambda \geq 1$.

**Proposition 1.** Consider a class of $\lambda$-scaled news-utility functions $(\mu_\lambda)_{\lambda \geq 1}$. For any $\lambda \geq 1$, the agent chooses one-shot information over gradual information when $v(c_A) = 0$, $v(c_B) = 1$. There exists some $\bar{\lambda} > 1$ so that for any $1 \leq \lambda \leq \bar{\lambda}$, the agent chooses gradual information over one-shot information when $v(c_A) = 1$, $v(c_B) = 0$.

Here the bound on loss aversion $\bar{\lambda}$ depends on the degree of diminishing sensitivity in $\tilde{\mu}_{pos}$, but it is always strictly larger than 1 when $\tilde{\mu}_{pos}$ is strictly concave.

Recall Ann and Bob from the introduction, who choose to undertake costly effort to change their informational environments. This result explains their behavior. In Ann’s job-application scenario, $T = 5$ and the event $\{X_t = 1\}$ corresponds to not getting a job offer from the firm on the $t$-th weekday. State $A$ is being rejected by the firm and state $B$ is being hired by the firm, so $v(c_A) < v(c_B)$. By Proposition 1, the gradual information inherent in her scenario is strictly worse than one-shot information, which Ann implements using the email filter.

In Bob’s scenario with the quarantined relative, $T = 7$ and $\{X_t = 1\}$ corresponds to his relative not falling sick on the $t$-th day from today. State $A$ is the relative not being infected by the virus, while state $B$ is the relative being infected, so $v(c_A) > v(c_B)$. By Proposition 1, if Bob’s loss aversion is low, he strictly prefers gradual information about the relative’s health over the one-shot information of his default informational environment. So, he is willing to pay a cost to switch from weekly calls to daily calls.

Proposition 1 carries implications about information choices at the population level and at the individual level.

At the population level, Proposition 1 shows that society can exhibit an endogenous diversity of information preferences, driven by an underlying diversity of consumption preferences. Individuals with the same news-utility function $\mu$ can nevertheless choose to learn about the state of the world in two different ways, if they have opposite rankings of the states in terms of their consumption levels. So heterogeneous consumption preferences generate heterogeneous information preferences.
This observation suggests a possible mechanism for media competition: if the realization of some state $A$ depends on a series of smaller events, then some news sources may cover these small events in detail as they happen, while other sources may choose to only report the final outcome. If there is a heterogeneity of tastes over states in the society, then viewers will sort between these two kinds of news sources based on how they rank states $A$ and $B$ in terms of consumption.

At the individual level, Proposition 1 shows that the same person may choose gradual information in one situation but one-shot information in another, even if his news-utility function remains stable. For example, if political candidate $X$ wins any debate when and only when she does not make a gaffe, an agent may choose to watch a debate between candidates $X$ and $Y$ but refuse to watch a debate between candidates $X$ and $Z$, because he prefers $X$ over $Y$ but $Z$ over $X$.

By contrast, related theories about behavioral information preference tend to predict that the agent either always prefers one-shot information in all situations, or always prefers every other information structure to one-shot information in all situations. The theories that do not allow situation-dependent information choice include news-utility without diminishing sensitivity (even if the agent is gain-loving instead of loss-averse, as in Goette, Graeber, Kellogg, and Sprenger (2020)), anticipatory utility, and suspense and surprise.

**Proposition 2.** The following models predict that the agent will not change his choice between gradual information and one-shot information when the sign of $v(c_A) - v(c_B)$ changes.

1. News utility with a two-part linear $\mu$, where $\mu(x) = x$ for $x \geq 0$ and $\mu(x) = \lambda x$ for $x < 0$, with any $\lambda \geq 0$.

2. Anticipatory utility where the agent gets either $u(\pi_t)$ or $u(1 - \pi_t)$ in period $t$ depending on his preference over states $A$ and $B$, with $u$ an increasing, weakly concave function.


A final application of Proposition 1 concerns the design of game shows. Consider a game show featuring a single contestant who will win either $100,000 or nothing depending on her performance across five rounds.\textsuperscript{4} The audience, empathizing with the contestant, derives news utility $\mu(\pi_t - \pi_{t-1})$ at the end of round $t$, where $\pi_t$ is the contestant’s probability of winning the prize based on the first $t$ rounds. One possible format (“sudden death”) features five easy rounds each with $w = 0.5^{1/5} \approx 87\%$ winning probability, where the contestant wins $100,000 if she wins all five rounds. Another possible format (“repêchage”) involves

\textsuperscript{4}This can be thought of as a stylized payout structure for game shows like American Ninja Warrior and Who Wants to Be a Millionaire.
five hard rounds each with $1 - w$ winning probability, but the contestant wins $100,000 as soon as she wins any round. Both formats lead to the same distribution over final outcomes and generate the same amount of suspense and surprise utilities à la Ely, Frankel, and Kamenica (2015). Proposition 1 shows the first format induces more news utility than one-shot information (which could correspond to not watching the game show and simply looking up the contestant’s outcome later) for audience members who are not too loss averse, while the second format is worse than one-shot information for all audience members. Consistent with this prediction, the vast majority of game shows resemble the first format more than the second format.

4 Benevolent Informed Sender with Commitment Power

In this section, we consider a sender-receiver framework where the agent with news-utility preference plays the role of the receiver. A benevolent sender with commitment power knows the state and communicates it to the receiver over $T$ periods. After setting up the environment, we provide a general inductive procedure to solve the sender’s problem. We show that information structures featuring gradual bad news, one-shot good news are strictly worse than one-shot resolution, then identify sufficient conditions that imply the optimal information structure features gradual good news, one-shot bad news. We illustrate these results with the quadratic news-utility specification, finding that the said sufficient conditions hold whenever diminishing sensitivity is sufficiently strong relative to loss aversion, and explicitly characterize the optimal information structure.

4.1 The Sender-Receiver Framework

In period 0, the sender chooses a finite message space $M$ and a strategy $\sigma = (\sigma_t)_{t=1}^{T-1}$, where $\sigma_t(\cdot \mid h^{t-1}, \theta) \in \Delta(M)$ is a distribution over messages in period $t$ that depends on the public history $h^{t-1} \in H^{t-1} := (M)^{t-1}$ of messages sent so far, as well as the true state $\theta$. The sender can commit to any information structure $(M, \sigma)$, which then becomes common knowledge between the players. At the start of period 1, the sender privately observes the state’s realization, then sends a message in each of the periods $1, 2, ..., T - 1$ according to the strategy $\sigma$. The sender is benevolent and maximizes the receiver’s expected total welfare.\(^5\)

At the end of period $t$ for $1 \leq t \leq T - 1$, the receiver forms the Bayesian posterior belief $\pi_t$ about the state after the on-path history $h^t \in H^t$ of $t$ messages. This belief is rational and

\(^5\)The problem of a benevolent sender with commitment power is equivalent to a single-agent framework where the receiver chooses an information structure for himself.
calculated with the knowledge of the information structure \((M, \sigma)\). In period \(T\), the receiver exogenously and perfectly learns the true state \(\theta\), consumes \(c_\theta\), and the game ends.

Since the receiver is Bayesian, the sender faces cross-state constraints in choosing the receiver’s belief paths. In view of diminishing sensitivity, one might conjecture that the sender should concentrate all bad news in period 1 if the state is bad, and deliver equally-sized pieces of good news in periods 1, 2, 3, ... if the state is good. But these belief paths are infeasible, since a Bayesian audience who knows this strategy and does not receive bad news in period 1 will conclusively infer that the state is good. The receiver would not judge subsequent communication from the sender as further good news or derive positive news utility from them. Indeed, if the sender wishes to use some message \(m \in M\) to convey positive but inconclusive news in the good state, then in the bad state the same message must also be sent with positive probability – otherwise, receiving this information in the first period would amount to conclusive evidence of the good state. These cross-state constraints imply distortions from perfect “consumption smoothing” of good news, as we show later.

When \(K = 2\), we label two states as Good and Bad, \(\Theta = \{G, B\}\), so that \(v(c_G) = 1\), \(v(c_B) = 0\). We also abuse the notation \(\pi_t\) to mean \(\pi_t(G)\) in the case of binary states.

In this environment, the sender has perfect information about the receiver’s future consumption level once she observes the state. Online Appendix OA 3.2 discusses an extension where the sender’s information is imperfect, so that there is residual uncertainty about the receiver’s consumption conditional on the state (i.e., given the sender’s private information).

### 4.2 A General Backwards-Induction Procedure

For \(f : \Delta(\Theta) \to \mathbb{R}\), let \(\text{cav} f\) be the concavification of \(f\) — that is, the smallest concave function that dominates \(f\) pointwise. Concavification plays a key role in solving this information design problem, just as in Kamenica and Gentzkow (2011) and Aumann and Maschler (1995).

For \(\pi_{T-2}, \pi_{T-1} \in \Delta(\Theta)\) two beliefs about the state, let \(U_{T-1}(\pi_{T-1} \mid \pi_{T-2})\) be the sum of the receiver’s expected news utilities in periods \(T-1\) and \(T\), if he enters period \(T-1\) with belief \(\pi_{T-2}\) and updates it to \(\pi_{T-1}\). More precisely,

\[
U_{T-1}(\pi_{T-1} \mid \pi_{T-2}) := N(\pi_{T-1} \mid \pi_{T-2}) + \sum_{\theta \in \Theta} \pi_{T-1}(\theta) \cdot N(1_\theta \mid \pi_{T-1}),
\]

where \(1_\theta\) is the degenerate belief putting probability 1 on the state \(\theta\). Note that by the martingale property of beliefs, if the receiver holds belief \(\pi_{T-1}\) at the end of period \(T-1\), then state \(\theta\) must then realize in period \(T\) with probability \(\pi_{T-1}(\theta)\).

Let \(U^*_{T-1}(\pi_{T-2}) := (\text{cav} U_{T-1}(\cdot \mid \pi_{T-2})) (\pi_{T-2})\). As we will show in the proof of Proposition
3. $U^*_t(\pi_{T-2})$ is the value function of the sender when the receiver enters period $T-1$ with belief $\pi_{T-2}$. Continuing inductively, using the value function $U^*_t(\cdot)$ for $t \geq 1$, we may define $U_t(\pi_t \mid \pi_{t-1}) := N(\pi_t \mid \pi_{t-1}) + U^*_{t+1}(\pi_t)$, which leads to the period $t$ value function $U^*_t(x) := (\text{ca}vU_t(\cdot \mid x))(x)$. The maximum expected news utility across all information structures is $U^*_1(\pi_0)$, and the sequence of concavifications give the optimal information structure.

**Proposition 3.** The maximum expected news utility across all information structures is $U^*_1(\pi_0)$. There is an information structure $(M, \sigma)$ with $|M| = K$ attaining this maximum, with the property that after each on-path public history $h^{t-1}$ associated with belief $\pi_{t-1}$, the sender’s strategy $\sigma_t(\cdot \mid h^{t-1}, \theta)$ induces posterior $q^k$ at the end of period $t$ with probability $w^k$, for some $q^1, \ldots, q^K \in \Delta(\Theta)$, $w^1, \ldots, w^K \geq 0$, satisfying $\sum_{k=1}^K w^k = 1$, $\sum_{k=1}^K w^k q^k = \pi_{t-1}$, and $U^*_t(\pi_{t-1}) = \sum_{k=1}^K w^k U_t(q^k \mid \pi_{t-1})$.

In practice, explicitly calculating the solution to the sender’s problem may be difficult, since the function to be concavified in each period depends on both current period’s belief and next period’s belief — a key feature of news utility. Nevertheless, in Section 4.5 we show that the optimal information structure can be tractably characterized for the quadratic news utilities we introduced. In Sections 4.3 and 4.4, we discuss some qualitative features of the receiver’s preference over information structures, including the ranking of one-shot resolution among all information structures.

A perhaps surprising implication is that the receiver only needs a binary message space if there are two states of the world, regardless of the shape or curvature of the news-utility function $N$. Figure 2 illustrates the concavification procedure in an environment with two equally likely states, $T = 5$, and the mean-based news-utility function $\mu(x) = \sqrt{x}$ for $x \geq 0$, $\mu(x) = -1.5\sqrt{-x}$ for $x < 0$. In the optimal information structure, there are two signals, $\{g, b\}$. The sender sends the conclusive bad-news signal $b$ in a random period when $\theta = B$, and sends the other signal $g$ in the other periods. Signal $g$ is sent in every period if $\theta = G$. This means $g$ is a partial good-news signal that makes the receiver more optimistic about $\theta = G$, but receiving the $b$ signal is conclusive bad news.

