Attention Capture

Andrew Koh∗  Sivakorn Sanguanmoo†
MIT  MIT

December 6, 2022

Abstract

We study the extent to which information can be used to extract attention from a decision maker (DM). All feasible stopping times—random times DM stops paying attention—are implementable by giving DM either full information at random times, or null messages which make DM increasingly uncertain. We then introduce a designer with nonlinear preferences over DM’s stopping time and characterize the set of achievable welfare outcomes. Designer-optimal structures leave DM with zero surplus and have a block structure in which DM is indifferent at the boundaries, and receives no information in the interior. Our results speak directly to the attention economy.

∗MIT Department of Economics; email: ajkoh@mit.edu
†MIT Department of Economics; email: sanguanm@mit.edu

First posted version: September 2022
We are grateful to Ian Ball, Alessandro Bonatti, Roberto Corrao, Matt Elliott, Drew Fudenberg, Giacomo Lanzani, Stephen Morris, and Ellen Muir for valuable discussions. We also thank participants at the MIT Summer Theory Lab, MIT Theory Lunch, and MIT Industrial Organization Lunch for helpful comments.
1 Introduction

The modern age is brimming with information—a vast and endless ocean which exponentially accumulates. Yet, our access to this information is, in practice, mediated by platforms which funnel our attention toward algorithmically curated streams of content—articles to digest, videos to watch, feeds to scroll—which we have little control over. We are inundated with information, yet not truly free to explore the ocean’s depths. These platforms, in turn, generate revenue primarily through advertisements; their business, therefore, is to capture, repackaging, and monetize attention.

Motivated by these developments, we study the extent to which information can be used to capture attention. A platform mediates a single decision maker (DM)’s knowledge by choosing some dynamic information structure. The DM values information instrumentally and at each point in time, chooses between observing more information generated by this dynamic information structure, or stopping and taking action based on her best available information. The platform faces the following trade off: if it gives too much information too quickly, the DM learns what she needs and takes action early; conversely, if it gives the DM too little information, the DM might not find it optimal to keep paying attention—perhaps because she dislikes advertisements, or because waiting is costly—and stops to act even though she is relatively uncertain. This setting is one in which the DM is not free to choose what to pay attention to, only whether to pay attention at all. Our contribution is to give a relatively complete characterization of (i) stopping times achievable through information; (ii) general properties of information structures which implement every achievable stopping time; (iii) designer-optimal structures under nonlinear preferences; and (iv) the set of achievable welfare outcomes.

1.1 Summary of results. Fix a dynamic information structure which, for each time period and each history, specifies a joint distribution over messages and states. This induces some distribution over DM stopping times: on some paths of message realizations, the DM might become confident enough that she stops early and takes action; on other paths, the DM might observe information for a long stretch. Our first result, and a tool for the rest of our analysis, is a reduction principle: for any distribution of stopping times achievable by some dynamic information structure, we can achieve the same distribution of stopping times through a dynamic information structure which (i) gives the DM full information at some random time, and the null message otherwise; and (ii) upon receipt of the null message, the DM prefers to continue paying attention. We call these simple and obedient dynamic information structures. This holds quite generally, and applies to any decision problem, for any time-dependent cost function and nests standard functional forms (e.g., per-period cost, exponential discounting). Simple and obedient structures have the attractive property that it induces a unique path of beliefs (conditional on the DM not stopping). We use this reduction to study the set of distri-
utions over stopping times achievable through information which we denote with $D$. $D$ is closed under a relation we call dynamic-FOSD which requires that at every time the conditional distribution going forward is FOSD. When costs are additively separable, every feasible stopping time can be implemented through simple and obedient structures which have increasing and extremal belief paths. The paths are increasing in the sense that conditional on not having learnt the state, the DM’s uncertainty increases over time, and her beliefs are progressively steered towards a region at which her value for information is maximized. The paths are extremal in the sense that along this path, her beliefs move as much as the martingale constraint on beliefs will allow. As time grows large, the DM’s beliefs also converge to this region.

We then explicitly introduce a designer whose value function is an arbitrary increasing function of the DM’s stopping time. We show that every designer-optimal structure leaves the DM with zero surplus and give a partial characterization of the set of feasible welfare outcomes for the DM and designer. We then detail a class of information structures which solves the designer’s problem for any increasing value and cost function; optimal solutions are simple and obedient, and take on a block structure: time is split into blocks of consecutive periods. On the interior of each block, the DM receives no information with probability one; on the boundaries of each block, the DM learns the state with probability which makes her exactly indifferent between paying attention and stopping. This block structure has the natural interpretation as the policy of a platform which controls the duration of advertisements (length of each block). Over the duration of each advertisement (interior of each block), the DM receives no information; at the end of each advertisement (boundary of each block), the designer gives the DM full information with some probability, and shows the DM yet another advertisement with complementary probability. Crucially, since the DM has already sunk her attention costs, conditional on being shown yet another advertisement, she finds it optimal to continue paying attention although she would have been better-off not paying attention in the first place.

When the designer’s value function is more concave than the receiver’s cost function, the optimum can be implemented by giving full information with probability one at a fixed, deterministic time, and no information otherwise. When the value function is more convex than the receiver’s cost function, the optimum can be implemented by giving the DM full information with positive probability at all time periods, where this probability decays exponentially at time-varying rates. These structures are two extremes on either end of the space of block structures. When the value function is ‘S-shaped’ as studied extensively in the marketing literature, the designer’s optimal structure can be implemented through a suitable mixture of

---

3i.e., as well-off as if she did not pay attention in the first place—this is straightforward in the “transferable utility” case where the designer’s value function is a linear transformation of the cost function.

4For instance, this is the form of advertisements (ads) on popular streaming platforms such as Youtube, Spotify, and Tiktok which frequently show users ‘unskippable’ ads. Such platforms vary both the type, duration, and number of ads users see; the popular streaming platform Youtube recently experimented with showing users up to 10 unskippable ads in a row.

5More precisely, the distribution of stopping times follows a monotone transformation of a geometric distribution where the transformation depends on the DM’s cost function.

6A sizeable literature in marketing [Bemmaor, 1984; Vakratsas, Feinberg, Bass, and Kalyanaram, 2004; Dubčę, Hitsch, and Manchanda, 2005] studies ‘S-shaped’ response curves in which there exists a threshold (i) below which consumers are essentially nonreactive to advertisement; and (ii) beyond which consumers are much more
the concave and convex optimal structures outlined above.

1.2 Related literature.

**Dynamic information design.** Our paper contributes to a burgeoning literature in dynamic information design. [Ball (2019)] studies a setting in which the sender and receiver have misaligned preferences and shows how information can be optimally employed as leverage to incentivize the sender’s preferred action. [Ely and Szydlowski (2020)] study a problem in which the principal’s payoff is linear in the time the agent spends working; the agent pays a per-unit cost for time worked, and receives a fixed reward if and only if she has worked long enough—above some unknown threshold $x$ which she learns about. In their setting, exerting costly effort and observing information tends to happen together but information is not costly *per se*. By contrast, we study a setting which links information and cost directly: as soon as DM learns the state perfectly, there is nothing left the designer can offer her and she stops to act immediately. Our setting is, in this regard, simpler since we abstract away from more complex principal-agent problems; but it is, in other regards, also more general because we (i) allow for arbitrary decision problems and time-dependant costs; (ii) focus on characterizing the distributions of stopping times which can be achieved as well as the properties of structures which implement them; and (iii) solve a designer’s problem for non-linear value functions.

The closest work to ours is that of [Knoeple (2020)] and the contemporaneous paper of [Hébert and Zhong (2022)], both of which independently study related settings. Although the motivation of these papers are similar, our papers differ in both results and techniques. [Knoeple (2020)] focuses on competition for attention among multiple designers. [Hébert and Zhong (2022)] study the case in which a designer wishes to maximise ‘engagement’—measured in the difference of a generalized entropy measure—rather than time spent paying attention. Further, they model the DM as having limited information processing capacity which constrains the extent to which beliefs can move between periods; this is the main friction in their model which is absent in ours.

Our contribution is distinct from [Knoeple (2020)] and [Hébert and Zhong (2022)] in several regards. First, we make progress on understanding all distributions of stopping times which can be achieved through information rather than solving for equilibria arising from particular economic settings. This set delineates possible outcomes across a wide class of models. For instance, we might imagine that the DM has the option to pay the information provider to receive information sooner (e.g., paid subscriptions which still show ads), or there might be several information providers who compete or collude. In richer environments such as these, all equilibrium outcomes must remain within the set. Furthermore, we derive general responsive before becoming saturated and tapering off.

Several papers study dynamic persuasion in which the sender’s preferences are over the DM’s action à la Bayesian persuasion ([Kamenica and Gentzkow (2011)], [Ely (2017)] and [Renault, Solan, and Vieille (2017)]) study the case in which the state transitions stochastically, and the DM takes action every period; [Guo and Shmaya (2018)] studies the case with multiple receiver types and show that dynamic persuasion can sometimes be optimal.

Several streaming giants (e.g., Netflix, Hulu, and Disney+) have launched ‘ad-supported’ plans: consumers enjoy a discount on full-price subscriptions, but are shown ads. Similarly, Youtube has launched a ‘Premium’ subscription service where consumers can pay to avoid ads.

This is in a similar spirit to the literature on informational robustness in games of incomplete information.
properties of information structures which implement all feasible stopping times.\(^\text{10}\) Second, we solve the problem in which the value of attention for the designer and the cost of attention for the DM are arbitrary increasing functions (e.g., ‘S-shaped’). This is motivated by an extensive literature in marketing and psychology. Within this nonlinear environment, we (i) characterize designer-optimal information structures; and (ii) show that all designer-optimal designs leave the DM with no utility as well as give a characterization of the feasible surplus pairs which can arise between the designer and DM. Third, the reduction principle is a helpful tool which substantially reduces the space of all dynamic information structures—this might be of wider applicability in future work.\(^\text{11}\)

**Information acquisition.** Our paper is related to the literature on rational inattention and costly information acquisition (Sims (2003), Pomatto, Strack, and Tamuz (2018), Morris and Strack (2019), Zhong (2022)). In these models, the DM is typically free to choose any information structure but pays a cost which varies with this choice (e.g., increasing in the Blackwell order). Our paper complements this literature by studying the opposite case in which the DM has no control over the information structure she faces, and whose cost of paying attention depends only on time.