**Remark 1.** The information-design problem imposes additional constraints relative to a habit-formation model. To see this, consider a “relaxed” version of the sender’s problem in the binary-states case where she simply chooses some $x_t \in [0, 1]$ each period for $1 \leq t \leq T-1$, depending on the realization of $\theta$. The receiver gets $\mu(x_t - x_{t-1})$ in period $1 \leq t \leq T$, with the initial condition $x_0 = \pi_0$ and the terminal condition $x_T = 1$ if $\theta = G$, $x_T = 0$ if $\theta = B$. One interpretation of the relaxed problem is that the sender chooses the receiver’s sequence of beliefs only subject to the constraint that the initial belief in period 0 is $\pi_0$ and the final belief in period $T$ puts probability 1 on the true state. The belief paths do not have to be
Figure 2: The concavifications giving the optimal information structure with horizon $T = 5$, mean-based news-utility function $\mu(x) = \begin{cases} \sqrt{x} & \text{for } x \geq 0 \\ -1.5\sqrt{-x} & \text{for } x < 0 \end{cases}$, prior $\pi_0 = 0.5$. The dashed vertical line in the $t$-th graph marks the receiver’s belief in $\theta = G$ conditional on not having heard any bad news by the start of period $t$. The $y$-axis shows the sum of news utility this period and the value function of entering next period with a certain belief. In the good state of the world, the receiver’s belief in $\theta = G$ grows at increasing rates across the periods, $0.5 \rightarrow 0.556 \rightarrow 0.626 \rightarrow 0.715 \rightarrow 0.834 \rightarrow 1$. In the bad state of the world, the receiver’s belief follows the same path as in the good state up until the random period when conclusive bad news arrives.
Bayesian. Another interpretation is that \( x_t \) is not a belief, but a consumption level for period \( t \). The receiver’s welfare in period \( t \) only depends on a gain-loss utility based on how current period’s consumption differs from that of period \( t-1 \). Provided \( \mu \) has diminishing sensitivity, Jensen’s inequality implies that in this relaxed problem, any sender strategy that induces an unevenly increasing \((x_t)_{t=0}^T\) when \( \theta = G \) (as in the optimal solution to the information-design problem in Figure 2) is strictly worse than choosing \( x_t = \pi_0 + \frac{1}{T}(1 - \pi_0) \) in period \( t \) when \( \theta = G \).

### 4.3 Diminishing Sensitivity and the Sub-Optimality of One-Shot Resolution

An information structure features one-shot resolution if \( \mathbb{P}[\pi_t \neq \pi_{t-1} \text{ for at most one } 1 \leq t \leq T] = 1 \). That is, the receiver’s belief path is almost surely constant in all except one period. In Kőszegi and Rabin (2009)’s model of news utility without diminishing sensitivity, one-shot resolution is optimal among all information structures.\(^6\) But this conclusion does not hold when we allow for diminishing sensitivity.

We give a sufficient condition on the news-utility function for one-shot resolution to be strictly suboptimal. Let \( \theta_H, \theta_L \in \Theta \) be the states with the highest and lowest consumption utilities. Let \( 1_H, 1_L \in \Delta(\Theta) \) represent degenerate beliefs in states \( \theta_H \) and \( \theta_L \) and let \( v_0 := \mathbb{E}_{\theta \sim \pi_0}(v(c_\theta)) \) be the ex-ante expected future consumption utility. The symbol \( \oplus \) denotes the mixture between two beliefs in \( \Delta(\Theta) \).

**Proposition 4.** For any \( T \) and \( \Theta \), one-shot resolution is strictly suboptimal if

\[
\lim_{\epsilon \to 0^+} \frac{N(1_H \mid (1 - \epsilon)1_H \oplus \epsilon 1_L)}{\epsilon} = N(1_H \mid \pi_0) - N(1_L \mid \pi_0) < \lim_{\epsilon \to 0^+} \frac{N(1_H \mid \pi_0) - N((1 - \epsilon)1_H \oplus \epsilon 1_L \mid \pi_0)}{\epsilon} - N(1_L \mid 1_H).
\]

For the mean-based news-utility model, this condition is equivalent to

\[
\mu(1 - v_0) - \mu(-v_0) + \mu'(0^+) - \mu'(1 - v_0) + \mu(-1) > 0.
\]

In fact, the proof of Proposition 4 shows that whenever its condition is satisfied, some information structure featuring “gradual good news and one-shot bad news” (to be defined precisely in the next subsection) strictly improves on one-shot resolution.

\(^6\)Kőszegi and Rabin (2009) showed this for their percentile-based model of news utility with binary states, while Dillenberger and Raymond (2020) proved the same also holds for arbitrarily many states.
We can interpret Proposition 4’s sufficient condition as “strong enough diminishing sensitivity relative to loss aversion.” Evidently, $\mu(1-v_0) - \mu(-v_0) > 0$, so the condition is satisfied whenever $\mu'(0^+) - \mu'(1-v_0) + \mu(-1) > 0$. We always have $\mu'(0^+) - \mu'(1-v_0) > 0$, and it increases when $\mu$ becomes more concave in the positive region. We have $\mu(-1) < 0$, but it increases when $\mu$ is more convex in the negative region. On the other hand, holding fixed $\mu'(0^+)$ and the curvature $\mu''(x)$ for $x \neq 0$, increasing the amount of loss aversion near 0 (i.e., $\mu'(0^-) - \mu'(0^+)$) decreases $\mu(-1)$. So, the condition holds if $\mu$ exhibits enough diminishing sensitivity, in the gains or losses domain, relative to the level of loss aversion. (It is easy to show that if $\mu$ is instead two-part linear, then one-shot resolution is optimal.)

The quadratic news utility provides a clear illustration of Proposition 4’s condition as a race between loss aversion and diminishing sensitivity: the condition holds if and only if there is enough curvature relative to the size of the “kink” at 0.

**Corollary 1.** If the receiver has quadratic news utility with $\alpha_n - \alpha_p < \beta_n + \beta_p$, then one-shot resolution is strictly suboptimal for any $T$.

The difference $\alpha_n - \alpha_p \geq \mu'(0^-) - \mu'(0^+)$, which corresponds to the amount of loss aversion in quadratic news utilities. On the other side, $\beta_p$ and $\beta_n$ control the amounts of curvature in the positive and negative regions, respectively, and correspond to the extent of diminishing sensitivity.

The sufficient condition in Proposition 4 is also satisfied by the most commonly used model of diminishing sensitivity, the power function. One could think of the power function specification as having “infinite” diminishing sensitivity near 0, as $\mu''(0^+) = -\infty$ and $\mu''(0^-) = \infty$.

**Corollary 2.** Suppose $\mu(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -\lambda \cdot |x|^{\beta} & \text{if } x < 0 \end{cases}$ for some $0 < \alpha, \beta < 1$ and $\lambda \geq 1$. Then one-shot resolution is strictly suboptimal for any $T$.

While Proposition 4 holds generally, we can find sharper results on the sub-optimality of one-shot resolution for specific news-utility models and environments. Köszegi and Rabin (2009)’s percentile-based news-utility model stipulates

$$N(\pi_t | \pi_{t-1}) = \int_0^1 \mu \left( v(F_{\pi_t}(p)) - v(F_{\pi_{t-1}}(p)) \right) dp,$$

where $F_{\pi_t}(p)$ and $F_{\pi_{t-1}}(p)$ are the $p$-th percentile consumption levels according to beliefs $\pi_t$ and $\pi_{t-1}$, respectively. Whenever $\mu$ exhibits diminishing sensitivity to gains and there are at least three states, one-shot resolution is strictly suboptimal. This result does not require any assumption about loss aversion or diminishing sensitivity in losses.
Proposition 5. In Kőszegi and Rabin (2009)’s percentile-based news-utility model, provided the gain-loss utility function $\mu$ satisfies $\mu''(x) < 0$ for all $x > 0$, one-shot resolution is strictly suboptimal for any $T$ and any $K \geq 3$.

Similar to the idea behind Proposition 4, the proof of Proposition 5 constructs an information structure to gradually deliver the good news that the state is the best one possible. Proposition 5 requires at least three distinct consumption levels, $K \geq 3$. In a binary-states world, the percentile-based news-utility function $N$ only depends on the value of $\mu$ at two non-zero points. Thus every increasing $\mu$ is behaviorally indistinguishable from a two-part linear one, meaning the percentile-based model cannot capture diminishing sensitivity in a setting with binary states.

4.4 Gradual Good News and Gradual Bad News

For the remainder of the paper, we focus on mean-based news-utility functions to study additional implications of diminishing sensitivity. Two classes of information structures will play important roles in the sequel. To define them, we write $v_t := E_{\theta \sim \pi_t}[v(c_\theta)]$ for the expected future consumption utility based on the receiver’s (random) belief $\pi_t$ at the end of period $t$. Partition states into two subsets, $\Theta = \Theta_B \cup \Theta_G$, where $v(c_\theta) < v_0$ for $\theta \in \Theta_B$ and $v(c_\theta) \geq v_0$ for $\theta \in \Theta_G$. Interpret $\Theta_B$ as the “bad” states and $\Theta_G$ as the “good” ones.

Definition 2. An information structure $(M, \sigma)$ features gradual good news, one-shot bad news if

- $\mathbb{P}_{(M, \sigma)}[v_t \geq v_{t-1} \text{ for all } 1 \leq t \leq T \mid \theta \in \Theta_G] = 1$ and
- $\mathbb{P}_{(M, \sigma)}[v_t < v_{t-1} \text{ for no more than one } 1 \leq t \leq T \mid \theta \in \Theta_B] = 1$.

An information structure $(M, \sigma)$ features gradual bad news, one-shot good news if

- $\mathbb{P}_{(M, \sigma)}[v_t \leq v_{t-1} \text{ for all } 1 \leq t \leq T \mid \theta \in \Theta_B] = 1$ and
- $\mathbb{P}_{(M, \sigma)}[v_t > v_{t-1} \text{ for no more than one } 1 \leq t \leq T \mid \theta \in \Theta_G] = 1$.

In the first class of information structures (“gradual good news, one-shot bad news”), the sender relays good news over time and gradually increases the receiver’s expectation of future consumption. When the state is bad, the sender concentrates all the bad news in one period. The “one-shot bad news” terminology comes from noting that when $\theta \in \Theta_B$, the single period $t$ where $v_t < v_{t-1}$ must satisfy $v_t = v(c_\theta)$ and $v_{t'} = v_t$ for all $t' > t$. The receiver gets negative information about his future consumption level for the first time in period $t$, and
his expectation stays constant thereafter. On the other hand, we use the phrase “gradual bad news, one-shot good news” to refer to the “opposite” kind of information structure.

One-shot resolution falls into both of these classes. To rule out this triviality, we say that an information structure features strictly gradual good news if

$$P_{(M,\sigma)}[v_t > v_{t-1} \text{ and } v_{t'} > v_{t'-1} \text{ for two distinct } 1 \leq t, t' \leq T | \theta \in \Theta_G] > 0.$$ 

That is, there is positive probability that the receiver’s expectation strictly increases at least twice in periods 1 through $T$. Similarly define strictly gradual bad news.

We now prove that whenever $\mu$ satisfies diminishing sensitivity and (weak) loss aversion, information structures featuring strictly gradual bad news, one-shot good news are strictly worse than one-shot resolution. This conclusion also applies to symmetric $\mu$ with diminishing sensitivity but not loss aversion, and holds for any state space $\Theta$, horizon $T$, and prior $\pi_0$.

**Proposition 6.** Suppose $\mu$ satisfies diminishing sensitivity and weak loss aversion. Any information structure featuring strictly gradual bad news, one-shot good news is strictly worse than one-shot resolution in expectation, and almost surely weakly worse ex-post.

Proposition 6 identifies a class of information structures that are worse than one-shot resolution for news utility with diminishing sensitivity, distinguishing it from other models of information preference where one-shot resolution is the worst possible information structure. Utility models that make this other prediction include suspense and surprise (Ely, Frankel, and Kamenica, 2015) and news utility with a two-part linear, gain-loving (instead of loss-averse) value function (Chapman, Snowberg, Wang, and Camerer, 2019; Goette, Graeber, Kellogg, and Sprenger, 2020).

Under some additional restrictions, the optimal information structure exhibits strictly gradual good news, one-shot bad news. For the rest of the paper, we specialize to the case of $K = 2$. The next result presents a necessary and sufficient condition for inconclusive bad news to be suboptimal when $T = 2$. We then verify the condition for quadratic news utility.

**Proposition 7.** For $T = 2$, information structures with $P_{(M,\sigma)}[\pi_1 < \pi_0 \text{ and } \pi_1 \neq 0] > 0$ are strictly suboptimal if and only if there exists some $q \geq \pi_0$ so that the chord connecting $(0, U_1(0 | \pi_0))$ and $(q, U_1(q | \pi_0))$ lies strictly above $U_1(p | \pi_0)$ for all $p \in (0, \pi_0)$.

**Corollary 3.** Quadratic news utility satisfies the condition of Proposition 7.

In particular, combining Corollaries 1 and 3, we infer that any optimal information structure for a receiver with quadratic news utility satisfying $\alpha_n - \alpha_p < \beta_n + \beta_p$ with $T = 2$ must feature strictly gradual good news, one-shot bad news. Furthermore, since there exists
an optimal information structure with binary messages by Proposition 3, in this environment there is an optimal information structure where the sender induces either belief 0 or belief $p_H > \pi_0$ in the only period of communication. The next subsection characterizes $p_H$ as a function of the parameters of quadratic news utility.

In summary, we have established a ranking between three kinds of information structures that point to a preference over the direction of news skewness. For any time horizon and any state space, provided the condition in Proposition 4 holds and $\mu$ satisfies diminishing sensitivity and weak loss aversion, some information structure featuring gradual good news, one-shot bad news gives more news utility than one-shot resolution, which in turn gives more news utility than any information structure featuring strictly gradual bad news, one-shot good news. Further, under the additional restrictions in Proposition 7, a gradual good news, one-shot bad news information structure is optimal among all information structures.

4.5 Explicit Solution with Quadratic News Utility

We illustrate Proposition 3’s concavification procedure by finding in closed-form the optimal information structure when the receiver has a quadratic news-utility function.

Suppose the parameters of $\mu$ satisfy $\alpha_n - \alpha_p < \beta_n + \beta_p$ in a $T = 2$ environment. From the arguments in Section 4.4, there is an optimal information structure induces either $\pi_1 = 0$ or $\pi_1 = p_H$ for some $p_H > \pi_0$. Proposition 3 implies $(\text{cav} U_1(\cdot | \pi_0))(x) > U_1(x | \pi_0)$ for all $x \in (0, p_H)$. The geometry of concavification shows the derivative of the value function at $p_H$, $\frac{d}{dx} U_1(x | \pi_0)(p_H)$, equals the slope of the chord from 0 to $p_H$ on the function $U_1(\cdot | \pi_0)$. We use this equality to derive $p_H$ as the solution to a cubic polynomial.

Proposition 8. For $T = 2$ and quadratic news utility satisfying $\alpha_n - \alpha_p < \beta_n + \beta_p$, the optimal partial good news $p_H > \pi_0$ satisfies

$$\pi_0(\alpha_n - \alpha_p) - (\beta_p + \beta_n)\pi_0^2 = p_H^2(\alpha_n - \alpha_p + \beta_n + \beta_p) - p_H^3(2\beta_p + 2\beta_n).$$

Let $c := \frac{\alpha_n - \alpha_p}{\beta_n + \beta_p}$. We have $\frac{dp_H}{dc} > 0$. Also, we have $\frac{dp_H}{d\pi_0} < 0$ when $\pi_0 < \frac{1}{2}c$, and $\frac{dp_H}{d\pi_0} > 0$ when $\pi_0 > \frac{1}{2}c$.

There is a tension between loss aversion near the reference point (captured by $\alpha_n - \alpha_p$) and diminishing sensitivity (captured by $\beta_n + \beta_p$) in shaping the optimal information structure. Fixing the prior belief, the optimal amount of partial good news is increasing in loss aversion but decreasing in diminishing sensitivity. To understand these comparative statics, recall that in the bad state the receiver will sometimes experience false hope as he gets interim good news. The sender chooses between the receiver getting (i) a larger piece of false hope
with lower probability, or (ii) a smaller piece of false hope with higher probability. When there is more loss aversion near the reference point, belief paths that feature a small piece of good news followed by a small piece of bad news become much more costly, so (i) is preferred. When there is more diminishing sensitivity, the utility gap between the positive components of (i) and (ii) narrows, so (ii) becomes more favorable.

The optimal partial good news is non-monotonic in the prior belief when $\mu$ exhibits loss aversion — in that case, $p_H$ decreases with the prior when the prior is low, but increases with the prior when it is high. Figure 3 illustrates. The intuition is that the sender faces different incentives in maximizing the receiver’s news utility conditional on $\omega = G$ and $\omega = B$. Conditional on $\omega = G$, the optimal interim good news is $\frac{1}{2}(1 + \pi_0)$, which exploits diminishing sensitivity by splitting the good news evenly across two periods. Conditional on $\omega = B$, the distortion from loss aversion discussed before pushes the sender towards sending a bigger piece of interim good news (with lower probability). For $\pi_0$ near 0, the receiver’s expected welfare is essentially determined by his welfare in the bad state, so the latter incentive dominates and $p_H$ is far above 0.5. As $\pi_0$ increases, the relative weight on the good state’s welfare increases, so $p_H$ converges to $\frac{1}{2}(1 + \pi_0)$. In the case of $\alpha_n = \alpha_p$, the distortion from loss aversion is absent, so we get $\frac{dp_H}{d\pi_0} > 0$ for any $\pi_0 \in (0, 1)$.

Figure 3: Left: Optimal partial good news with $T = 2$, prior $\pi_0 = 0.1$, quadratic news utility parameters $\alpha_p = 2$, $\beta_p = 1$, $\beta_n = 1$, as a function of $\alpha_n$. The optimal $p_H$ monotonically increases with the amount of loss aversion near 0. Right: Optimal partial good news with $T = 2$, quadratic news utility parameters $\alpha_p = 2$, $\alpha_n = 3$, $\beta_p = 1$, $\beta_n = 1$, as a function of the prior belief. The dashed blue line shows $\pi_0 \mapsto \frac{1}{2}(\pi_0 + 1)$, the midpoint between the prior and 1. The optimal partial good news is decreasing in the prior before $\pi_0 = 0.25$, and increasing afterwards.

5 Diminishing Sensitivity and the Credibility Problem

Section 4 studied the optimal disclosure of news when the sender has commitment power. When the commitment assumption is dropped, some information structures that improve on one-shot resolution cannot be implemented in the equilibrium of the cheap-talk game.
If the sender wishes to gradually reveal the good state to a Bayesian receiver over multiple periods, then she must also sometimes provide false hope in the bad state due to the cross-state constraints on beliefs. But without commitment, the benevolent sender may strictly prefer giving false hope over telling the truth in the bad state. This deviation improves the total news utility of a receiver with diminishing sensitivity, if the positive utility from today’s good news outweighs the additional future disappointment from higher expectations. In fact, when news utility exhibits diminishing sensitivity and low enough loss aversion, the above credibility problem is so severe that every equilibrium is payoff-equivalent to the babbling equilibrium. This lack-of-commitment problem is a unique implication of diminishing sensitivity — without it, the commitment solution and the equilibrium outcome coincide.

Sufficiently strong loss aversion can restore the equilibrium credibility of good-news messages. We show that the highest equilibrium payoff when the sender lacks commitment may be non-monotonic in the extent of loss aversion, in contrast to the conclusion that more loss-averse receivers are always strictly worse off when the sender has commitment power. We also completely characterize the class of equilibria that feature (a deterministic sequence of) gradual good news in the good state and study the equilibrium rate of learning. With the quadratic or the square-root news-utility function, the sender always releases progressively larger pieces of good news over time, so the receiver’s belief in the good state grows at an increasing rate.

5.1 Equilibrium Analysis When the Sender Lacks Commitment

We continue to maintain that state space $\Theta = \{G, B\}$ is binary. To study the case where the sender lacks commitment, we analyze the perfect-Bayesian equilibria of the cheap talk game between the two parties. Formally, the equilibrium concept is as follows.

**Definition 3.** Let a finite set of messages $M$ be fixed. A *perfect-Bayesian equilibrium* consists of sender’s strategy $\sigma^* = (\sigma_t^*)_{t=1}^{T-1}$ together with receiver’s beliefs $p^*: \bigcup_{t=0}^{T-1} H^t \rightarrow [0,1]$, where:

- For every $1 \leq t \leq T - 1$, $h_{t-1}^t \in H^{t-1}$ and $\theta \in \{G, B\}$, $\sigma^*$ maximizes the receiver’s total expected news utility in periods $t, ..., T - 1, T$ conditional on having reached the public history $h_{t-1}^t$ in state $\theta$ at the start of period $t$.

- $p^*$ is derived by applying the Bayes’ rule to $\sigma^*$ whenever possible.

We make two belief-refinement restrictions:

- If $t \leq T - 1$, $h^t$ is a continuation history of $h_t$, and $p^*(h^t) \in \{0, 1\}$, then $p^*(h^t) = p^*(h_t^t)$.  

20
• The receiver’s belief in period $T$ when state is $\theta$ satisfies $\pi_T = 1_\theta$, regardless of the preceding history $h^{T-1} \in H^{T-1}$.

We will abbreviate a perfect-Bayesian equilibrium satisfying our belief refinements as an “equilibrium.” Our definition requires that once the receiver updates his belief to 0 or 1, this belief stays constant through the end of period $T - 1$. In other words, the support of his belief is non-expanding through the penultimate period.\(^7\) In period $T$, the receiver updates his belief to reflect full confidence in the true state of the world, regardless of his (possibly dogmatically wrong) belief at the end of period $T - 1$.

The receiver derives news utility in periods $1 \leq t \leq T$ based on changes in his belief, as in the model with commitment. The sender is benevolent, with the total news utility of the receiver as her objective. This implies the sender expects different equilibrium payoffs from sending the same sequence of messages in different states, as the state determines $\pi_T$ and hence the receiver’s news utility in the final period.\(^8\)

Let $V_{\mu,M,T}(\pi_0) \subseteq \mathbb{R}$ denote the set of equilibrium payoffs with news-utility function $\mu$, message space $M$, time horizon $T$, and prior $\pi_0$. Clearly, $V_{\mu,M,T}(\pi_0)$ is non-empty. There is always the babbling equilibrium, where the sender mixes over all messages uniformly in both states and the receiver’s belief never updates from the prior belief until period $T$. Denote the babbling equilibrium payoff by

$$V^{\text{Bab}}_\mu(\pi_0) := \pi_0 \mu(1 - \pi_0) + (1 - \pi_0) \mu(-\pi_0)$$

and note it is independent of $M$ and $T$.

We state two preliminary properties of the equilibrium payoffs set $V_{\mu,M,T}(\pi_0)$.

**Lemma 1.** We have:

1. For any finite $M$, $V_{\mu,M,T}(\pi_0) \subseteq V_{\mu,\{g,b\},T}(\pi_0)$

2. If $T \leq T'$, then $V_{\mu,M,T}(\pi_0) \subseteq V_{\mu,M,T'}(\pi_0)$.

The first statement says any equilibrium payoff achievable with an arbitrary finite message space is also achievable with a binary message space. The second statement says the set of equilibrium payoffs weakly expands with the time horizon.

\(^7\)This standard refinement was first used in Grossman and Perry (1986). It rules out pathological off-path belief updates if the sender deviates and sends a message perfectly indicative of one state following a history where the receiver is fully convinced of the other state.

\(^8\)In particular, this is not a cheap-talk game with state-independent sender payoffs, as in Lipnowski and Ravid (2020).
5.2 The Credibility Problem and Babbling

To understand the source of the credibility problem, let \( N_B(x; \pi) := \mu(x - \pi) + \mu(-x) \) denote the total amount of news utility across two periods when the receiver updates his belief from \( \pi \) to \( x > \pi \) today and updates it from \( x \) to 0 tomorrow. Suppose there exists a period \( T - 2 \) public history \( h^{T-2} \in H^{T-2} \) with \( p^*(h^{T-2}) = \pi \) and some \( x > \pi \) satisfying \( N_B(x; \pi) > N_B(0; \pi) \). Then, the sender strictly prefers to induce belief \( x \) rather than belief 0 after arriving at the history \( h^{T-2} \) in the bad state. A good-news message \( m_x \) inducing belief \( x \) and a bad-news message \( m_0 \) inducing belief 0 cannot both be on-path following \( h^{T-2} \), else the sender would strictly prefer to send \( m_x \) with probability 1 in the bad state.

Yet, the inequality \( N_B(0; \pi) < N_B(x; \pi) \) automatically holds for any \( x > \pi \), provided \( \mu \) is strictly concave in the positive region and symmetric around 0.

Lemma 2. If \( \mu \) is symmetric around 0 and \( \mu''(x) < 0 \) for all \( x > 0 \), then for any \( 0 < \pi < x < 1 \) it holds \( N_B(0; \pi) < N_B(x; \pi) \).