**Sequential learning.** Our paper also contributes to the literature on sequential learning, experimentation, and optimal stopping starting from Wald (1947) and Arrow et al. (1949). One set of papers explore settings in which the DM’s attention is optimally allocated across several exogenous and stationary information sources (Austen-Smith and Martinelli 2018, Che and Mierendorff 2019, Gossner, Steiner, and Stewart 2021, Liang, Mu, and Syrgkanis 2022). Also related is the paper of Fudenberg, Strack, and Strzalecki (2018) who study the tradeoff between a DM’s speed and accuracy within an uncertain-difference drift-diffusion model driven by exogenous Brownian signals. In our setting, the DM only faces a single source of information and our motivation is to understand how distributions of stopping times vary with the information structure.

**Direct recommendations.** Our reduction principle bears some resemblance to the revelation principle (Gibbard, 1973; Myerson, 1982, 1986). More recently, Ely (2017) develops an ‘obfuscation principle’ which, roughly, states that the designer can simply tell the DM what her beliefs should be after every history.\(^\text{12}\) Our reduction principle is in this spirit, but goes

---

\(^{10}\)We show that *every feasible stopping* time is implementable through simple and obedient structures which progressively make the DM more uncertain in a deterministic fashion (increasing belief paths). This is related to the independent result of Hebert and Zhong (2022) who show that in a linear setting (value and cost functions are linear) with capacity constraints, an optimal structure can be implemented through what they call ‘suspensive signals’ in which the DM’s beliefs remain in a region at which the DM is ‘more uncertain than the prior’.

\(^{11}\)Indeed, the ‘All-or-nothing’ structure studied in Knoeple (2020) is a special case. Similarly, the solution of Hebert and Zhong (2022) might be loosely viewed as a continuous time analog of simple and obedient structures (albeit with limits on ‘jumps’ because of the DM’s processing capacity).

\(^{12}\)This allows Ely (2017) to optimize over stochastic processes (beliefs) rather than information policies; see also Renault, Solan, and Vieille (2017) who implicitly do this. Our approach for much of this paper will similarly be to work directly on the space of beliefs rather than information structures. However, in our setting the DM’s optimal decision at time \(t\) depends not just on her current belief, but also on the information structure she faces going forward. This feature is absent in both Ely (2017) and Renault, Solan, and Vieille (2017) where the DM
a step further: instead of telling the DM what her beliefs should be, we explicitly modify the
DM’s beliefs (i) on histories at which she stops; and (ii) histories at which she continues to
pay attention. This yields a unique path of beliefs (conditional on paying attention so far). We
draw a more explicit connection after we state the reduction principle formally.

**Information design with nonlinear preferences.** Finally, there is recent interest in static
persuasion when preferences over DM’s actions are nonlinear ([Dworczak and Kolotilin](#) 2022,
Kolotilin, Corrao, and Wolitzky, 2022). We complement this by studying a dynamic setting
with nonlinearities in preferences over stopping times. We employ different tools and, after
reducing the nonlinear problem to optimizing over simple and obedient structures, introduce
the notion of conditional concavification to exploit the temporal structure of the problem.

**1.3 A simple example.** To illustrate a few of the results in our paper, we begin with a simple
eexample. All supporting calculations are in Online Appendix VI.

**Example 1. Part 1: Reduction.** Suppose that DM’s utility from taking action $a \in \{0, 1\}$ given
the state is $\theta \in \{0, 1\}$ at time $t = 0, 1, 2, \ldots$ is given by

$$v(a, \theta, t) = -(a - \theta)^2 - (c \cdot t)$$

for some constant $c > 0$. Denote DM’s belief at time $t$ after the history $(m_i)_{i=1}^t$ with $\mu_t|(m_i)_{i=1}^t := \mathbb{P}(\theta = 1|(m_i)_{i=1}^t)$, letting $\mu_0 = 1/2$ be her initial belief. Suppose DM faces the following dy-
namic information structure: at time $t \geq 1$, message $m_t$ takes on a binary value according to

$$\mathbb{P}(m_t = 1|\theta = 1) = \mathbb{P}(m_t = 0|\theta = 0) = \begin{cases} 2/3 & \text{for } t \leq 3 \\ 1/2 & \text{otherwise} \end{cases}$$

for the first three periods the receiver obtains iid draws of noisy but informative signals; after
$t = 4$, messages are completely uninformative. We represent this information structure as a
tree in Figure 1(a). One can verify that for sufficiently low $c > 0$, the DM finds it optimal to
stop at the nodes marked in Figure 1(a), and optimal to continue at all other nodes.

First consider the following modification: on the nodes at which the DM stops in the original
information structure, we instead give the DM full information. Call the resultant information
structure $I'$; this modification is depicted in Figure 1(b) where we represent full information
as the node with ‘F’. Full information can be given in such a way that if the DM does not
learn the state, her posterior must be equal to that of the corresponding node in the original
structure. Clearly, the DM’s incentives to continue paying attention weakly increase at each

---

13Explicitly, full information corresponds to two messages; upon receipt of either, the DM learns that the state
is $\theta = 1$ or $\theta = 0$.
14This can always be done by choosing the probabilities that the DM learns the state under each state suitably.
This is because there are two equations (the first equation ensures full information is given with probability 5/9,
the second ensures that beliefs conditional on not receiving full information is preserved) and two unknowns.
Figure 1: Illustration of the modifications $I \rightarrow I' \rightarrow I''$

Note: The number inside each node represents DM’s belief given the history of messages leading up to the node, and the number on each branch represents the probability of moving to the next node with respect to DM’s current belief (i.e., the belief on the node the branch stems from).

non-terminal node since the modification to full information at terminal nodes must improve her continuation value.

Next consider the further modification depicted in Figure 1(c) which is constructed by taking the information structure in Figure 1(b) and collapsing the two branches stemming from histories $\{m_1 = 1\}$ and $\{m_1 = 0\}$ into a single branch. In particular, DM now receives no information (e.g., the null message $\emptyset$) with probability one at time 1. We need to check that if DM did find it optimal to continue paying attention after the histories $\{m_1 = 1\}$ and $\{m_1 = 0\}$ in information structure $I'$, she must also find it optimal to continue paying attention after the history $\{m_1 = \emptyset\}$ in information structure $I''$. Notice that the DM’s belief under $I''$ following history $\{m_1 = \emptyset\}$ is 1/2 which is a mean-preserving contraction of the DM’s beliefs on $I'$ at history $\{m_1 = 1\}$ (belief = 2/3) and $\{m_1 = 0\}$ (belief = 1/3). But since (i) the DM found it optimal to continue paying attention under $I'$ on those histories; and (ii) the continuation payoff conditional on not stopping before $t = 1$ is identical under $I'$ and $I''$, an elementary application of Jensen’s inequality implies the DM must also find it optimal to continue paying attention under $I''$ on history $\{m_1 = \emptyset\}$. We have thus replicated the stopping time under $I$ with $I''$—this structure is simple and obedient.

**Part 2: Optimization.** Suppose a designer is interested in maximizing some increasing function $f$ of the DM’s stopping time. We now optimize among simple and obedient structures. We will suppose that the DM’s per-period cost $c = 1/4$. A preliminary observation is that at the start of $t = 0$, the DM’s value of learning the state is $1/2^{15}$. Now consider the following structure depicted as $I_1$ in Figure 2(a): at $t = 1$, give the null message with probability 1; at

---

15 This is the value of taking action while fully informed minus the value of taking action under her prior $\mu_0$.
Suppose, instead, that the designer’s value of the DM’s attention is non-linear. For instance, the DM willingness to pay might be an S-shaped function of time she spent viewing ads, as shown in the curve in Figure 2(b).

Now consider an alternate information structure in which we give the DM full information with probability 1/2 at time $t = 1$, but conditional on not doing so, give the DM no information with probability 1 at $t = 2$, and full information with probability 1 at $t = 3$. This structure is depicted as $I_2$ in Figure 2(a). One can check that at every node at which the DM does not receive full information, the DM prefers to continue paying attention. Furthermore, $E[\tau(I_2)] = 2$ where we use $\tau(I)$ to denote the stopping time induced by the structure $I$. Convexity of $f$ for $t \leq 3$ then allow us to conclude that $I_2$ is an improvement over $I_1$; this is depicted in Figure 2(b). Furthermore, it turns out that $I_2$ is optimal across all dynamic information structures. For instance, the stopping time $\tau$ which places probability of stopping on times 0 and 3 such that $E[\tau(I_2)] = 2$, as depicted in Figure 2(c) is infeasible because it violates some obedience constraint.

---

$^{16}$A common technique in the related literature (Knoeplle, 2020; Hébert and Zhong, 2022) is to only focuses on the time 0 obedience constraint and show that the bound can be obtained. However, this only works when the designer’s and DM’s objectives are diametrically opposed e.g., $c$ and $f$ are both linear. In Section 4 we show how to solve the problem directly.

$^{17}$The curve interpolates our discrete increments $f(1), f(2), \ldots$

$^{18}$It might be helpful to compare this with the well-known setting of Bayesian persuasion in which, over the space of beliefs, the designer is free to choose any distribution of posteriors subject only to a martingale condition on the first moment. In our problem, however, the set of feasible stopping times is not closed mean preserving spreads.
Example was quite special with constant per-period cost, quadratic payoffs, and symmetric, history-independent information structures. It turns out that none of these is essential for the reduction from $I$ to $I'$. Similarly, the information structure $I_2$ we found which solves the designer’s non-linear problem has a ‘block structure’—we will later show that this class of information structures solve the designer’s problem for any value function. The rest of the paper is organized as follows. Section 2 details the general model. Section 3 introduces the reduction principle and applies it to develop properties of the set of feasible stopping times and structures which implement them. Section 4 introduces a designer whose value for the DM’s attention is any increasing function. It develops results on the set of feasible welfare outcomes, as well as solves the designer’s problem. Section 5 concludes.