The intuition is that when the state is bad, the sender knows the receiver will inevitably get conclusive bad news in period \( T \). Giving false hope in period \( T - 1 \) (i.e., inducing belief \( x > \pi \) instead of 0) provides positive news utility at the cost of greater disappointment in the final period. Diminishing sensitivity limits the incremental cost of this additional disappointment.

The credibility problem implies that the babbling payoff is the unique equilibrium payoff.

Proposition 9. Suppose \( \mu \) is symmetric around 0 and \( \mu''(x) < 0 \) for all \( x > 0 \). For any \( M, T, \pi_0 \), the only equilibrium payoff is the babbling payoff, \( V_{\mu,M,T}(\pi_0) = \{V_{\mu}^{Bab}(\pi_0)\} \).

The babbling equilibrium is unique up to payoffs, even though the players share the same payoff function. In a cheap-talk setting with instrumental information and anticipatory utility, Kőszegi (2006) shows that a benevolent sender also distorts equilibrium communication relative to the commitment benchmark. The breakdown in communication is more complete in our setting, for the players get the same payoffs as when communication is impossible.

To understand why communication fails despite an apparent lack of conflicting interests, observe that the interaction has the same set of equilibria as the following non-psychological auxiliary game: in every period \( t \geq 0 \), the receiver plays some action \( a_t \in [0, 1] \). At the end of the game, the receiver gets \( \sum_{t=0}^{T} -(a_t - 1_{\{\theta = G\}})^2 \) — that is, the receiver is incentivized to play the action that corresponds to the Bayesian probability of the state being good in every period, and gets no other sources of payoffs. The sender gets utility \( \mu(a_t - a_{t-1}) \) in period \( t \geq 1 \). In this equivalent model where the receiver takes actions, we see that the players
have different objective functions over these actions, so we should expect some distortion of communication in all equilibria.

We now explore what happens when $\mu$ is asymmetric around 0 due to loss aversion. Say $\mu$ exhibits greater sensitivity to losses if $\mu'(x) \leq \mu'(-x)$ for all $x > 0$. We first establish a robustness check to Proposition 9 within this class of news-utility functions: when loss aversion is sufficiently weak relative to diminishing sensitivity in a $T = 2$ model, the babbling equilibrium remains unique up to payoffs.

**Proposition 10.** Suppose $\mu$ exhibits greater sensitivity to losses. If $\min_{z \in [0,1-\pi_0]} \frac{\mu'(z)}{\mu'(-\pi_0 + z)} > 1$, then $V_{\mu,M,2}(\pi_0) = \{V_{\mu}^{Bab}(\pi_0)\}$ for any $M$.

When $\mu$ is symmetric and does not exhibit strict loss aversion, diminishing sensitivity implies $\mu'(-\pi_0 + z)) = \mu'(\pi_0 + z) < \mu'(z)$ for every $z \in [0,1-\pi_0]$, so the inequality condition in Proposition 10 is always satisfied. This condition continues to hold if $\mu$ is slightly asymmetric due to a “small enough” amount of loss aversion relative to the size of the sensitivity gap $\mu'(z) - \mu'(\pi_0 + z)$. This interpretation is clearest for the $\lambda$-scaled news-utility functions, as formalized in the following corollary.

**Corollary 4.** Suppose for some $\tilde{\mu}_{pos} : [0,1] \to \mathbb{R}_+$ and $\lambda \geq 1$, the news-utility function $\mu$ satisfies $\mu(x) = \tilde{\mu}_{pos}(x)$, $\mu(-x) = -\lambda \tilde{\mu}_{pos}(x)$ for all $x \geq 0$. Provided $\lambda < \min_{z \in [0,1-\pi_0]} \frac{\tilde{\mu}'_{pos}(z)}{\tilde{\mu}'_{pos}(\pi_0 + z)}$, $V_{\mu,M,2}(\pi_0) = \{V_{\mu}^{Bab}(\pi_0)\}$ for any $M$.

When $\mu$ is strictly concave in the positive region, Corollary 4 gives a non-degenerate interval of loss-aversion parameters for which the conclusion of Proposition 9 extends in a $T = 2$ setting. If $\tilde{\mu}_{pos}$ contains more curvature, then $\tilde{\mu}'_{pos}(z)/\tilde{\mu}'_{pos}(\pi_0 + z)$ becomes larger and the interval of permissible $\lambda$’s expands.

What happens when loss aversion is high? The next proposition says a new equilibrium that payoff-dominates the babbling one exists for large $\lambda$, provided the marginal utility of an infinitesimally small piece of good news is infinite — as in the power-function specification.

**Proposition 11.** Fix $\tilde{\mu}_{pos} : [0,1] \to \mathbb{R}_+$ strictly increasing and concave, continuously differentiable at $x > 0$, $\tilde{\mu}_{pos}(0) = 0$, and $\lim_{x \to 0} \tilde{\mu}'_{pos}(x) = \infty$. Consider the family $\lambda$-indexed news-utility functions $\mu(x) = \tilde{\mu}_{pos}(x)$, $\mu(-x) = -\lambda \tilde{\mu}_{pos}(x)$ for $x \geq 0$. For each $\pi_0 \in (0,1)$, there exists $\tilde{\lambda} \geq 1$ so that whenever $\lambda \geq \tilde{\lambda}$ and for any $T \geq 2$, $|M| \geq 2$, there exists $V \in V_{\mu,M,T}(\pi_0)$ with $V > V_{\mu}^{Bab}(\pi_0)$.

To help illustrate these results, suppose $\mu(x) = \sqrt{x}$ for $x \geq 0$, $\mu(x) = -\lambda \sqrt{-x}$ for $x < 0$, $T = 2$, and $\pi_0 = \frac{1}{2}$. Corollary 4 implies whenever $\lambda < \sqrt{2}$, the babbling equilibrium is unique up to payoffs. On the other hand, Proposition 11 says when $\lambda$ is sufficiently high, there is
Figure 4: The babbling equilibrium is essentially unique for low values of $\lambda$, but there exists an equilibrium with gradual good news for $\lambda \geq 2.414$. Due to the role of loss aversion in sustaining credible partial news, a receiver with higher loss aversion may experience higher or lower expected news utility in equilibrium than a receiver with lower loss aversion.

another equilibrium with strictly higher payoffs. In fact, a non-babbling equilibrium first appears when $\lambda = 2.414$. Figure 4 plots the highest equilibrium payoff for different values of $\lambda$. Receivers with higher $\lambda$ may enjoy higher equilibrium payoffs. The reason for this non-monotonicity is that for low values of $\lambda$, the babbling equilibrium is unique and increasing $\lambda$ decreases expected news utility linearly. When the new, non-babbling equilibrium emerges for large enough $\lambda$, the sender’s behavior in the new equilibrium depends on $\lambda$. Higher loss aversion carries two countervailing effects: first, a non-strategic effect of hurting welfare when $\theta = B$, as the receiver must eventually hear the bad news; second, an equilibrium effect of changing the relative amounts of good news in different periods conditional on $\theta = G$. Receivers with an intermediate amount of loss aversion enjoy higher expected news utility than receivers with low loss aversion, as the equilibrium effect leads to better “consumption smoothing” of good news across time. But, the non-strategic effect eventually dominates and receivers with high loss aversion experience worse payoffs than receivers with low loss aversion.

5.3 Deterministic Gradual Good News Equilibria

An equilibrium $(M, \sigma^*, p^*)$ features deterministic\(^9\) gradual good news (GGN equilibrium) if there exist a sequence of constants $p_0 \leq p_1 \leq \ldots \leq p_{T-1} \leq p_T$ with $p_0 = \pi_0$, $p_T = 1$, and the receiver always has belief $p_t$ in period $t$ when the state is good. By Bayesian beliefs, in the

\(^9\)This class of equilibria is slightly more restrictive than the gradual good news, one-shot bad news information structures from Definition 2, because the sender may not randomize between several increasing paths of beliefs in the good state.
bad state of any GGN equilibrium the sender must induce a belief of either 0 or \( p_t \) in period \( t \), as any message not inducing belief \( p_t \) is a conclusive signal of the bad state.

The class of GGN equilibria is non-empty, for it contains the babbling equilibrium where \( \pi_0 = p_0 = p_1 = \ldots = p_{T-1} < p_T = 1 \). The number of intermediate beliefs in a GGN equilibrium is the number of distinct beliefs in the open interval \((\pi_0, 1)\) along the sequence \( p_0, p_1, \ldots, p_{T-1} \). The babbling equilibrium has zero intermediate beliefs.

The next proposition characterizes the set of all GGN equilibria with at least one intermediate belief.

**Proposition 12.** Let \( P^*(\pi) \subseteq (\pi, 1] \) be those beliefs \( x \) satisfying \( N_B(x; \pi) = N_B(0; \pi) \). Suppose \( \mu \) exhibits diminishing sensitivity and loss aversion. For \( 1 \leq J \leq T-1 \), there exists a gradual good news equilibrium with the \( J \) intermediate beliefs \( q^{(1)} < \ldots < q^{(J)} \) if and only if \( q^{(j)} \in P^*(q^{(j-1)}) \) for every \( j = 1, \ldots, J \), where \( q^{(0)} := \pi_0 \).

To interpret, \( P^*(\pi) \) contains the set of beliefs \( x > \pi \) such that the sender is indifferent between inducing the two belief paths \( \pi \rightarrow x \rightarrow 0 \) and \( \pi \rightarrow 0 \). Recall that when \( \mu \) is symmetric, Lemma 2 implies this indifference condition is never satisfied, which is the source of the credibility problem for good-news messages. The same indifference condition pins down the relationship between successive intermediate beliefs in GGN equilibria.

We illustrate this result with the quadratic news utility.

**Corollary 5.** 1) With quadratic news utility, \( P^*(\pi) = \left\{ \pi \cdot \frac{\beta_p+\beta_n}{\beta_p-\beta_n} - \frac{\alpha_n-\alpha_p}{\beta_p-\beta_n} \right\} \cap (\pi, 1) \).

2a) If \( \beta_n > \beta_p \), there cannot exist any gradual good news equilibrium with more than one intermediate belief.

2b) If \( \beta_n < \beta_p \), there can exist gradual good news equilibria with more than one intermediate belief. For a given set of parameters of the quadratic news-utility function and prior \( \pi_0 \), there exists a uniform bound on the number of intermediate beliefs that can be sustained in equilibrium across all \( T \).

3) In any GGN equilibrium with quadratic news utility, intermediate beliefs in the good state grow at an increasing rate.

Combined with Proposition 12, part 1) of this corollary says that in every GGN equilibrium, the successive intermediate beliefs are related by the linear map \( x \mapsto x \cdot \frac{\beta_p+\beta_n}{\beta_p-\beta_n} - \frac{\alpha_n-\alpha_p}{\beta_p-\beta_n} \). When \( \beta_n > \beta_p \), this map has a negative slope, so there cannot exist any GGN equilibrium with more than one intermediate belief. When \( \beta_p > \beta_n \), this map has a slope strictly larger than 1. As a result, after eliminating periods where no informative signal is released, every GGN equilibrium releases progressively larger pieces of good news in the good state, \( q^{(j+1)} - q^{(j)} > q^{(j)} - q^{(j-1)} \). Since equilibrium beliefs in the good state grow at an increasing
Figure 5: The longest possible sequence of GGN intermediate beliefs starting with prior $\pi_0 = \frac{1}{3}$. For quadratic news utility, equilibrium GGN beliefs always increase at an increasing rate in the good state.

rate, there exists some uniform bound $\bar{J}$ on the number of intermediate beliefs depending only on the prior belief $\pi_0$ and parameters of the news-utility function.

As an illustration, consider the quadratic news utility with $\alpha_p = 2$, $\alpha_n = 2.1$, $\beta_p = 1$, and $\beta_n = 0.2$. Starting at the prior belief of $\pi_0 = \frac{1}{3}$, Figure 5 shows the longest possible sequence of intermediate beliefs in any GGN equilibrium for arbitrarily large $T$. Since the $P^*$ sets are either empty sets or singleton sets for the quadratic news utility, Figure 5 also contains all the possible beliefs in any state of any GGN equilibrium with these parameters.

The result that GGN equilibria release increasingly larger pieces of good news generalizes to other news-utility functions with diminishing sensitivity. The basic intuition is that if the sender is indifferent between providing $d$ amount of false hope and truth-telling in the bad state when the receiver has prior belief $\pi_L$ (i.e., $\pi_L + d \in P^*(\pi_L)$), then she strictly prefers providing the same amount of false hope over truth-telling at any more optimistic prior belief $\pi_H > \pi_L$. The false hope generates the same positive news utility in both cases, but an extra $d$ units of disappointment matters less when added a baseline disappointment level of $\pi_H$ rather than $\pi_L$, thanks to diminishing sensitivity.

The next proposition formalizes this idea. It shows that when diminishing sensitivity is combined with a pair of regularity conditions, intermediate beliefs grow at an increasing rate in any GGN equilibrium.

**Proposition 13.** Suppose $\mu$ exhibits diminishing sensitivity, $|P^*(\pi)| \leq 1$ and $\frac{\partial}{\partial x} N_B(x; \pi)|_{x=\pi} > 0$ for all $\pi \in (0, 1)$. Then, in any GGN equilibrium with intermediate beliefs $q^{(1)} < ... < q^{(J)}$, we get $q^{(j)} - q^{(j-1)} < q^{(j+1)} - q^{(j)}$ for all $1 \leq j \leq J - 1$. 

26
The first regularity condition requires that the sender is indifferent between the belief paths \( \pi \rightarrow x \rightarrow 0 \) and \( \pi \rightarrow 0 \) for at most one \( x > \pi \). It is a technical assumption that lets us prove our result, but we suspect the conclusion also holds under some relaxed conditions. The second regularity condition says in the bad state, the total news utility associated with an \( \epsilon \) amount of false hope is higher than truth-telling for small \( \epsilon \). These conditions are satisfied by the power-function news utility with \( \alpha = \beta \), for example.

**Corollary 6.** In any GGN equilibrium with power-function news utility with \( \alpha = \beta \) and any \( \lambda \geq 1 \), intermediate beliefs in the good state grow at an increasing rate.

## 6 Related Literature and Predictions of Other Belief-Based Utility Models

### 6.1 Related Work on New Utility

Since Kőszegi and Rabin (2009), several other authors have analyzed the implications of news utility in different settings: asset pricing (Pagel, 2016), life-cycle consumption (Pagel, 2017), portfolio choice (Pagel, 2018), and mechanism design (Duraj, 2019). These papers focus on Bayesian agents with two-part linear gain-loss utilities and do not study the role of diminishing sensitivity to news.

Our model of diminishing sensitivity over the magnitude of news shares the same psychological motivation as Kahneman and Tversky (1979), who base their theory of human responses to monetary gains and losses on Weber’s law and psychological experiments about human responses to changes in physical attributes like temperature or brightness. Interpreting monetary gains and losses as news about future consumption, experiments that show risk-seeking behavior when choosing between loss lotteries and risk-averse behavior when choosing between gain lotteries provide evidence for diminishing sensitivity over consumption news (see e.g., Rabin and Weizsäcker (2009)). In the same vein, papers in the finance literature that use diminishing sensitivity over monetary gains and losses to explain the disposition effect (Shefrin and Statman, 1985; Kyle, Ou-Yang, and Xiong, 2006; Barberis and Xiong, 2012; Henderson, 2012) also provide indirect evidence for diminishing sensitivity over consumption news.

We are not aware of existing work that focuses on how diminishing sensitivity matters for information design with news utility. In fact, except for the work on disposition effect in finance, very few papers deal with diminishing sensitivity in *any* kind of reference-dependent preference. One exception is Bowman, Minehart, and Rabin (1999), who study
a consumption-based reference-dependent model with diminishing sensitivity. A critical difference is that their reference points are based on past habits, not rational expectations. In their environment, a consumer who knows their future income optimally concentrates all consumption losses in the first period if income will be low, but spreads out consumption gains across multiple periods if income will be high. As discussed in Remark 1, the analog of this strategy cannot be implemented in our setting since the receiver derives news utility from changes in rational Bayesian beliefs.

While some of our results apply to Kőszegi and Rabin (2009)'s model of news utility or to a more general class of such models (e.g., Proposition 3, Proposition 4, Proposition 5), we mostly focus on the simplest model of news utility where the agent derives gain-loss utility from changes in expected future consumption utility. This mean-based model lets us concentrate on the implications of diminishing sensitivity, but differs from Kőszegi and Rabin (2009)'s model where agents make a percentile-by-percentile comparison between old and new beliefs. Fully characterizing the optimal information structure using this percentile-based model is out of reach for us, but our numerical simulations in Online Appendix OA 3.2.2 suggest the answers would be very similar.

6.2 Predictions of Other Belief-Based Utility Models

In general, papers on belief-based utility have highlighted two sources of felicity: levels of belief about future consumption utility ("anticipatory utility," e.g., Kőszegi (2006); Eliaz and Spiegler (2006); Schweizer and Szech (2018)) and changes in belief about future consumption utility ("news utility" and "suspense and surprise" (Ely, Frankel, and Kamenica, 2015)). For the latter, some function of both the prior belief and the posterior belief serves as the carrier of utility, while a given posterior belief brings the same anticipatory utility for all priors (Eliaz and Spiegler, 2006). The rich dynamics of the optimal information structure under news utility with diminishing sensitivity contrast against more stark predictions of the other commonly used models of belief-based utility in the behavioral literature.

6.2.1 News Utility without Diminishing Sensitivity

The literature on reference-dependent preferences and news utility has focused on two-part linear gain-loss utility functions, which violate diminishing sensitivity. If \( \mu \) is two-part linear with loss aversion, then it follows from the martingale property of Bayesian beliefs that one-shot resolution is weakly optimal for the sender among all information structures. If there is strict loss aversion, then one-shot resolution does strictly better than any information structure that resolves uncertainty gradually. As our results have shown, more nuanced
information structures emerge as optimal when the receiver exhibits diminishing sensitivity.

6.2.2 Anticipatory Utility

In our setup, a receiver who experiences anticipatory utility gets $A(\sum_{\theta \in \Theta} \pi_t(\theta) \cdot v(c_\theta))$ if she ends period $t$ with posterior belief $\pi_t \in \Delta(\Theta)$, where $A : \mathbb{R} \to \mathbb{R}$ is a strictly increasing anticipatory-utility function. When $A$ is the identity function (as in Kőszegi (2006)), the solution to the sender’s problem would be unchanged if we modified our model and let the receiver experience both anticipatory utility and news utility. This is because by the martingale property, the receiver’s ex-ante expected anticipatory utility in a given period is the same across all information structures. So, the ranking of information structures entirely depends on the news utility they generate.

For a general $A$, if the receiver only experiences anticipatory utility, not news utility, then the sender has an optimal information structure that only releases information in $t = 1$, followed by uninformative babbling in all subsequent periods (see Online Appendix OA 4.1). The rich dynamics of the optimal information structure in our news-utility model are thus absent in an anticipatory-utility model.

6.2.3 Suspense and Surprise

Ely, Frankel, and Kamenica (2015) study dynamic information design with a Bayesian receiver who derives utility from suspense or surprise. They propose and study an original utility function over belief paths where larger belief movements always bring greater felicity. By contrast, because our states are associated with different consumption consequences, changes in beliefs may increase or decrease the receiver’s utility depending on whether the news is good or bad. While one-shot resolution is suboptimal in both Ely, Frankel, and Kamenica (2015)’s problem and our problem (under some conditions), the optimal information structure differs. The optimal information structure in our problem is asymmetric, a key implication of diminishing sensitivity. Another difference is that information structures featuring gradual bad news, one-shot good news are worse than one-shot resolution in our problem, while one-shot resolution is the worst possible information structure in Ely, Frankel, and Kamenica (2015)’s problem.

Ely, Frankel, and Kamenica (2015) also discuss state-dependent versions of suspense and surprise utilities, but this extension does not embed our model. Suppose there are two states, $\Theta = \{G, B\}$, and the agent has the suspense objective $\sum_{t=0}^{T-1} u(\sum_{\theta} \alpha_\theta \cdot (\pi_{t+1}(\theta) - \pi_t(\theta))^2)$ or the surprise objective $\sum_{t=1}^{T} u(\sum_{\theta} \alpha_\theta \cdot (\pi_t(\theta) - \pi_{t-1}(\theta))^2)$, where $\alpha_G, \alpha_B > 0$ are state-dependent scaling weights. We must have $\pi_{t+1}(G) - \pi_t(G) = -(\pi_{t+1}(B) - \pi_t(B))$, so pathwise
\[(\pi_{t+1}(G) - \pi_t(G))^2 = (\pi_{t+1}(B) - \pi_t(B))^2.\] This shows that the new objectives obtained by applying two possibly different scaling weights \(\alpha_G \neq \alpha_B\) to states \(G\) and \(B\) are identical to the ones that would be obtained by applying the same scaling weight \(\alpha = \frac{\alpha_G + \alpha_B}{2}\) to both states. Due to this symmetry in preference, the optimal information structure for entertaining an agent with state-dependent suspense or surprise utility treats the two states symmetrically, in contrast to a central prediction of diminishing sensitivity in our model.

### 6.2.4 Designing Beliefs through Non-Informational Channels

Brunnermeier and Parker (2005) and Macera (2014) study the optimal design of beliefs for agents with belief-based utilities that differ from the news-utility setup we consider. Another important distinction is that we focus on the design of information: changes in the receiver’s belief derive from Bayesian updating an exogenous prior, using the information conveyed by the sender. Macera (2014) considers a non-Bayesian agent who freely chooses a path of beliefs, while knowing the actual state of the world. Brunnermeier and Parker (2005) study the “opposite” problem to ours, where the agent freely chooses a prior belief (over the sequence of state realizations) at the start of the game, then updates belief about future states through an exogenously given information structure.

### 6.3 Related Decision-Theoretic Work on Information Preference

Several paper in decision theory have studied models of preference over dynamic information structures. Dillenberger (2010) shows that preference for one-shot resolution of uncertainty is equivalent to a weakened version of independence, provided the preference satisfies recursivity. This result does not apply here because our mean-based model of news utility violates recursivity — it can be shown that a news-utility agent may strictly prefer a 0% chance of winning a prize over a 1% chance of winning it, if he will gradually learn about the outcome of the lottery and has high enough loss aversion (see Online Appendix OA 3.1).

Dillenberger and Raymond (2020) axiomatize a general class of additive belief-based preferences in the domain of two-stage lotteries, relaxing recursivity and the independence axiom. In the case of \(T = 2\), our news-utility model belongs to the class they characterize. Under this specialization, our work may be thought of as studying the information design problem, with and without commitment, using some of Dillenberger and Raymond (2020)’s additive belief-based preferences. Dillenberger and Raymond (2020) also provide high-level conditions for additive belief-based preferences to exhibit preference for one-shot resolution.

---

\(^{10}\)Dreyfuss, Heffetz, and Rabin (2019) show that the percentile-based news-utility model can also predict preference over dominated lotteries. They use this result to explain the empirical evidence of people making dominated choices in strategy-proof mechanisms.
We are able to find more interpretable and easy-to-verify conditions for the sub-optimality of one-shot resolution, working with a specific sub-class of their preferences. Gul, Natenzon, and Pesendorfer (2019) axiomatize a class of preferences over non-instrumental information called risk consumption preferences, including a novel “peak-trough” utility specification. In contrast, we study the implications diminishing sensitivity, a classical assumption from the behavioral economics literature. Our model is not a risk consumption preference (see Online Appendix OA 4.2). Finally, our work differs from the strand of decision theory literature in that we also study a cheap-talk game with an informed sender who lacks commitment power.

6.4 Related Work in Dynamic Information Design

In a setting without behavioral preferences, Li and Norman (2021) and Wu (2018) consider a group of senders with commitment power, sequentially sending signals to persuade a single receiver. The receiver takes an action after observing all signals. This action, together with the true state of the world, determines the payoffs of every player. While these authors study a dynamic environment, only the final belief of the receiver at the end of the last period matters for payoffs. Indeed, every equilibrium in their setting can be converted into a payoff-equivalent “one-step” equilibrium where the first sender sends the joint signal implied by the old equilibrium, while all subsequent senders babble uninformatively. In our setting, the distribution of the receiver’s final belief is already pinned down by the prior belief at the start of the first period. Yet, different sequences of interim beliefs cause the receiver to experience different amounts of total news utility. The stochastic process of these interim beliefs constitutes the object of design. We provide a general procedure for computing the optimal dynamic information structure in this new setting.

Lipnowski and Mathevet (2018) study a static model of information design with a psychological receiver whose welfare depends directly on posterior belief. They discuss an application to a mean-based news-utility model without diminishing sensitivity in their Appendix A, finding that either one-shot resolution or no information is optimal. We focus on the implications of diminishing sensitivity and derive specific characterizations of the optimal information structure. Our work also differs in that we study a dynamic problem and examine equilibria without commitment.