2 Model

Time is discrete and infinite, indexed by $T = 0, 1, 2, \ldots$. The state space $\Theta$ is finite with $\theta \in \Theta$ as a typical member.

2.1 Dynamic information structures. A dynamic information structure is a series of functions

$$\left\{ p_t(\cdot | \theta, H_t) \right\}_{t \in T, \theta \in \Theta, H_t \in \mathcal{H}}$$

where $p_t(m | \theta, H_t)$ denotes the probability that message $m \in M$ is realised given that the state is $\theta$ and given the history of past message realizations $H_t \in \mathcal{H}_t := \prod_{s=0}^{t} M$. We shall assume that our message space $M$ is large but finite and the same across every period. Let $I$ denote the space of all dynamic information structures where we use $I_2$ to denote a typical member.

2.2 Decision maker’s payoffs. A decision maker (DM) has prior $\mu_0 \in \Delta(\Theta)$ which we shall assume has full support over $\Theta$. DM faces a single decision problem to maximize $u : A \times \Theta \rightarrow \mathbb{R}$ where $\theta \in \Theta$ is the true state, and $a \in A$ is her action. In addition, DM observes information about the state generated by a dynamic information structure. If DM has observed information up to time $t$ (i.e., the messages $H_t = (m_s)_{s=1}^t$), she decides between stopping and taking her action under her current belief $\mu_t | H_t \in \Delta(\Theta)$, or delaying her decision to observe more information.

Delay is costly and DM’s utility taking action $a$ at time $t$ given that the state is $\theta$ is given by $v : A \times \Theta \times T \rightarrow \mathbb{R}$ where we impose

(i) (Impatience) $t > t' \implies v(a, \theta, t) < v(a, \theta, t')$; and

(ii) (Time invariance) $u(a, \theta) \geq u(a', \theta) \implies v(a, \theta, t) \geq v(a', \theta, t)$.

Our formulation of $v(a, \theta, t)$ is quite general\(^{19}\) and nests (i) per-period cost: $v(a, \theta, t) = u(a, \theta) - $

\(^{19}\)See Fishburn and Rubinstein (1982) who study axioms and utility representations for decision problems in which the DM receives a single outcome (as opposed to, say, streams of consumption over time) at a particular time (in our setting, this is when the DM takes action). Property (i) maps to axiom A3; property (ii) maps to axiom A2.
\(c \cdot t\); and (ii) exponential discounting: \(v(a, \theta, t) = \delta^t u(a, \theta)\) for non-negative \(u\).\(^{20}\)

It will be helpful to define \(v^*(\theta, t) := \max_{a \in A} v(a, \theta, t)\), as the DM’s payoff under state \(\theta\) at time \(t\) when she chooses an optimal action.

### 2.3 Measures under different dynamic information structures.
We will often vary the dynamic information structure to understand how variables of interest (e.g., probabilities of histories, incentive compatibility constraints etc.) change. To this end, we will use \(\mathbb{E}^I[\cdot]\), \(\mathbb{E}^I(\cdot)\), and \(\mathbb{P}^I(\cdot)\) to denote the unconditional and conditional expectations and probabilities under dynamic information structure \(I \in \mathcal{I}\). For instance, \(\mathbb{P}^I(H_t)\) and \(\mathbb{P}^{I'}(H_t)\) are the unconditional probabilities that history \(H_t \in \mathcal{H}_t\) realizes under dynamic information structures \(I\) and \(I'\) respectively. Throughout this paper we use superscripts to track dynamic information structures, and subscripts to track time periods.

### 2.4 Decision maker’s optimization problem.
Facing dynamic information structure \(I \in \mathcal{I}\), the DM solves the following optimization problem:

\[
\sup_{\tau, a_\tau} \mathbb{E}^I[v(a_\tau, \theta, \tau)],
\]

where \(\tau\) is a stopping time and \(a_\tau\) is a (stochastic) action under the natural filtration.\(^{21}\) Throughout we will assume that the DM breaks indifferences in favour of not stopping. This will ensure that the set of feasible stopping times is closed though it is not essential for the main economic insights.

### 2.5 Feasible distributions of stopping times.
For a given information structure \(I \in \mathcal{I}\), this induces the DM to stop at random times. We call this the stopping time induced by \(I\), and denote it with \(\tau(I)\) which is a \(\mathcal{T} \cup \{+\infty\}\)-valued random variable. It will also be helpful to denote the distribution of \(\tau(I)\) with \(d(I) \in \Delta(\mathcal{T} \cup \{+\infty\})\).

**Definition 1** (Feasible stopping time). The stopping time \(\tau\) is feasible if there exists an information structure \(I \in \mathcal{I}\) such that \(\tau \overset{d}\rightarrow \tau(I)\).

Define

\[
\mathcal{D} := \{d(I) : I \in \mathcal{I}\} \subseteq \Delta(\mathcal{T} \cup \{+\infty\})
\]

as the set of all distributions of feasible stopping times.

### 2.6 Simple and obedient dynamic information structures.
We now introduce a special class of information structures.

\(^{20}\)These special cases are stationary in the sense that shifting in time does not change preferences over outcomes (action-state pairs) and times: \(v(a, \theta, t) = v(d, \theta', t+k) \iff v(a, \theta, t') = v(a', \theta', t'+k)\) for any \(k, t, t' \in \mathcal{T}\) and \((a, \theta), (a', \theta') \in A \times \Theta\). All results in this paper do not depend on stationarity.

\(^{21}\)A sequential formulation of this problem is given in Appendix when we prove Proposition.
**Definition 2** (Simple and obedient dynamic information structures). A dynamic information structure is simple and obedient if,

(i) (Simplicity) There exists a set of distinct messages \( \{m_\theta \}_{\theta \in \Theta} \cup \{m_0\} \subseteq M \) such that

\[
p_t(m_\theta|\theta, H_t) + p_t(m_0|\theta, H_t) = 1 \quad \text{for all } \theta \in \Theta, t \in T, H_t \in \mathcal{H}_t;
\]

(ii) (Obedience) DM (weakly) prefers to continue paying attention on every history \( H \in \{(m_\theta)^t : t \in T\} \) which is reached with positive probability.

Part (i) is the simplicity condition which states that at every time \( t \) and history \( H_t \), the DM either (i) learns the state—upon receipt of \( m_\theta \), she places full probability on \( \theta \); or (ii) receives the null message \( m_0 \). As such, upon receipt of message \( m_\theta \) for any \( \theta \), the DM learns the state immediately and stops paying attention. We will refer to dynamic information structures which fulfil condition (i) (but not necessarily (ii)) as simple. Part (ii) is the obedience condition which states that upon receipt of the null message \( m_0 \), the DM must always weakly prefer to continue paying attention. We note that simplicity (condition (i)) is purely a property of the dynamic information structure, whereas obedience (condition (ii)) is a joint property of the dynamic information structure, the decision problem, and the DM’s prior belief. Denote the set of all simple and obedient dynamic information structures with \( I^* \subset I \).

Figure 3: Illustration of simple structures

![Figure 3: Illustration of simple structures](image)

Figure 3 illustrates the general path of beliefs under simple information structures. Here we set \( \Theta = \{1, 2, \ldots, n\} \). For any time \( t \in T \), if the DM has not yet learnt the state (i.e., has seen the history \( (m_\theta)^t \)), we denote her belief on that history with \( \mu_t \). If the DM decides to continue paying attention, her beliefs in period \( t + 1 \) transitions to either (i) certainty in which she puts full probability on some state \( i \in \Theta \)—we denote this belief with \( \mu^i \); or (ii) \( \mu_{t+1} := \mu_{t+1} \mid (m_\theta)^{t+1} \).

Simple information structures (condition (i) in Definition 2) take this form. Obedience additionally requires that on every history \( (m_\theta)^t \), the DM must prefer to continue paying attention for an additional period.
3 Properties of feasible distributions.

In this section we make progress on understanding $\mathcal{D}$, the set of all feasible distributions of stopping times induced by some information structure. Because, the set of all dynamic information structures is large, our first step will be to develop a reduction result which states that all stopping times achieved by $\mathcal{I}$ can be achieved by $\mathcal{I}^+$, the set of simple and obedient dynamic information structures.

3.1 Reduction to simple and obedient structures.

**Proposition 1** (Reduction principle). For any $d \in \mathcal{D}$, there exists a simple and obedient dynamic information structure $I \in \mathcal{I}^+$ such that $d(I) = d$.

Proposition 1 states that every distribution of stopping times which is achievable by some dynamic information structure is also achievable by a simple dynamic information structure. As such, restricting our attention to the set $\mathcal{I}^+$ is sufficient to characterise $\mathcal{D}$:

$$\mathcal{D} := \{d(I) : I \in \mathcal{I}\} = \{d(I) : I \in \mathcal{I}^+\}.$$  

The proof of Proposition 1 proceeds in two steps. Fix any $I \in \mathcal{I}$. The first step modifies the histories on which DM optimally stops so that the DM now receives full information with probability one on those histories. This increases the DM’s incentives to stop at histories on which she originally stopped, and decreases DM’s incentives to stop at histories on which she originally did not. This step is analogous to the modification from $I$ to $I'$ in Example 1.

The second step collapses all branches comprising the tree of possible histories together. The key observation underlying this step is that when we collapse several time-$t$ histories—say, $H_1^t, H_2^t, \ldots, H_n^t$—together into the history $H_1^*$, if the DM did not find it optimal to stop at the histories $H_1^t, H_2^t, \ldots H_n^t$, then she cannot find it optimal to stop at the collapsed history $H_1^*$. Here we offer a sketch of the basic argument. First observe that on each history, the DM’s incentives to stop depends on her expected payoff from taking her best action under the current belief at that history. Furthermore, note that the maximum operator is convex. Finally, note that the beliefs on the collapsed history $H_1^*$, $\mu[H_1^*]$ must be a mean-preserving contraction of $\mu[H_1^t, \ldots, \mu[H_n^t]$. An elementary application of Jensen’s inequality then implies that the DM must do weakly worse by stopping with belief $\mu[H_1^*]$ than a hypothetical lottery over stopping with beliefs $\mu[H_1^t, \mu[H_2^t], \ldots, \mu[H_n^t]$ with mean $\mu[H_1^*]$. As such, if the DM found it optimal to continue paying attention at time $t$ on each of the original histories, the linearity of the expectation operator implies that she must also find it optimal to do so at the collapsed history. This step is analogous to the modification from $I'$ to $I''$ in Example 1. Our sketch leaves out subtle details; the full proof of Proposition 1 is in Appendix A.