6.5 Experiments on Information Preference

A number of experimental papers have tested whether people prefer one-shot resolution by asking subjects to choose how they wish to learn about their prize for the experiment, with
one-shot information as a feasible information structure. The empirical results are mixed. After accounting for preference over the timing of resolution,\textsuperscript{11} Falk and Zimmermann (2017) and Bellemare, Krause, Kröger, and Zhang (2005) find evidence that subjects prefer one-shot resolution, while Nielsen (2020); Masatlioglu, Orhun, and Raymond (2017); Zimmermann (2014); Budescu and Fischer (2001) find evidence against it. News utility with diminishing sensitivity may explain these mixed results, as it predicts one-shot resolution is neither the best nor the worst information structure, so it may or may not be chosen depending on what other information structures are feasible in a particular experiment. On the other hand, these experimental results are harder to reconcile with theories that either predict agents always choose one-shot resolution or predict agents always avoid it.

Two experiments have examined people’s preference over the skewness of news, with mixed results. Tables 10 and 11 in Nielsen (2020) report that subjects prefer negatively skewed news, as predicted by news utility with diminishing sensitivity. But, Masatlioglu, Orhun, and Raymond (2017) find that agents prefer positively skewed news. In showing that a classical assumption of reference dependence leads to a prediction about preference over news skewness, we hope to stimulate further empirical work on this topic.

In the concluding discussion below, we propose an experimental test of news utility with diminishing sensitivity, based on Proposition 1’s prediction about within-subject variations in the preference for one-shot resolution across treatments that associate prizes to states of the world in different ways.

7 Concluding Discussion

In this work, we have studied how an informed sender optimally communicates with a receiver who derives diminishingly sensitive gain-loss utilities from changes in beliefs. If we think that diminishing sensitivity to the magnitude of news is psychologically realistic in this domain, then the stark predictions of the ubiquitous two-part linear models may be misleading. In the presence of diminishing sensitivity, richer informational preferences emerge.

An agent’s consumption preference over the states can determine his preference between an information structure that delivers news gradually and another that results in one-shot resolution. When all information structures are feasible, one-shot resolution is neither the best one nor the worst one — skewness matters. One-shot resolution is strictly better than information structures with strictly gradual bad news, one-shot good news. But, it

\textsuperscript{11}Information structures that reveal the prize gradually will resolve uncertainty earlier than a one-shot information structure that reveals the prize at the end of the experiment, but later than a one-shot information structure that reveals the prize immediately.
is strictly worse than the optimal information structure, which is asymmetric and features strictly gradual good news, one-shot bad news (under the conditions we identified).

If the sender lacks commitment power, diminishing sensitivity leads to novel credibility problems that inhibit any meaningful communication when the receiver has no loss aversion. High enough loss aversion can restore the equilibrium credibility of good-news messages, and the receiver’s equilibrium welfare may be non-monotonic in loss aversion. We construct a family of non-babbling equilibria when loss aversion is high enough, finding that the sender must communicate increasingly larger pieces of good news over time in the good state.

We have considered news utility as a subclass of reference-dependent preferences, exploring the implications of diminishing sensitivity for informational preference. We have abstracted away from probability weighting, another aspect of prospect theory that is sometimes applied together with reference dependence. This is partly to focus on the under-explored assumption of diminishing sensitivity, and partly because probability weighting seems to face more foundational issues than the reference-dependent aspect of prospect theory. In Kahneman and Tversky (1979)’s original formulation of nonlinear probability weighting, breaking up a positive outcome into two equally likely positive outcomes that slightly differ from each other leads to a discontinuous increase in the utility of the prospect, and can thus generate a violation of first-order stochastic dominance. On the other hand, Quiggin (1982)’s rank-dependent probability weighting has met recent experimental challenges (Bernheim and Sprenger, 2020).

Some of our predictions can empirically distinguish news utility with diminishing sensitivity from other models of belief-based preference over non-instrumental information, including the two-part linear news-utility model. Proposition 1, for example, suggests a laboratory experiment where a sequence of binary events determines whether a baseline state or an alternative state realizes, with the alternative state happening if and only if all of the binary events are “successful.” Consider two treatments that have the same success probabilities for the binary events, but differ in terms of whether subjects get a better prize or a worse prize in the alternative state compared with the baseline state. In each treatment, the experiment asks whether each subject wishes to gradually learn about the intermediate binary events or only learn about the final state, and elicits the subject’s willingness to pay to experience their preferred information structure rather than the other one. Diminishing sensitivity over news predicts that subjects should have a higher willingness-to-pay for one-shot resolution when consumption is lower in the alternative state than when it is higher in the alternative state, a hypothesis we plan to test in future work.
References


Appendix

The Online Appendix may be accessed at: https://kevinhe.net/papers/ds_oa.pdf

A Proofs of the Main Results

In the proofs, we will often use the following fact about news-utility functions with diminishing sensitivity. We omit its simple proof.

Fact 1. Let $d_1, d_2 > 0$ and suppose $\mu(0) = 0$.

- (sub-additivity in gains) If $\mu''(x) < 0$ for all $x > 0$, then $\mu(d_1 + d_2) < \mu(d_1) + \mu(d_2)$.
- (super-additivity in losses) If $\mu''(x) > 0$ for all $x < 0$, then $\mu(-d_1 - d_2) > \mu(-d_1) + \mu(-d_2)$

A.1 Proof of Proposition 1

Proof. Consider an agent who prefers B over A. In state A, he gets $\mu(-\rho_0)$ with one-shot information, but $\sum_{t=1}^T \mu(\rho_t - \rho_{t-1})$ with gradual information. For each $t$, $\rho_t - \rho_{t-1} < 0$, and furthermore $\sum_{t=1}^T \rho_t - \rho_{t-1} = -\rho_0$ by telescoping and using the fact that $\rho_T = 0$. Due to super-additivity in losses, we get that $\mu(-\rho_0) > \sum_{t=1}^T \mu(\rho_t - \rho_{t-1})$. In state B, he gets $\mu(1 - \rho_0)$ with one-shot information. With gradual information, let $\hat{T} \leq T$ be the first period where the coin toss comes up tails. His news utility is $\left[\sum_{t=1}^{\hat{T}-1} \mu(\rho_t - \rho_{t-1})\right] + \mu(1 - \rho_{\hat{T}-1})$ where each $\rho_t - \rho_{t-1} < 0$ for $1 \leq t \leq \hat{T} - 1$. Again by super-additivity in losses, $\sum_{t=1}^{\hat{T}-1} \mu(\rho_t - \rho_{t-1}) < \mu(\rho_{\hat{T}-1} - \rho_0)$. By sub-additivity in gains, $\mu(1 - \rho_{\hat{T}-1}) < \mu(\rho_0 - \rho_{\hat{T}-1}) + \mu(1 - \rho_0) \leq -\mu(\rho_{\hat{T}-1} - \rho_0) + \mu(1 - \rho_0)$, where the weak inequality follows since $\lambda \geq 1$. Putting these pieces together,

$$
\left[\sum_{t=1}^{\hat{T}-1} \mu(\rho_t - \rho_{t-1})\right] + \mu(1 - \rho_{\hat{T}-1}) < \mu(\rho_{\hat{T}-1} - \rho_0) - \mu(\rho_{\hat{T}-1} - \rho_0) + \mu(1 - \rho_0) = \mu(1 - \rho_0)
$$

as desired.

Now consider an agent who prefers A over B. We show that when $\lambda = 1$, the agent strictly prefers gradual information to one-shot information. By continuity of news utility in $\lambda$, the same strict preference must also hold for $\lambda$ in an open neighborhood around 1.
In state A, the agent gets $\mu(1 - \pi_0)$ with one-shot information, but $\sum_{t=1}^{T} \mu(\pi_t - \pi_{t-1})$ with gradual information. For each $t$, $\pi_t - \pi_{t-1} > 0$, and furthermore $\sum_{t=1}^{T} \pi_t - \pi_{t-1} = 1 - \pi_0$ by telescoping and using the fact that $\pi_T = 1$. Due to sub-additivity in gains, we get that $\sum_{t=1}^{T} \mu(\pi_t - \pi_{t-1}) > \mu(1 - \pi_0)$. In state B, he gets $\mu(-\pi_0)$ with one-shot information. With gradual information, let $\hat{T} \leq T$ be the first period where the $X_T = 0$. His news utility is 
$[\sum_{t=1}^{\hat{T}-1} \mu(\pi_t - \pi_{t-1})] + \mu(-\pi_{\hat{T}-1})$ where each $\pi_t - \pi_{t-1} > 0$ for $1 \leq t \leq \hat{T} - 1$. Again by sub-additivity in gains, $\sum_{t=1}^{T} \mu(\pi_t - \pi_{t-1}) > \mu(\pi_{\hat{T}-1} - \pi_0)$. By super-additivity in losses, $\mu(-\pi_{\hat{T}-1}) > \mu(-(\pi_{\hat{T}-1} - \pi_0)) + \mu(-\pi_0) = -\mu(\pi_{\hat{T}-1} - \pi_0) + \mu(-\pi_0)$, where the equality comes from the fact that $\lambda = 1$ so $\mu$ is symmetric about 0. Putting these pieces together,

$$\left[\sum_{t=1}^{\hat{T}-1} \mu(\pi_t - \pi_{t-1})\right] + \mu(-\pi_{\hat{T}-1}) > \mu(\pi_{\hat{T}-1} - \pi_0) - \mu(\pi_{\hat{T}-1} - \pi_0) + \mu(-\pi_0) = \mu(-\pi_0)$$

as desired. \hfill $\Box$

### A.2 Proof of Proposition 4

**Proof.** Suppose $T = 2$. Consider the following family of information structures, indexed by $\epsilon > 0$. Order the states based on $E_\epsilon \sim \pi_0(v(c))$ and label them $\theta_L, \theta_2, \ldots, \theta_{K-1}, \theta_H$. Let $M = \{m_L, m_2, \ldots, m_{K-1}, m_H\}$. Let $\sigma_t(\theta_k)(m_k) = 1$ for $2 \leq k \leq K - 1$, $\sigma_t(\theta_H)(m_H) = 1$, and $\sigma_t(\theta_L)(m_L) = x$, $\sigma_t(\theta_L)(m_H) = 1 - x$ for some $x \in (0,1)$ so that the posterior belief after observing $m_H$ is $(1 - \epsilon)1_H \oplus \epsilon 1_L$.

For every $\epsilon > 0$, the information structure just described leads to one-shot resolution of states $\theta \notin \{\theta_L, \theta_H\}$. The difference between its expected news utility and that of one-shot resolution is $W(\epsilon)$, given by

$$\pi_0(\theta_H) \cdot [N((1 - \epsilon)1_H \oplus \epsilon 1_L \mid \pi_0) + N(1_H \mid (1 - \epsilon)1_H \oplus \epsilon 1_L) - N(1_H \mid \pi_0)]$$

$$+ \frac{\epsilon}{1 - \epsilon} \pi_0(\theta_H) \cdot [N((1 - \epsilon)1_H \oplus \epsilon 1_L \mid \pi_0) + N(1_L \mid (1 - \epsilon)1_H \oplus \epsilon 1_L) - N(1_L \mid \pi_0)].$$

$W$ is continuously differentiable away from 0 and $W(0) = 0$. To show that $W(\epsilon) > 0$ for some $\epsilon > 0$, it suffices that $\lim_{\epsilon \to 0^+} W'(\epsilon) > 0$. Using the continuous differentiability of $N$ except when its two arguments are identical, this limit is

$$\lim_{\epsilon \to 0^+} \frac{N((1 - \epsilon)1_H \oplus \epsilon 1_L \mid \pi_0) - N(1_H \mid \pi_0)}{\epsilon} + \lim_{\epsilon \to 0^+} \frac{N(1_H \mid (1 - \epsilon)1_H \oplus \epsilon 1_L)}{\epsilon}$$

$$+ N(1_H \mid \pi_0) + N(1_L \mid 1_H) - N(1_L \mid \pi_0).$$

Simple rearrangement gives the expression from Proposition 4. The expression for the case
of mean-based $\mu$ follows by algebra, noting that $N((1 - x)1_H \oplus x1_L \mid \pi_0) = \mu((1 - x) - v_0)$ for $x \in [0, 1]$.

If $T > 2$, then note the sender’s $T$-period problem starting with prior $\pi_0$ has a value at least as large as the 2-period problem with the same prior. On the other hand, one-shot resolution brings the same total expected news utility regardless of $T$. \hfill \Box

### A.3 Proof of Proposition 6

**Proof.** We show that one-shot resolution gives weakly higher news utility conditional on each state, and strictly higher news utility conditional on at least one $\theta \in \Theta_B$.