We now employ the reduction to
study properties of \( D \).

### 3.2 Conditional first-order stochastic dominance.

A preliminary observation is that \( D \) is convex—this is immediate by randomizing over dynamic information structures. Our next result develops an ordering over stopping times such that if an analyst knows that \( d \in D \), then all distributions which \( d \) dominates under this ordering is also implementable through information.

**Definition 3** (Conditional First Order Stochastic Dominance (C-FOSD)). We say that

\[
\tau \overset{\text{C-FOSD}}{\geq} \tau' \quad \text{if, for all times } t \in T, \quad \tau|\tau > t \overset{\text{FOSD}}{\geq} \tau'|\tau' > t.
\]

We will sometimes abuse notation and use \( d \overset{\text{C-FOSD}}{\geq} d' \) to refer to the C-FOSD relation between random variables distribution according to \( d \) and \( d' \) respectively. C-FOSD is equivalent to a decreasing ratio of CDFs between the dominant and dominated stopping times:

\[
\tau \overset{\text{C-FOSD}}{\geq} \tau' \iff \text{for every } t \in T, \quad \frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau' > t + 1)} \geq \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau' > t)}.
\]

The proof of this fact is in Appendix B. Conditional first order stochastic dominance is linked to the DM’s problem through the following intuition. When we consider simple and obedient information structures, the DM’s expected continuation utility at time \( t \) after history \( H_t \) depends only on the distribution of times at which she learns the state. But at time \( t \), this distribution is conditional on having not learnt the state up to time \( t \); C-FOSD reflects exactly this.

**Proposition 2.** If \( d \in D \) and \( d \overset{\text{C-FOSD}}{\geq} d' \) then \( d' \in D \).

Proposition 2 states that if \( d \) is feasible and \( d \) conditionally first-order stochastically dominates \( d' \), then \( d' \) is also feasible. This offers a tight characterisation of the set of feasible interior distributions i.e., those which are not conditionally FOSD dominated. We illustrate this relation in Figure 4. Panel (a) depicts the anti-CDFs (i.e., one minus the CDF) corresponding to the distribution of stopping times \( \tau \) (blue), \( \tau' \) (red), and \( \tau'' \) (green). Panel (b) plots the ratio of the anti-CDF between \( \tau \) and \( \tau' \), and that between \( \tau \) and \( \tau'' \). Clearly, both \( \tau' \) and \( \tau'' \) are first-order stochastically dominated by \( \tau \). Now note that since the ratio \( \mathbb{P}(\tau > t)/\mathbb{P}(\tau' > t) \) (red line in panel (b)) is increasing in \( t \), \( \tau \) conditionally FOSD \( \tau' \); however, the ratio \( \mathbb{P}(\tau > t)/\mathbb{P}(\tau'' > t) \) (green line in panel (b)) initially increases, but decreases back to 1 before increasing again. As such, \( \tau \overset{\text{C-FOSD}}{\geq} \tau' \) although \( \tau \not\overset{\text{C-FOSD}}{\geq} \tau'' \) i.e., we can conclude that \( \tau' \) is implementable but are not guaranteed the same for \( \tau'' \).

The reduction principle will allow us to sidestep information structures work directly on stopping times. This is helpful whenever the designer’s payoff depends solely on the DM’s stopping

---

The DM the single recommendation ‘continue’. However, the DM’s disobedience utility from stopping and taking her best action under her new (coarsened) belief is no longer additive; we handle this by invoking the convexity of the max operator.
time, as in the case of a platform aiming to maximize users’ impressions or views (see Section 4 where we consider this problem). This is analogous to standard techniques in static information design which takes a belief-based approach since the set of feasible posterior beliefs is tightly described by a martingale condition (Stiglitz and Rothschild [1970], Gentzkow and Kamenica [2016]). However, we note that because of the DM’s dynamic incentives, $D$ is not closed under mean-preserving spreads.\footnote{Consider, for instance, a setting with constant per-unit costs and the stopping time which puts probability 1 on $t = 2$ against the stopping time which puts probability $p$ on $t = 1$ and $1 - p$ on $t = (2 - p)/(1 - p)$. For sufficiently high $p$, the latter is not implementable although the former might be. The problem is that if the DM does not receive full information after $t = 1$, it is not sequentially optimal to continue paying attention.}

In Section 4 we take an alternative approach to solve a designer’s nonlinear problem.

### 3.3 Belief paths.

Our reduction principle showed that it is without loss to restrict our attention to simple and obedient structures. An attractive property of such structures is that conditional on continuing to pay attention at time $t$, the DM’s beliefs at time $t$ is deterministic; we call these belief paths.

**Definition 4** (Belief path). A belief path associated with the simple and obedient structure $I \in I^*$ is a unique sequence of beliefs $(\mu_t)_{t \in T} \in \Delta(\Theta)^T$ such that for any $\theta \in \Theta$,

$$\mu_t(\theta) = \mathbb{P}^I(\theta | H_t = (m_0)^t)$$

for all histories $(m_0)^t$ reached with positive probability.\footnote{Uniqueness is up to histories reached with positive probability. We leave beliefs on histories reached with zero probability unspecified.}

Observe that the only history at which the DM pays attention up to time $t$ is the one in which she receives the null message for every $s < t$. As such, conditional on paying attention up to $t$,
her belief at \( t \) is uniquely pinned down by \( \mu_t \). To make progress on understanding belief paths and its relationship to feasible stopping times, we will maintain the following assumption.

**Assumption (Additively separable cost).** We say that costs are additively separable if we can write

\[
\Pi(a, \theta, t) = u(a, \theta) - c(t)
\]

for some strictly increasing function \( c : \mathcal{T} \to \mathbb{R}_{\geq 0} \) where we normalize \( c(0) = 0 \) and assume \( \lim_{t \to \infty} c(t) = +\infty \). \(^{25}\)

Additively separability adds structure to the problem by separating the DM’s value from receiving information from the time at which she does so. More precisely, the DM’s incentives to continue paying attention at time \( t \in \mathcal{T} \) after history \( H_t \in \mathcal{H}_t \) under information structure \( I \in \mathcal{I}^* \) is given by the condition

\[
\max_{a \in A} \mathbb{E}[u(a, \theta, t) | H_t] \leq \mathbb{E}[u^*(\theta, \tau) | \tau > t].
\]

There are two sources of uncertainty within the expectation operator on the right: uncertainty over the time at which the receiver learns the state, as well as the payoff the receiver will receive then, which depends on what the true state is. When \( \Pi(a, \theta, t) \) is an arbitrary function, these two sources of uncertainty can interact in complicated ways. Additive separability allows us to rewrite the above condition:

\[
\mathbb{E}[u^*(\theta) | H_t] - \max_{a \in A} \mathbb{E}[u(a, \theta) | H_t] = (m_{\theta})^T \geq \mathbb{E}[c(\tau) | \tau > t] - c(t)
\]

where \( u^*(\theta) := \max_{a \in A} u(a, \theta) \). Define the following functional over the space of beliefs \( \phi : \Delta(\Theta) \to \mathbb{R}_{\geq 0} \) such that for all \( \mu \in \Delta(\Theta) \),

\[
\phi(\mu) := \mathbb{E}_{\mu}[u^*(\theta)] - \max_{a \in A} \mathbb{E}_{\mu}[u(a, \theta)].
\]

\( \phi(\mu) \) is the additional value the DM gains from full information under the belief \( \mu \). \( \phi \) is continuous\(^{26}\) and concave since \( \mathbb{E}_{\mu}[u^*(\theta)] \) is linear and \( \max_{a \in A} \mathbb{E}_{\mu}[u(a, \theta)] \) is convex. Define:

\[
\Phi^* := \arg\max_{\mu \in \Delta(\Theta)} \phi(\mu) \subseteq \Delta(\Theta) \quad \phi^* := \max_{\mu \in \Delta(\Theta)} \phi(\mu).
\]

\( \Phi^* \) is the set of beliefs for which the DM’s benefit from obtaining full information relative to stopping immediately is maximized; \( \phi^* \) is the maximum of \( \phi \). Since \( \phi \) is concave, \( \Phi^* \) is convex.

\(^{25}\)The following separability condition (also known as the Thomsen condition) over outcomes [Debreu 1960]:

\[
((a, \theta), t) \sim ((a', \theta'), s) \text{ and } ((a', \theta'), u) \sim ((a'', \theta''), t) \implies (a, \theta, u) \sim ((a'', \theta''), s)
\]

for all \((a, \theta), (a', \theta'), (a'', \theta'') \in A \times \Theta \) and \( t, s, u \in \mathcal{T} \), together with basic regularity axioms are sufficient to imply representation through an additively separable \( \Pi \) [Fishburn and Rubinstein 1982]. Separability is much weaker than stationarity; see Section 5 of [Fishburn and Rubinstein 1982] for justification of the separability condition in the setting in which the DM obtains a single outcome at a particular time period.

\(^{26}\)We endow \( \Delta(\Theta) \) with the topology of weak convergence, which makes it metrizable and compact. Berge’s maximum theorem implies that \( \phi \) is continuous.
We will often refer to $\Phi^*$ as the basin of uncertainty.

Simple and obedient structures can be thought of as choosing both a stopping time $\tau$ with distribution in $\Delta(T)$, as well as a belief path $(\mu_t)_{t \in T}$. These two objects impose constraints on each other and are interlinked in the following sense. Fixing a particular belief path $(\mu_t)_t$, whether the stopping time is feasible depends on whether, at each belief $\mu_t$, facing the distribution of stopping times going forward, the DM prefers to continue paying attention. We call this the obedience constraint. Furthermore, fixing a particular distribution of stopping times, not every path of beliefs will be compatible—there are paths of beliefs which violate the martingale requirement that the expectation of the posterior is the prior. Call this the boundary constraint which we illustrate through Example 2 below.