When $\theta \in \Theta_B$, $\mathbb{P}_{(M, \sigma)}$-almost surely the expectations in different periods form a decreasing sequence $v_0 \geq v_1 \geq ... \geq v_T = v(c_0)$. By super-additivity in losses, $\sum_{t=1}^{T} \mu(v_t - v_{t-1}) \leq \mu(v_T - v_0) = \mu(v(c_0) - v_0)$. This shows $\mathbb{P}_{(M, \sigma)}$-almost surely the ex-post news utility in state $\theta$ is no larger than $\mu(v(c_0) - v_0)$, the news utility from one-shot resolution.

Let $E$ be the event where the receiver’s expectation strictly decreases two or more times. From the definition of strict gradual bad news, there exists some $\theta^* \in \Theta_B$ so that $\mathbb{P}_{(M, \sigma)}[E \mid \theta^*] > 0$. On $E \cap \{\theta^*\}$, $\sum_{t=1}^{T} \mu(v_t - v_{t-1}) < \mu(v(c_{\theta^*}) - v_0)$ from super-additivity in losses, which means the expected news utility conditional on $E \cap \{\theta^*\}$ is strictly lower than that of one-shot resolution. Combined with the fact that the ex-post news utility in state $\theta^*$ is always weakly lower than $\mu(v(c_{\theta^*}) - v_0)$, this shows expected news utility in state $\theta^*$ is strictly lower than that of one-shot resolution.

Conditional on any state $\theta \in \Theta_G$, there is some random period $t^* \in \{0, ..., T - 1\}$ so that $v_t$ is weakly decreasing up to $t = t^*$ and $v_t = v(c_0)$ for $t > t^*$. If $t^* = 0$, then this belief path yields the same news utility as one-shot resolution. If $t^* \geq 1$, then the total news utility is $\sum_{t=1}^{t^*} \mu(v_t - v_{t-1}) + \mu(v(c_0) - v_{t^*})$. By sub-additivity in gains, $\sum_{t=1}^{t^*} \mu(v_t - v_{t-1}) \leq \mu(v_{t^*} - v_0)$, and for the same reason, $\mu(v(c_0) - v_{t^*}) \leq \mu(v_0 - v_{t^*}) + \mu(v(c_0) - v_0)$ as we must have $v_{t^*} \leq v_0$. Total news utility is therefore bounded above by $\mu(v_{t^*} - v_0) + \mu(v_0 - v_{t^*}) + \mu(v(c_0) - v_0)$. By weak loss aversion, $\mu(v_{t^*} - v_0) + \mu(v_0 - v_{t^*}) \leq 0$, therefore total news utility is no larger than that of one-shot resolution, $\mu(v(c_0) - \pi_0)$. \hfill \Box

### A.4 Proof of Proposition 7

**Proof.** Suppose the condition in Proposition 7 holds. So in particular, it holds for $q = \pi_0$. Consider any information structure $(M, \sigma)$ and its induced distribution over posterior beliefs in state $G$, $\eta \in \Delta([0, 1])$. If there exists $0 < x < \pi_0$ such that $\eta(x) > 0$, then we can “split posterior $x$ into 0 and $\pi_0$”: that is, we can construct another information structure $(\tilde{M}, \tilde{\sigma})$ with induced distribution $\tilde{\eta} \in \Delta([0, 1])$, so that $\tilde{\eta}(x) = 0$, $\tilde{\eta}(0) = \eta(0) + \eta(x)(1 - \frac{x}{\pi_0})$, 39
\( \tilde{\eta}(\pi_0) = \eta(\pi_0) + \eta(x) \frac{\pi_0}{\alpha_0}. \) Information structure \((\tilde{M}, \tilde{\sigma})\) gives strictly higher news utility than \((M, \sigma)\), since the condition implies \( U_1(x \mid \pi_0) < (1 - \frac{\pi_0}{\alpha_0}) \cdot U_1(0 \mid \pi_0) + \frac{\pi_0}{\alpha_0} \cdot U_1(\pi_0 \mid \pi_0). \)

Conversely, suppose every information structure \((M, \sigma)\) with induced posterior distribution \(\eta\) such that \(\eta(x) > 0\) for some \(0 < x < \pi_0\) is strictly suboptimal. Then, there must exist an optimal information structure, \((\tilde{M}, \tilde{\sigma})\) with posterior distribution \(\tilde{\eta}\), so that \(\tilde{\eta}\) is supported on two points: \(0\) and some \(q > \pi_0\). If there exists some point \(p \in (0, \pi_0)\) that violates the condition of Proposition 7 for this \(q\), that is \(U_1(p \mid \pi_0)\) is at least as large as the height of the chord connecting \((0, U_1(0 \mid \pi_0))\) and \((q, U_1(q \mid \pi_0))\), then an information structure inducing the posterior beliefs \(p\) and \(q\) would strictly dominate the optimal information structure, which is impossible. 

\[ \Box \]

### A.5 Proof of Proposition 8

**Proof.** We have

\[
\frac{d}{dp} U(p \mid \pi_0) = 2\alpha_p - \alpha_n - \beta_p + 2\beta_p \pi_0 + p(-2\alpha_p + 2\beta_p + 2\alpha_n + 2\beta_n) + p^2(-3\beta_p - 3\beta_n)
\]

Further, \(p\) times slope of chord is:

\[
U(p \mid \pi_0) - U(0 \mid \pi_0) = U(p \mid \pi_0) - (\beta_n \pi_0^2 - \alpha_n \pi_0)
= \pi_0(-\alpha_p + \alpha_n + \pi_0^2(-\beta_p - \beta_n) + p(2\alpha_p - \alpha_n - \beta_p)
+ p^2(-\alpha_p + \beta_p + \alpha_n + \beta_n) + p^3(-\beta_p - \beta_n) + p\pi_0(2\beta_p)
\]

Equating \(p \cdot \frac{d}{dp} U(p \mid \pi_0) = U(p \mid \pi_0) - U(0 \mid \pi_0)\), we get

\[
\pi_0(\alpha_n - \alpha_p) - (\beta_p + \beta_n)\pi_0^2 = p^2(\alpha_n - \alpha_p + \beta_p + \beta_n) - p^3(2\beta_p + 2\beta_n).
\]

Define \(c = \frac{\alpha_n - \alpha_p}{\beta_n + \beta_p}\). Note that \(c \in [0, 1)\) by assumptions in the statement of the Proposition.

Corollary 3 allows us to define \(p(\pi_0, c)\) as an implicit function through \(\pi_0 c - \pi_0^2 = p^2(1 + c) - 2p^3\).

We characterize first the derivative of \(p\) w.r.t. \(\pi_0\).

We check the conditions of the implicit function theorem in our setting: define the function \(f(\pi_0, p, c) = p^2(1 + c) - 2p^3 - \pi_0 c + \pi_0^2\) with domain \((0, 1)^3\). We look at the case \(c = 0\) separately in the end. We need \(\partial_{p}f(\pi_0, p, c) \neq 0\). If this is true, then we can solve for \(p(\pi_0, c)\) locally and also calculate its derivative. We note that \(\partial_{p}f(\pi_0, p, c) = 2p(1 + c) - 6p^2\). Hence, \(\partial_{p}f(\pi_0, p, c)\) is zero if \(p = \frac{1 + c}{3} =: \hat{p} \in \left[\frac{1}{3}, \frac{3}{3}\right]\). Now, for a fixed \(c\), we see if there is a \(\pi_0\)
that would give \( \hat{p} \). This involves solving for \( \pi_0 \) in quadratic equation

\[
\pi_0^2 - \pi_0 c + \frac{1}{27}(1 + c)^3 = 0. \tag{1}
\]

The discriminant as a function of \( c \) is given as \( D(c) = c^2 - \frac{4}{27}(1 + c)^3 \). Note that \( D'(c) = \frac{2}{9}(2 - c)(2c - 1) \). In particular, \( D \) is decreasing from \( c = 0 \) to \( c = \frac{1}{2} \) and increasing from then on until \( c = 1 \). We note also that \( D(0) < 0, D(1) < 0 \) so that overall it follows that \( D(c) < 0 \) for all \( c \in [0, 1] \). In particular, it holds that Equation (1) has no solution. This means that \( \partial_p f \) never changes sign in \( (0, 1)^3 \setminus \{(p, \pi_0, c) : \pi_0 c - \pi_0^2 = p^2(1 + c) - 2p^3\} \). This also implies that \( p(\pi_0) > \frac{1+c}{3} \), for all \( c, \pi_0 \in (0, 1) \). Recall here that \( f \) is a smooth function on its domain. Thus, implicit function theorem is applicable for all \((\pi_0, c) \in (0, 1)^2\).

Totally differentiating, we get:

\[
d\pi_0 \cdot (\alpha_n - \alpha_p) - (\beta_p + \beta_n)2\pi_0 \cdot d\pi_0 = 2p \cdot dp \cdot (\alpha_n - \alpha_p + \beta_n + \beta_p) - 3p^2 \cdot dp \cdot (2\beta_p + 2\beta_n),
\]

which can be rearranged to \( \frac{dp}{d\pi_0} = \frac{c - 2\pi_0}{2p^2(1 + c) - 6p^3} \). The steps above showed that the denominator of this expression never changes sign. Given that we know it is negative at \( c = 0 \) and \( f \) is continuously differentiable, we conclude that the denominator is always negative for all \( c \) and all \( \pi_0 \in (0, 1) \). It follows that unless \( c = 0 \), \( p(\pi_0) \) is falling until the prior characterized in the statement of the Proposition and increasing afterwards.

For the case \( c = 0 \) one looks at the implicitly defined function \( p(\pi_0) \) through \( 2p^3(\pi_0) - p^2(\pi_0) = \pi_0^2 \). Solving with similar steps as above one finds, it is strictly increasing and it is strictly above \( \frac{1}{3} \) for all \( \pi_0 \in (0, 1) \).

Summarizing, the amount of optimal good news can always be solved explicitly under the condition that \( c \in [0, 1) \) and it is always strictly above \( \max\{\pi_0, \frac{1}{3}\} \). It is strictly increasing in the prior in the absence of loss aversion and otherwise U-shaped.

Next, we focus on the derivative of \( p \) w.r.t. \( c \) as defined implicitly through \( \pi_0 c - \pi_0^2 = p^2(1 + c) - 2p^3 \). We fix \( \pi_0 \in (0, 1) \) and we look at the function \( f : (0, 1) \times (0, 1) \to \mathbb{R} \) given by \( f(c, p) = p^2(1 + c) - 2p^3 - (\pi_0 c - \pi_0^2) \). The calculations above show that \( f_p \neq 0 \) for every \( (c, p) \in (0, 1)^2 \). Hence, the implicit function theorem is applicable and we can calculate

\[
p \frac{dp}{dc} = \frac{\pi_0 - p^2}{2(1 + c) - 6p}.
\]

Above we showed that the denominator of this expression is strictly negative. Next we show that \( p(c) > \sqrt{\pi_0} \) for all \( c \in (0, 1) \). Note that the function \( p \mapsto p^2(1 + c) - 2p^3 \) is strictly decreasing in \( p \) for \( p > \frac{1+c}{3} \). This can be established by looking at first order derivatives. We note also that \( \pi_0(1 + c) - 2\sqrt{\pi_0}\pi_0 > \pi_0 c - \pi_0^2 \) is equivalent to \( 1 - 2\sqrt{\pi_0} + \pi = (1 - \sqrt{\pi_0})^2 > 0 \),
which is true by virtue of $\pi_0 < 1$. This, and the strict monotonicity of $p \mapsto p^2(1 + c) - 2p^3$ for $p > \frac{1+c}{3}$ implies that $p(c) > \sqrt{\pi_0}$ for all $c, \pi_0 \in (0, 1)$. This establishes the result.

Summarizing, the amount of optimal good news is increasing in loss aversion as measured by $\alpha_n - \alpha_p$ and decreasing in the amount of diminishing sensitivity, as measured by $\beta_n + \beta_p$. \qed

### A.6 Proof of Proposition 9

We begin by giving some additional definition and notation.

For $p, \pi \in [0, 1]$, let $N_G(p; \pi) := \mu(p - \pi) + \mu(1 - p)$.

We state a preliminary lemma about $N_G$ and $N_B$.