It turns out that the obedience and boundary constraints offer a tight characterization of feasible stopping times. We formalize this in the following proposition which is proved in Appendix B.

**Proposition 3.** The following are equivalent:

1. There exists a simple and obedient information structure $I \in I^*$ which induces stopping time $\tau$ and belief path $(\mu_t)_{t \in T} \in (\Delta(\Theta))^T$.

2. The following conditions are fulfilled:

   (i) (Obedience constraint) $\phi(\mu_t) \geq \mathbb{E}[c(\tau) | \tau > t] - c(t)$ for every $t \in T$; and

   (ii) (Boundary constraint) $\mathbb{P}(\tau > t + 1|\mu_{t+1}(\theta)) \leq \mathbb{P}(\tau > t|\mu_t(\theta))$ for every $t \in T$ and $\theta \in \Theta$.

Condition (i) is the usual constraint that at each time $t$, the DM must find it optimal to continue paying attention upon receipt of the message $m_\theta$. Condition (ii) imposes a constraint on the degree to which beliefs can move between periods in terms of the probability that the DM learns the state at time $t + 1$ conditional upon paying attention up to time $t$. We emphasise that this boundary constraint is not an artefact of working in discrete time and would still obtain in continuous time if we work on the space of RCLL belief martingales.

We illustrate the boundary constraint through the following example.

**Example 2** (Illustration of boundary constraints). Suppose $\Theta = \{0, 1\}$ and for all $t > 0$, define $\mu_t = \mathbb{P}(\theta = 1|(m_\theta)^{t-1})$. Further suppose $\mu_0 = 1/4$. By Proposition 1, it is without loss to work on simple and obedient information structures in which at each period, the DM either learns the state, or receives the null message. Now suppose we would like the DM’s beliefs to move from $1/4$ in period 0 to $1/2$ in period 1. By the martingale condition on beliefs, we require that the probability that the DM receives full information in period 1 is be greater than $1/2$ (under the prior). If this were not so, then the probability of the null message, denoted $p$, is strictly greater than $1/2$. The DM’s expected posterior belief is then

$$\mathbb{E}_{\mu_0}[\mu_1] = (1/2) \cdot p + C \geq (1/2) \cdot p > 1/4 = \mu_0$$

The analogous condition in continuous time would be the degree to which belief can move over the interval $[a, b] \subset [0, +\infty)$ is constrained by the probabilities $\mathbb{P}(\tau \geq b)$ and $\mathbb{P}(\tau \geq a)$.
which violate the martingale condition for beliefs, where here \( C \geq 0 \) is the contribution from the extreme posteriors at which she learns the state. As such, for any fixed distribution of stopping times such that \( \mathbb{P}(\tau > 1) \geq 1/2 \), we are not able to choose \( \mu_1 = 1/2 \) without violating the boundary constraint.

Proposition 3 is a helpful auxiliary result which simplifies the analysis: any candidate stopping time is implementable through some dynamic information structure if and only if we can find a supporting belief path and stopping time which satisfy the obedience and boundary constraints. However, belief dynamics are of independent interest. The next result develops general properties of belief paths i.e., those which accompany any feasible stopping time.

**Proposition 4.** Every distribution \( d \in D \) is induced by a simple and obedient information structure \( I \in I^* \) for which

(i) (Increasing paths) \( \phi(\mu_j(m_0)') \) is increasing in \( t \); and

(ii) (Extremal paths) for every \( t \in \mathcal{T} \), either \( \mu_{t+1} \in \Phi^* \) or \( \mathbb{P}(\tau > t+1) = \min_{\theta \in \Theta} \frac{\mu_{t+1}(\theta)}{\mu_{t+1}(\theta)} \).

Figure 5: Illustration of increasing and extremal paths

\[ \Delta(\Theta) \]

(a) Increasing belief paths

(b) Extremal belief paths

Part (i) states that it is without loss to consider belief paths which are increasing in \( \phi \) i.e., the longer the DM goes without learning the state, the more uncertain she is. The underlying intuition is that by steering the DM’s belief toward \( \Phi^* \), this increases the value of full information for the DM—and hence this slackens future obedience constraints. However, the boundary constraint complicates the issue by imposing a limit on the degree to which beliefs can move between periods.\(^{28}\) Nonetheless, we can implement any feasible stopping time through in-
creasing paths. This is illustrated in Figure 5(a) in which the dotted lines are contour lines of $\phi$ (beliefs on the same dotted line have the same value of $\phi$). Starting from an initial belief path $(\mu_t)_t$ depicted by the solid red line, this might be strictly decreasing in $\phi$ for some times, as depicted by the blue portions. It turns out that we can always find an alternate sequence of beliefs $(\mu'_t)_t$ which still fulfills the obedience and boundary constraints, but which is now increasing in $\phi$. The new sequence of beliefs is illustrated by the dotted red line. Part (ii) states that it is without loss to consider extremal belief paths—those for which before reaching the basin $\Phi^*$, the boundary constraint binds. This is illustrated in Figure 5(b). For belief $\mu_t$ at time $t$, the shaded triangle shows the set of feasible beliefs $\mu_{t+1}$. Part (ii) states that it is without loss to consider paths for which $\mu_{t+1}$ is on the boundary of this set, depicted by the black dotted line.

The proof of Proposition 4 is quite technical and we defer it to Appendix B. Starting from an arbitrary $d \in D$, by the reduction principle it is without loss to consider a simple and obedient structure with implements $d$. This structure is associated with an arbitrary belief path $(\mu_t)_t$—the key challenge is to modify this belief path in a way that (i) preserves obedience at every time $t \in T$; while (ii) the modified belief path is both increasing in $\phi$ and extremal. A natural corollary of Proposition 4 is that at least for binary states, if the DM pays attention for long enough, her beliefs eventually reaches the basin of uncertainty.

**Proposition 5.** If $|\Theta| = 2$, there exists $T < +\infty$ such that either $P(\tau > T) = 0$ or $\mu_t \in \Phi^*$ for every $t > T$.

This follows straightforwardly from the increasing and extremal property of belief paths developed in Proposition 4. The proof is deferred to Appendix B but is illustrated in Figure 4. We conjecture that this property is also true for an arbitrary number of states; in Online Appendix III we provide a sufficient condition which guarantees it.

### 4 Introducing a designer with preferences over $\Delta(T)$

Our discussion thus far has abstracted away principal-agent problems to attempt to understand $D \subseteq \Delta(T)$, the set of feasible distributions achievable through information, as well as general properties of belief paths which implement them. In this section we introduce a designer (e.g., a platform) whose payoff is increasing in the amount of time the DM pays attention.
4.1 Designer’s problem. The designer’s problem is

\[ \sup_{i \in I} \mathbb{E}^i [f(\tau)] \]

where the designer’s value function \( f : T \rightarrow \mathbb{R} \) is a strictly increasing function of \( \tau \). For the problem to be well-defined (i.e., supremum is attainable), we maintain the following regularity assumption on the shape of \( f \) and \( c \).

**Assumption** (Regularity assumptions on \( f \) and \( c \)). Assume:

(i) \((f/c \text{ is bounded})\) There exist \( M > 0 \) such that \( f(t)/c(t) \leq M \) for all \( t \in T \).

(ii) \((c \text{ is regular})\) \( \sup_{t \in T} |c(t + 1) - c(t)| < +\infty \) and \( \inf_{t \in T} |c(t + 1) - c(t)| > 0 \).

Assumptions (i) and (ii) guarantee that \( \max_{i \in I} \mathbb{E}^i [f(\tau)] = \sup_{i \in I} \mathbb{E}^i [f(\tau)] \)\(^{29}\) Assumption (i) says that \( f \) cannot be much greater than \( c \); assumption (ii) says that increments in costs are bounded above and below.

Implicit in our formulation is that the designer can commit to dynamic information structures. We view this as reasonable in a wide range of settings in which the information structure—the content of news feeds, or the number of ads users are shown—is generated by an algorithm. For instance, the information structure we saw in Example 1 in which the DM receives no information with probability 1 at time 1, then full information at time 2 has a natural interpretation of a un-skippable video advertisement which lasts for two periods\(^{30}\).

4.2 Smoothing out \( f \) and \( c \). The coarseness of increments in the value and cost functions are a byproduct of working in discrete time. This makes results less clean, and can obscure the underlying economics. To this end, we will maintain the assumption that \( f \) and \( c \) are both ‘sufficiently smooth’: \( |f(t + 1) - f(t)| \) and \( |c(t + 1) - c(t)| \) can be made arbitrarily small. This assumption is solely to aid the exposition, and has no economic content. To see this, note that starting from an arbitrary increasing function \( f' \) and \( c' \) which fulfils the regularity assumptions above, we can define a sequence of models which linearly interpolates the increments, and then relabel times such that they once again run from 0, 1, 2, . . . . This is illustrated for \( f' \) in Figure 7\(^{31}\). The results for the rest of this section will be stated assuming both \( |f(t + 1) - f(t)| \) and \( |c(t + 1) - c(t)| \) are sufficiently small i.e., we have interpolated at sufficiently many points before relabelling time.

\(^{29}\)See Online Appendix for a formal treatment.

\(^{30}\)Indeed, Youtube closely monitors its advertisements’ ‘abandonment rate’—the proportion of users which exit the website when faced with long, un-skippable advertisements.

\(^{31}\)In Online Appendix we formally introduce the sequence of interpolating models and the deterministic time change.
4.3 Feasible combinations of DM and designer surplus. We first study how dynamic information structures influence the designer’s and DM’s surplus. Define

\[ S(I) := \left( \sup_{\tau, a_{\tau}} \mathbb{E}^I[g(a_{\tau}, \theta, \tau)], \mathbb{E}^I[f(\tau)] \right) \in \mathbb{R}_{\geq 0} \]

as the expected DM and designer surplus induced by information structure \( I \in \mathcal{I} \). We are interested in the Pareto frontier of achievable surplus outcomes which we denote with

\[ \mathcal{P} := \left\{ S(I) : \text{There is no } S(I’) \text{ such that } S(I’) > S(I) \right\} \]

where we use > to mean that the inequality is strict for at least one coordinate. We note that it is straightforward to move from the Pareto frontier to the set of feasible surplus pairs through a suitable mixing.