**Lemma A.1.** Suppose $\mu$ exhibits diminishing sensitivity and greater sensitivity to losses. Then, $p \mapsto N_G(p; \pi)$ is strictly increasing on $[0, \pi]$ and symmetric on the interval $[\pi, 1]$. For each $p_1 \in [\pi, 1]$, there exists exactly one point $p_2 \in [\pi, 1]$ so that $N_G(p_1; \pi) = N_G(p_2; \pi)$. For every $p_L < \pi$ and $p_H \geq \pi$, $N_G(p_L; \pi) < N_G(p_H; \pi)$. Also, $N_B(p; \pi)$ is symmetric on the interval $[0, \pi]$. For each $p_1 \in [0, \pi]$, there exists exactly one point $p_2 \in [0, \pi]$ so that $N_B(p_1; \pi) = N_B(p_2; \pi)$.

Consider any period $T - 2$ history $h_{T-2}$ in any equilibrium $(M, \sigma^*, p^*)$ where $p^*(h_{T-2}) = \pi \in (0, 1)$. Let $P_G$ and $P_B$ represent the sets of posterior beliefs induced at the end of $T - 1$ with positive probability, in the good and bad states. The next lemma gives an exhaustive enumeration of all possible $P_G, P_B$.

**Lemma A.2.** The sets $P_G, P_B$ belong to one of the following cases.

1. $P_G = P_B = \{\pi\}$
2. $P_G = \{1\}, P_B = \{0\}$
3. $P_G = \{p_1\}$ for some $p_1 \in (\pi, 1)$ and $P_B = \{0, p_1\}$
4. $P_G = \{\pi, 1\}$ and $P_B = \{0, \pi\}$
5. $P_G = \{p_1, p_2\}$ for some $p_1 \in (\pi, \frac{1+c}{2})$, $p_2 = 1 - p_1 + \pi$, $P_B = \{0, p_1, p_2\}$.

We now give the proof of Proposition 9.

**Proof.** Consider any period $T - 2$ history $h_{T-2}$ with $p^*(h_{T-2}) \in (0, 1)$. By Lemma 2, $N_B(p; p^*(h_{T-2})) > N_B(0; p^*(h_{T-2}))$ for all $p \in (p^*(h_{T-2}), 1]$. Therefore, cases 3 and 5 are ruled out from the conclusion of Lemma A.2. This shows that after having reached history $h_{T-2}$, the receiver will get total news utility of $\mu(1 - p^*(h_{T-2}))$ in the good state and
\[ \mu(-p^*(h^{T-2})) \] in the bad state. This conclusion applies to all period \( T - 2 \) histories (including those with equilibrium beliefs 0 or 1). So, the sender gets the same utility as if the state is perfectly revealed in period \( T - 1 \) rather than \( T \), and the equilibrium up to period \( T - 1 \) form an equilibrium of the cheap talk game with horizon \( T - 1 \). By backwards induction, we see that along the equilibrium path, whenever the receiver’s belief updates, it is updated to the dogmatic belief in \( \theta \).

\[ \square \]

A.7 Proof of Proposition 10

Proof. The conclusions of Lemmas A.1 and A.2 continue to hold, since these only depend on \( \mu \) exhibiting greater sensitivity to losses. As in the proof of Proposition 9, we only need to establish \( N_B(p; \pi_0) > N_B(0; \pi_0) \) for all \( p \in (\pi_0, 1) \) to rule out cases 3 and 5 from Lemma A.2 and hence establish our result.

For \( p = \pi_0 + z \) where \( z \in (0, 1 - \pi_0) \), \( N_B(p; \pi_0) - N_B(0; \pi_0) = \mu(z) + \mu(-\pi_0 + z) - \mu(-\pi_0) \). Consider the RHS as a function \( D(z) \) of \( z \). Clearly \( D(0) = 0 \), and \( D'(z) = \mu'(z) - \mu'(-\pi_0 + z) \). Since \( \min_{z \in [0, 1 - \pi_0]} \frac{\mu'(z)}{\mu'(-\pi_0 + z)} > 1 \), we get \( D'(z) > 0 \) for all \( z \in [0, 1 - \pi_0] \), thus \( D(z) > 0 \) on the same range.

\[ \square \]

A.8 Proof of Proposition 11

Proof. By the proof of Proposition 12, which does not depend on this result, there is a GGN equilibrium with one intermediate belief \( p \in (\pi_0, 1) \) whenever \( N_B(p; \pi_0) = N_B(0; \pi_0) \). In this equilibrium, the sender induces a belief of either \( p \) or 0 by the end of period 1, then babbles in all remaining periods of communication. Since the sender is indifferent between inducing belief \( p \) or 0 in the bad state, this equilibrium gives the same payoff as the babbling one in the bad state. But, since \( \mu(p - \pi_0) + \mu(1 - p) > \mu(1 - \pi_0) \) due to strict concavity of \( \hat{\mu}_{pos} \), the receiver gets strictly higher news utility in the good state.

To find \( \tilde{\lambda} \) that guarantees the existence of a \( p \) solving \( N_B(p; \pi_0) = N_B(0; \pi_0) \), let \( D(p) := N_B(p; \pi_0) - N_B(0; \pi_0) \). We have \( D(\pi_0) = 0 \) and \( \lim_{p \to \pi_0^+} D'(p) = \lim_{x \to 0^+} \hat{\mu}'_{pos}(x) - \mu'(-\pi_0) = \lim_{x \to 0^+} \hat{\mu}'_{pos}(x) - \lambda \mu'(\pi_0) \). For any finite \( \lambda \), this limit is \( \infty \), since \( \lim_{x \to 0^+} \hat{\mu}'_{pos}(x) = \infty \). On the other hand, \( D(1) = \mu(1 - \pi_0) + \mu(-1) - \mu(-\pi_0) = \hat{\mu}_{pos}(1 - \pi_0) - \lambda(\hat{\mu}_{pos}(1) - \hat{\mu}_{pos}(\pi_0)) \). Since \( \hat{\mu}_{pos}(1) - \hat{\mu}_{pos}(\pi_0) > 0 \), we may find a large enough \( \tilde{\lambda} \geq 1 \) so that \( \hat{\mu}_{pos}(1 - \pi_0) - \tilde{\lambda}(\hat{\mu}_{pos}(1) - \hat{\mu}_{pos}(\pi_0)) < 0 \). Whenever \( \lambda \geq \tilde{\lambda} \), we therefore get \( D(\pi_0) = 0 \), \( \lim_{p \to \pi_0^+} D'(p) = \infty \), and \( D(1) < 0 \). By the intermediate value theorem applied to the continuous \( D \), there exists some \( p \in (\pi_0, 1) \) so that \( D(p) = 0 \).
A.9 Proof of Proposition 12

Proof. Let $J$ intermediate beliefs satisfying the hypotheses be given. We construct a gradual good news equilibrium where $p_t = q^{(t)}$ for $1 \leq t \leq J$, and $p_t = q^{(J)}$ for $J + 1 \leq t \leq T - 1$.

Let $M = \{g, b\}$ and consider the following strategy profile. In period $t \leq J$ 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 $\sigma(h^{t-1}; B)(b) = 1$. Finally, if the period is $t > J$, then $\sigma(h^{t-1}; G)(b) = 1$ and $\sigma(h^{t-1}; B)(b) = 1$. In terms of beliefs, suppose $h^t$ has $t \leq J$ and every message so far has been $g$. Such histories are on-path and get assigned the Bayesian posterior belief. If $h^t$ has $t \leq J$ and contains at least one $b$, then it gets assigned belief 0. Finally, if $h^t$ has $t > J$, then $h^t$ gets assigned the same belief as the subhistory constructed from its first $J$ elements. 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 $t \leq J$ 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 $g$. The receiver will get some $I$ periods of good news before the bad state is revealed, either by the sender or by nature in period $T$. That is, the equilibrium news utility with $I$ periods of good news is given by $\sum_{i=1}^{I} \mu(p_{t-1+i} - p_{t-2+i}) + \mu(-p_{t-1+i})$. Since $p_{t-1+i} \in P^*(p_{t-2+i})$, we have $N_B(p_{t-1+i}; p_{t-2+i}) = N_B(0; p_{t-2+i})$, that is to say $\mu(p_{t-1+i} - p_{t-2+i}) + \mu(-p_{t-1+i}) = \mu(-p_{t-2+i})$. We may therefore rewrite the receiver’s total news utility as $\sum_{i=1}^{I-1} \mu(p_{t-1+i} - p_{t-2+i}) + \mu(-p_{t-2+i})$. 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 $I$’s realization, 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 the receiver’s total news utility in the good state along the equilibrium path is given by $\sum_{i=1}^{J-(t-1)} \mu(p_{t-1+i} - p_{t-2+i}) + \mu(1 - p_{t-1+i})$. 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 gives the total news utility $\mu(-p_{t-1}) + \mu(1)$. We have

$$\mu(1) < \mu(1 - p_{t-1}) + \mu(p_{t-1}) \leq \mu(1 - p_{t-1}) - \mu(-p_{t-1}),$$

44
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 \( b \) or a history with length \( K \) or longer, the receiver’s belief is the same at all continuation histories. So the sender has no deviation incentives since no deviations affect future beliefs.

For the other direction, suppose by way of contradiction there exists a gradual good news equilibrium with the \( J \) intermediate beliefs \( q^{(1)} < \ldots < q^{(J)} \). For a given \( 1 \leq j \leq J \), find the smallest \( t \) such that \( p_t = q^{(k-1)} \) and \( p_{t+1} = q^{(k)} \). At every on-path history \( h^t \in H^t \) with \( p^*(h^t) = p_t \), we must have \( \sigma^*(h^t; B) \) inducing both 0 and \( q^{(j)} \) with strictly positive probability. Since we are in equilibrium, we must have \( \mu(-q^{(j-1)}) \) being equal to \( \mu(q^{(j)} - q^{(j-1)}) \) plus the continuation payoff. If \( j = J \), then this continuation payoff is \( \mu(-q^{(j)}) \) as the only other period of belief movement is in period \( T \) when the receiver learns the state is bad. If \( j < J \), then find the smallest \( \bar{t} \) so that \( p_{\bar{t}+1} = q^{(j+1)} \). At any on-path \( h^{\bar{t}} \in H^{\bar{t}} \) which is a continuation of \( h^t \), we have \( p^*(h^{\bar{t}}) = q^{(j)} \) and the receiver has not experienced any news utility in periods \( t + 2, \ldots, \bar{t} \). Also, \( \sigma^*(h^{\bar{t}}; B) \) assigns positive probability to inducing posterior belief 0, so the continuation payoff in question must be \( \mu(-q^{(j)}) \). So we have shown that \( \mu(-q^{(j-1)}) = \mu(q^{(j)} - q^{(j-1)}) + \mu(-q^{(j)}) \), that is \( N_B(q^{(j)}; q^{(j-1)}) = N_B(0; q^{(j-1)}) \).

\[ \square \]

### A.10 Proof of Proposition 13

**Proof.** Since \( N_B(p; \pi) - N_B(0; \pi) = 0 \) for \( p = \pi \) and \( \frac{\partial}{\partial p} N_B(p; \pi)|_{p=\pi} > 0 \), \( N_B(p; \pi) - N_B(0; \pi) \) starts off positive for \( p \) slightly above \( \pi \). Given that \( |P^*(\pi)| \leq 1 \), if we find some \( p' > \pi \) with \( N_B(p'; \pi) - N_B(0; \pi) > 0 \), then any solution to \( N_B(p; \pi) - N_B(0; \pi) = 0 \) in \( (\pi, 0) \) must lie to the right of \( p' \).

If \( q^{(j)} \), \( q^{(j+1)} \) are intermediate beliefs in a GGN equilibrium, then by Proposition 12, \( q^{(j)} \in P^*(q^{(j-1)}) \) and \( q^{(j+1)} \in P^*(q^{(j)}) \). Let \( p' = q^{(j)} + (q^{(j)} - q^{(j-1)}) \). Then,

\[
N_B(p'; q^{(j)}) - N_B(0; q^{(j)}) = \mu(p' - q^{(j)}) + \mu(-p') - \mu(-q^{(j)}) \\
= \mu(q^{(j)} - q^{(j-1)}) + \mu(-q^{(j)} - (q^{(j)} - q^{(j-1)})) - \mu(-q^{(j)}) \\
> \mu(q^{(j)} - q^{(j-1)}) + \mu(-q^{(j-1)} - (q^{(j)} - q^{(j-1)})) - \mu(-q^{(j-1)}),
\]

where the last inequality comes from diminishing sensitivity. But, the final expression is \( N_B(q^{(j)}; q^{(j-1)}) - N_B(0; q^{(j-1)}) \), which is 0 since \( q^{(j)} \in P^*(q^{(j-1)}) \). This shows we must have \( q^{(j+1)} - q^{(j)} > q^{(j)} - q^{(j-1)} \). \[ \square \]