First suppose that the DM’s costs function is a linear transformation of the designer’s value function i.e., \( c(t) = a \cdot f(t) \) for any \( t \in \mathcal{T} \) and some \( a > 0 \). Because we showed that it sufficed to consider simple and obedient information structures in which the DM only stops paying attention when she is certain of the state, it is easy to see that the structure which maximizes the designer’s surplus must leave the DM with zero surplus since argmax_{\tau \in \mathcal{T}} \mathbb{E}^I[f(\tau)] = argmin_{\tau \in \mathcal{T}} \mathbb{E}^I[-c(\tau)] =: \bar{T} \] i.e., information structures which maximize the designer’s value are also those which minimize the DM’s total (ex-ante) surplus:

\[ 0 = \min_{\tau \in \bar{T}} \mathbb{E}^I[f(\tau)] = \mathbb{E}^I[f(\tau)] \quad \text{for any } \bar{T} \in \mathcal{T} \]

where the first equality follows because we can always find a simple and obedient structure which leaves the DM indifferent between paying attention and not in period 0.

However, when \( c \) and \( f \) are not linear transformations of each other, this argument no longer
works since the structure which maximizes the DM’s value function is not necessarily the one which maximizes the DM’s cost. In particular, since the designer extracts attention only dynamically, we cannot ex-ante rule out the possibility that the DM’s dynamic obedience constraints might force an optimizing sender to leave the DM with some positive utility. The next proposition extends the simple linear result to nonlinear settings: for any increasing $f$ and $c$, all designer-optimal structures leave the DM with zero surplus.

Before we state our result on welfare, we develop some additional notation. Let $f_{\mu_0}^\ast := \max_{\tau \in \mathcal{T}} \mathbb{E}[f(\tau)]$ be the maximum surplus attainable by the designer when the prior is $\mu_0$. Let $f^\ast := f_{\mu}^\ast$ for $\mu \in \Phi^*$. Finally, let

$$F(\phi) = \left\{ (\phi - c(t), f(t)) : t \in \mathcal{T}, c(t) \leq \phi \right\}$$

be the surplus pairs which can be attained by revealing full information only once at time $t$ with probability 1 when the value of information to the DM is $\phi$.

**Proposition 6.** For any cost function $c$, value function $f$, and prior $\mu_0$,

(i) all designer optimal structures leave the DM with zero surplus i.e., $(0, f_{\mu_0}^\ast) \in \mathcal{P}$

(ii) $\mathcal{P} \leq \overline{\mathcal{P}}$ where

$$\overline{\mathcal{P}} = \text{co} \left( F(\phi^\ast) \cup \{(0, f^\ast)\} \right)$$

and $\text{co}(\cdot)$ denotes the concave envelope. Furthermore, $\mathcal{P} = \overline{\mathcal{P}}$ when $\mu_0 \in \Phi^*$.

Part (i) states that even in nonlinear environments, every designer-optimal design extracts all available surplus from the consumer—the DM’s dynamic incentive compatibility constraints are no bulwark against full surplus extraction. Part (ii) gives a partial characterization of the Pareto frontier and, in so doing, illustrates the power of information to shape welfare outcomes in markets for attention.

Figure 8 illustrates part (ii) of Proposition 6 for the cases where $f$ is either more concave (panel (a)) or more convex (panel (b)) than $c$. First consider the case where $f$ is more concave than $c$—for instance, $f$ might be linear and $c$, the disutility from viewing ads, might be convex. In this case, the Pareto frontier is concave and there are ‘gains from trade’ whenever transfers are possible and utilities are quasilinear in money, the socially efficient outcome is typically obtained by an interior point on the frontier which can be implemented by giving the DM full information at a fixed, deterministic time, yet leaves her with positive surplus. This might indicate that a regulator concerned with social welfare might wish to regulate the choice of information structure directly. Now consider the case in which $f$ is more concave than $c$—for instance, $f$ might be concave if there are diminishing returns to advertising. In this case, we

---

32 Hébert and Zhong (2022) independently find that in an environment in which the sender’s constrained entropy minimization is effectively linear in the DM’s stopping time and the DM’s costs are constant, the DM receives zero utility. This is a straightforward consequence of the exposition above.

33 For $\mu, \mu' \in \Phi^*$. $f_{\mu'}^\ast \neq f_{\mu'}^\ast$. To see this, observe that if $\mu_0 \in \Phi^*$, we can relax the obedience constraint by replacing $\phi(\mu_0)$ by $\phi'$ and eliminate the boundary constraint. Solving the relaxed optimization problem gives the same solution because we can choose $\mu_t = \mu_0$ for every $t \in \mathcal{T}$. Thus, the optimal designer’s value does not depend on belief $\mu_0$ whenever $\mu_0 \in \Phi^*$.

34 $\tau \in \text{argmax}_{\tau \in \mathcal{T}} \mathbb{E}[f(\tau)]$ implies $\sup_{\tau \in \mathcal{T}} \mathbb{E}[v(a, \theta, \tau)] = 0$.

35 This is true for any $\mu_0$: when $f$ is more concave than $c$, we always have $\mathcal{P} = F(\phi(\mu_0))$. 
see that the Pareto frontier is linear. If $f^{**}$ is in absolute amount greater than $\phi^*$ i.e., the value of attention to the designer is greater than the DM’s value for information, then the Pareto frontier is obtained by letting the designer extract the DM’s attention, then making transfers ex-post. This might offer a tentative argument for regulators to leave the choice of information structure to the designer.

The proof of Proposition 6 is technical and deferred to Appendix C; we give a rough sketch of part (i) here. The first step is to suppose, towards a contradiction, that the DM strictly prefers to pay attention at period 0 (i.e., she obtains strictly positive surplus) yet the designer is optimizing some value function $f$. We then explicitly construct a new stopping time supported by a new belief path which (i) fulfills the obedience and boundary constraints; and (ii) the sender’s expected value is strictly higher. The second step is to note that fixing a stopping time $\tau$, simple and obedient structures maximizes the DM’s utility. This allows us to extend the result from simple and obedient information structures to all designer-optimal information structures.

We have thus far developed general properties of structures which implement feasible stopping times, as well as some welfare implications when a designer chooses the information structure. We now give an explicit characterization of the information structures which solve the designer’s problem.

### 4.4 Designer’s value function is more concave than DM’s cost

We begin with the simple case in which the designer’s value function is more concave than the DM’s cost function. This nests cases in which both the designer’s value function and the DM’s cost function are both linear, variants of which are studied by Knoeple (2020) and Hébert and Zhong (2022).

---

36This is for $\mu_0 \in \Phi^*$; for $\mu \notin \Phi^*$, we conjecture that the Pareto frontier takes the same form as $\bar{P}$ but replacing $f^{**}$ with $f^{*}_{\mu_0}$ and $F(\phi^*)$ with $F(\phi(\mu_0))$ though we have not been able to prove it.
Because time is discrete, define \( c(\mathcal{T}) := \{ c(t) : t \in \mathcal{T} \} \) as the range of the DM’s cost function \( c \), and define the lower and upper inverses \( \overline{c}^{-1} : \mathbb{R} \to \mathcal{T} \) and \( \underline{c}^{-1} : \mathbb{R} \to \mathcal{T} \) as follows. For any \( x \in \mathbb{R} \),

\[
\overline{c}^{-1}(x) := \inf \{ t \in \mathcal{T} : c(t) \geq x \} \quad \text{and} \quad \underline{c}^{-1}(x) := \sup \{ t \in \mathcal{T} : c(t) \leq x \}.
\]

**Proposition 7** (Designer-optimal: \( f \) more concave than \( c \)). Suppose the sender’s value function is less convex than the DM’s cost function i.e., \( f \circ c^{-1} \) is concave.

(i) For any \( \epsilon > 0 \),

\[
\max_{I \in \mathcal{I}} \mathbb{E}[f(\tau)] - f(\overline{c}^{-1}(\phi(\mu_0))) \leq \epsilon.
\]

(ii) A dynamic information structure which attains \( f(\overline{c}^{-1}(\phi(\mu_0))) \) for the sender has the following form: give full information with full probability at time \( \overline{c}^{-1}(\phi(\mu_0)) \), and no information before that; if \( c \) is convex, the stopping times induced by this information structure also second-order stochastically dominates every other distribution in \( \mathcal{D} \).

Proposition 7 states that whenever costs are additively separable and the designer’s value function is less convex than the DM’s cost function, an optimal dynamic information structure gives full information with full probability at a fixed time in the future, and no information at other times. The \( \epsilon \)-bound in part (i) is arbitrary since we assumed that \( |f(t + 1) - f(t)| \) and \( |c(t + 1) - c(t)| \) can be made arbitrarily small by interpolating and a suitable time change.

**Proof of Proposition 7** We first derive an upper bound on the maximum expected stopping time achievable across all dynamic information structures. Recall we defined the functional \( \phi(\mu) := \mathbb{E}_\mu[u^*(\theta)] - \max_{a \in \mathcal{A}} \mathbb{E}_\mu[u(a, \theta)] \) as the value of full information under belief \( \mu \). Now considering the DM’s incentives at time 0, we have the following necessary condition for \( \tau \) to be strictly positive:

\[
\phi(\mu_0) \geq \mathbb{E}^I[c(\tau)] = \mathbb{E}^I[c \circ f^{-1} \circ f(\tau)] \geq (c \circ f^{-1})\left(\mathbb{E}^I[f(\tau)]\right)
\]

where the last inequality is from Jensen.

This implies that across all information structures \( I \in \mathcal{I} \), \( \mathbb{E}^I[f(\tau)] \leq f(\overline{c}^{-1}(\phi(\mu_0))) \). Then observe that by our assumption that \( f \) is ‘sufficiently smooth’, \( |f(\overline{c}^{-1}(\phi(\mu_0))) - f(\underline{c}^{-1}(\phi(\mu_0)))| \leq \epsilon \) which gives the upper bound. Now consider the information structure \( I’ \) which gives full information at time \( \overline{c}^{-1}(\phi(\mu_0)) \) and no information otherwise. Clearly \( \phi(\mu_0) \geq \mathbb{E}^{I’}[c(\overline{c}^{-1}(\phi(\mu_0)))] \) so it is incentive compatible for the DM to continue paying attention at time 0. Furthermore, it is easy to see that for any time \( 0 < t < \overline{c}^{-1}(\phi(\mu_0)) \), it remains incentive compatible to continue paying attention. But then this implements the distribution which puts full probability on \( \overline{c}^{-1}(\phi(\mu_0)) \) which gives the lower bound in part (i).

For part (ii), the first part has already been proven. To see that the stopping time \( \tau(\tau’) \) induced by this information structure \( I’ \) second-order stochastically dominates everything in \( \mathcal{D} \), notice
that when \( c \) is convex and \( f \) is the identity map, \( f \circ c^{-1} = c^{-1} \) is concave so the condition above holds, and the the information structure attains \( \sup_t \mathbb{E}^t[\tau] \). But since the distribution of \( \tau(I') \) puts a full probability on \( c^{-1}(x) \), we are done.

The proof of Proposition 7 relied on showing that the period 0 obedience constraint can be made tight. Indeed, this is also employed by Knoeple (2020) and Hébert and Zhong (2022) to study settings in which the DM’s cost and designer’s value are linear. Here our formulation is slightly more general to handle cases in which the designer’s value function is more concave than the DM’s cost. Nonetheless, this approach is quite special and does not generalize to non-linear cost and value functions which we now turn to.

4.5 Arbitrary value and cost functions. For much of the analysis, we will assume that \( \mu_0 \in \Phi^* \) i.e., the DM’s prior belief begins in the basin of uncertainty. This assumption allows us to optimize over stopping times without considering belief paths. There are several ways to interpret this. The literal interpretation is that the DM’s prior is in fact in the basin. This might correspond to cases in which the basin of uncertainty is large as when the DM might have a safe option to guarantee herself a fixed utility and maximum payoffs are the same across states or when the prior is uniform and actions are symmetric across states. Alternatively, we might think of starting in the basin as the second half of an optimal dynamic structure i.e., an optimal structure conditional on reaching the basin. Indeed, we know from Propositions 4 and 5 that it is without loss to consider belief paths which move towards the basin, and that beliefs are guaranteed to reach the basin as time grows large.

A starting point for our analysis will be to rewrite the obedience constraint explicitly in terms of the distribution of stopping times and the DM’s cost function.

**Lemma 1.** If \( \mu_0 \in \Phi^* \) and \( \tau \) is a stopping time, the following statements are equivalent:

(i) \( \tau \) is feasible.

(ii) The obedience constraint holds for every period: \( \phi^* \geq \mathbb{E}[c(\tau) \mid \tau > t] - c(t) \) for every \( t \in \mathcal{T} \).

(iii) There exists a sequence \( (C_t)_{t \in \mathcal{T}} \in [0, \phi^*] \) such that

\[
\mathbb{P}(\tau > t + 1 \mid \tau > t) = \frac{C_t + c(t) - c(t + 1)}{C_{t+1}},
\]

for every \( t \in \mathcal{T} \).

---

37 As we saw in Proposition 3, belief paths impose restrictions on stopping times and vice versa. By part (a) of Proposition 5, \( \mu_t \in \phi^* \) allows us to pick \( \mu_t = \mu_0 \) for all \( t \in \mathcal{T} \) without loss and focus on optimizing over stopping times. We leave the general case of solving the sender’s problem for arbitrary \( f, c \), and \( \mu_0 \) for future work.

38 For instance consider a binary state, binary action setting in which the DM obtains payoff 1 if she matches her action to the state, and zero otherwise. Now suppose the DM can additionally take a safe action which delivers payoff 0.75 in all states. The basin of uncertainty is then \([0.25, 0.75]\).

39 Formally, for every bijection of states \( \sigma_0 : \Theta \rightarrow \Theta \), there exists a bijection of actions \( \sigma_A : A \rightarrow A \) such that \( u(a, \theta) = u(\sigma_A(a), \sigma_0(\theta)) \) for every \( \theta \in \Theta \) and \( a \in A \). This is fulfilled by the binary state, binary action setting.
When part (iii) of the lemma is fulfilled, we note that the associated stopping time \( \tau \) is such that \( C_t = \mathbb{E}[c(\tau) | \tau > t] - c(t) \). \( C_t \) then takes on the interpretation as the additional expected cost of paying attention until the DM learns the state. Our approach will be to work directly over sequences \((C_t)_{t \in T}\) and back out feasible stopping times associated with it. Note that a sequence \((C_t)_{t \in T}\) induces a well-defined stopping time if and only if \( \frac{C_{t+1} + c(t) - c(t+1)}{C_{t+1}} \in [0, 1] \) and \( C_t \in [0, \phi^*] \) which can be equivalently restated as
\[
\begin{align*}
(i) \quad & C_t - C_{t+1} \leq c(t + 1) - c(t); \\
(ii) \quad & C_t \in \left[ c(t + 1) - c(t), \phi^* \right].
\end{align*}
\]

We say that a sequence \((C_t)_{t \in T}\) is an obedient sequence if the above two conditions hold. The following algorithm constructs sequences \((C_t)_{t \in T}\) corresponding to feasible stopping times which solve the designer’s problem for any value function \( f \) and any DM cost function \( c \).

**Definition 5** (Algorithm generating designer-optimal structures). Take an increasing sequence of times \((t_i)_{i=1}^n \in T^n\). We allow \( n = +\infty \) but if \( n < +\infty \), additionally choose a pair of terminal times \( t_{n+1}, \bar{t} \) where \( t_m < t_{m+1} < \bar{t} \) such that \( \phi^* + c(t_{n+1}) \geq c(\bar{t}) \geq \phi^* + c(t_n) \geq c(t_{m+1}) \). Choose \((C_t)_{t \in T}\) as follows:
\[
(i) \quad \text{(Construction of indifference points)} \quad \text{Set } C_0 = \phi^*, \quad C_{t_i} = \phi^* \text{ for every } i \in \{1, \ldots, n\}, \quad \text{and } C_{t_{n+1}} = c(\bar{t}) - c(t_{n+1}).
\]
\[
(ii) \quad \text{(Construction of terminal stage)} \quad \text{For } 1 \leq t \leq \bar{t}, \text{ define } i(t) := \max\{i \in \{0, 1, \ldots, n + 1\} : t_i \leq t\}, \text{ where } t_0 = 0, \text{ and set } C_t = C_{i(t)} - \left( c(t) - c(t_{i(t)}) \right).
\]

It is easy to verify that the resultant sequence \((C_t)_t\) is obedient and hence, from Lemma 1, corresponds to a stopping time \( \tau \) which is feasible.

Define \( \mathcal{A} : T^{Inc} \rightarrow \mathcal{I}^* \) as the map which takes in a finite or infinite increasing subsequence which takes values in \( T^{[0]} \) applies the algorithm above, and returns the corresponding simple and obedient information structure. Information structures generated by \( \mathcal{A} \) are comprised of two stages.
\[
(i) \quad \text{(Indifference stage)} \quad \text{This stage runs from times } 0 \text{ to } t_n, \text{ and is comprised of blocks—consecutive time periods. For any } i = 1, 2, \ldots, n - 1, \text{ the periods } t_i, t_{i+1}, \ldots, t_{i+1} \text{ comprise a single block. For the periods interior to each block (e.g., if the block is between } t_i \text{ and } t_{i+1}, \text{ these are the periods } t_i + 1, \ldots, t_{i+1} - 1, \text{ the DM receives no information. To see this, notice that we write } C_t \text{ as chosen by the algorithm as}
\]
\[
\mathbb{E}[c(\tau) | \tau > t] - c(t) = C_t = C_{i(t)} - \left( c(t) - c(t_{i(t)}) \right)
\]
\[
= \mathbb{E}[c(\tau) | \tau > t_{i(t)}] - c(t)
\]

\footnote{If the subsequence is finite, we require, as in the algorithm, that } \( \bar{t} \) \text{ be such that } \phi^* + c(t_n) \leq c(\bar{t}) \leq \phi^* + c(t_{n+1}).
so between periods $t_i$ and $t$, there is zero probability the DM learns the state. Furthermore, at boundaries of each block (e.g., if the block is $t_i \ldots t_{i+1}$, these are the times $t_i$ and $t_{i+1}$), the DM is indifferent between stopping and continuing to pay attention. To see this, notice that we set $C_{t_i} = \mathbb{E}[c(\tau) | \tau > t] - c(t) = \phi^*$ for all $i = 1, 2, \ldots n$. In light of the above discussion, we note that the probability the DM learns the state at times $t_{n} = 1$ is pinned down:

$$P(\tau > t_i | \tau > t_i - 1) = \frac{\phi^* + c(t_{i-1}) - c(t_i)}{\phi^*}.$$ 

Figure 9 illustrates the indifference stage.

(ii) (Terminal stage if $n < +\infty$) This stage runs from times $t_n + 1$ to $\bar{t}$, and offers the DM two opportunities to obtain full information: once with positive probability at $t_{n+1}$, and once with probability one at time $\bar{t}$. The designer faces the following tradeoff: increasing the probability that full information is revealed at time $t_{n+1}$ pushes down the probability that the DM pays attention until time $\bar{t}$; on the other hand, doing so allows us the designer to choose a $\bar{t}$ further into the future. Indeed, for a given pair $(t_{n+1}, \bar{t})$, the distribution of stopping times is completely pinned down:

$$P(\tau > t_{n+1} | \tau > t_{n+1} - 1) = \frac{\phi^* + c(t_{n+1}) - c(t_{n+1})}{c(\bar{t}) - c(t_{n+1})}$$

$$P(\tau > \bar{t} | \tau > \bar{t} - 1) = 0.$$ 

Figure 10 illustrates the terminal stage.

Let $\mathcal{A}(T^{inc})$ be the set of information structures generated by this algorithm. Such information structures exploit the fact that conditioned on paying attention up to time $t$, the DM’s costs are already sunk: her continuation incentives depend only on her current belief (which
Figure 10: Illustration of terminal stage

Proposition 8 (Designer-optimal: \( f \) and \( c \) arbitrary). If \( \mu_0 \in \Phi^\ast \) then for every strictly increasing value function \( f : T \to \mathbb{R} \), there exists \( I \in \mathcal{A}(T^{\text{inc}}) \) such that \( E^{I}[f(\tau)] = \max_{I' \in \mathcal{I}} E^{I'}[f(\tau)] \).

Proposition 8 states that the set of distributions which are sufficient for maximization under any increasing value function is generated by simple and obedient structures of the form above. Such information structures comprises blocks for which the DM receives no information with probability one in the interior, and receives full information with positive probability as to induce indifference on the boundaries. The block structure generated by the algorithm admits a natural interpretation in the context of platforms: the DM faces a random number of advertisements. Each advertisement corresponds to a single block, and whenever the advertisement ends, the DM has some chance to learn the information she wants. The platform controls (i) the duration of the advertisements (length of the block) as well as the (ii) probabilities at which users are made to watch consecutive advertisements (probability that the DM learns the state at the end of each block). Indeed, the current ad algorithm on Youtube, a popular video streaming platform, shows users a countdown for the current ad they are being shown, but not the total number of ads they will be shown; only at the end of every advertisement, does the DM learn whether she has to watch another ad—but by then her costs are already sunk.

The solution to the designer’s problem considered above, in which the value function is less convex than the cost function can be viewed as an extreme case of the block structure: we skip the indifference stage (no indifference blocks) by picking \( t_{n+1} = 0 \), and \( \tau \) as the value which makes the DM indifferent between continuing to pay attention and stopping at period \( t_n = 0 \). We now show that when the designer’s value function is more convex than the DM’s
cost function, the solution is implemented through the other extreme: we have no terminal stage, and pick \( (t_i)_{i=1}^{\infty} = \mathcal{T} \) i.e., \( t_1 = 1, t_2 = 2, \ldots \) for the algorithm above i.e., at every time \( t \), conditional on receiving the null message up to \( t \), the DM is indifferent between stopping and continuing. We denote the resultant information structure with \( \mathcal{A}(\mathcal{T}) \) and note that under this structure, the probability that the DM pays attention at time \( t+1 \) given that she has already paid attention up to time \( t \) is

\[
\mathbb{P}(m_0 \mid (m_0)^t) = 1 - \frac{c(t+1) - c(t)}{\delta^*}.
\]

**Proposition 9** (Designer-optimal: \( f \) more convex than \( c \)). Suppose the the sender’s value function \( f \) is is more convex than the DM’s cost function \( c \) i.e., \( f \circ c^{-1} \) is convex and \( \mu_0 \in \Phi^* \). Then

\[
\mathbb{E}^{\mathcal{A}(\mathcal{T})}[f(\tau)] = \max_{I \in \mathcal{I}} \mathbb{E}^I[f(\tau)].
\]

We defer the proof of Proposition 9 to Appendix C though we sketch out the main ideas here. From Proposition 8, we know that a solution to the problem takes a block structure. We then show that it is suboptimal for blocks to have length greater than 1 i.e., have a non-empty interior. Fixing an original sequence of times \( \{t_i\}_i \), suppose that there exists some \( k \) such that \( t_{k+1} - t_k > 1 \). Now consider instead an alternate sequence \( \{t_i'\}_i \) such that (i) \( t_i' = t_i \) for all \( i \leq k \); (ii) \( t_{k+1}' = t_{k+1} - 1 \); and (iii) \( t_i' = t_{i-1} \) for all \( i > k + 1 \) i.e., we simply add an extra element at time \( t_{k+1} - 1 \). This is depicted in Figure 11 (a).

Figure 11: Illustration of the modification from \( \{t_i\} \) to \( \{t_i'\} \)

Let \( \tau = \tau(\mathcal{A}(\{t_i\}_i)) \) be the stopping times corresponding to the original sequence, and \( \tau' = \tau(\mathcal{A}(\{t_i'\}_i)) \) be the new stopping time. Our modification has the following effect on the distribution of stopping times: the DM now has an extra opportunity to learn the state at time \( t_{k+1} - 1 \). This weakens the obedience constraints for all earlier times. As such, the designer
now has an extra degree of freedom to increase the probability that the DM pays attention beyond times $t_{k+1} + 1$. Finally, to see the effect of this modification on the designer’s expected value, we note that $\phi^* - E[c(\tau)] = \phi^* - E[c(\tau')] = 0$. Then since $\phi$ is more convex than $c$, we can conclude that $\tau'$ is preferred by the designer than $\tau$. This is depicted in Figure 11 (b).

4.6 S-shaped value functions. We now explicitly solve the case in which the designer’s value function is S-shaped and the DM faces linear costs.

**Definition 6** (S-shaped value function). $f$ is S-shaped if there exists $t^*$ such that $f(t + 1) - f(t) < f(t) - f(t - 1)$ if and only if $t > t^*$.

S-shaped value functions are those which exhibit increasing differences up to $t^*$, and decreasing differences after $t^*$. An interpretation of S-shaped functions is that users are initially unresponsive to advertising but, after crossing a threshold (consumer ‘wear in’), begin to respond strongly; at some point they become saturated and their demand once again tapers off (‘wear out’).

**Definition 7.** (Conditional concavification) Suppose $f$ is S-shaped. For every time $t \in T$, if $s^*(t) := \min \{ s > t : f(s + 1) - f(s) < \frac{f(s) - f(t)}{s - t} \}$ exists, define the conditional concavification from $t$, $\text{cconv}_t(f) : \{ s \in T : s \geq t \} \to \mathbb{R}_+$ such that

$$
\text{cconv}_t(f)(s) = \begin{cases} 
\frac{s^*(t) - s}{s^*(t) - t} f(t) + \frac{s - 1}{s^*(t) - t} f(s^*(t)), & \text{for } s \leq s^*(t) \\
\frac{s - 1}{s^*(t) - t} f(s), & \text{otherwise.}
\end{cases}
$$

$c\text{conv}_t(f)$ is depicted in Figure 12 (a) for $t = t_1$ (red) and $t = t_2$ (blue).

**Proposition 10** (Designer-optimal: $f$ is S-shaped). Suppose the sender’s value function $f$ has S-shaped and $\phi^*/c$ is an integer. Define $\hat{t} = \min \{ t \in T : s^*(t) - t \leq \phi^*/c \}$. Then, there exists $t_* \in \{ \hat{t}, \hat{t} - 1 \}$ such that a sender’s optimal structure is induced by the algorithm which takes in an increasing sequence $1, 2, \ldots, t_*$ with a terminal time $\hat{t} = t_* + \phi^*/c$.

Proposition 10 explicitly solves for the designer’s optimal structure when $f$ is S-shaped which is depicted in Figure 12 (b). The proof is deferred to Online Appendix II. The first stage runs up to time $t_*$ comprises indifference blocks of length 1 i.e., the DM has some chance to learn the state at times $t = 0, 1, \ldots, t_*$. Conditional on having paid attention up to time $t_*$, the DM then receives no information with probability 1 for the times $t_* + 1, t_* + 2, \ldots t_* + \phi^*/c - 1$, and then learns the state with probability 1 at time $t_* + \phi^*/c$.

\[ More explicitly, the distribution of $\tau$ for times $t \leq t_k$ is unchanged, and the probability that $\tau > t_k$ is similarly unchanged. The law of total expectation gives the result.

\[ S-shaped response functions have been enormously influential within economics, marketing, and consumer psychology. The reader is referred to Fennis and Stroebele (2015) and references therein for more details.

\[ Optimizing over S-shaped response functions has antecedents in the operations research literature (Freeland and Weinberg 1980). However, the dynamic obedience constraints in our problem poses additional complications which we handle in Appendix III. \]
5 Concluding remarks

We have studied the set of feasible distributions of stopping times attainable through information. This presents a benchmark for settings in which the DM does not control the dynamic information structure she faces. Our reduction principle, as well as the results of increasing and extremal belief paths, show that every stopping time can be implemented through information structures which special properties: they either give the DM full information, or give the DM the null message which increases the DM’s uncertainty and hence value for information. We used this class of structures to solve the problem of a platform who might be interested in maximizing any increasing function of the DM’s stopping time and showed that optimal structures are block structures which exploit the observation that conditional on paying attention up to time $t$, the DM’s costs are already sunk. We believe there is more work to be done in this direction. In particular, we have focused only on the case in which the designer’s preferences are over $\Delta(\mathcal{T})$. While we think that this is a natural starting point for understanding platforms’ incentives to capture users’ attention, platforms might raise revenue from both persuasion (DM’s action $a \in A$) e.g., the choice of which product to buy, as well as attention (DM’s stopping time $\tau \in \mathcal{T}$) e.g., time spent viewing advertisements. We leave the question of what can be implemented within $\Delta(\mathcal{T} \times A \times \Theta)$ for future work.

References

ARROW, K. J., D. BLACKWELL, AND M. A. GIRSHICK (1949): “Bayes and minimax solutions of sequential decision problems,” Econometrica, Journal of the Econometric Society, 213–244.

AUSTEN-SMITH, D. AND C. MARTINELLI (2018): “Optimal exploration.”
BALL, I. (2019): “Dynamic information provision: Rewarding the past and guiding the future,” Available at SSRN 3103127.

BEMMAOR, A. C. (1984): “Testing alternative econometric models on the existence of advertising threshold effect,” Journal of Marketing Research, 21, 298–308.

BERGEMANN, D., B. BROOKS, AND S. MORRIS (2015): “The limits of price discrimination,” American Economic Review, 105, 921–57.

——— (2017): “First-price auctions with general information structures: Implications for bidding and revenue,” Econometrica, 85, 107–143.

BERGEMANN, D. AND S. MORRIS (2013): “Robust predictions in games with incomplete information,” Econometrica, 81, 1251–1308.

CHE, Y.-K. AND K. MIERENDORFF (2019): “Optimal dynamic allocation of attention,” American Economic Review, 109, 2993–3029.

CORRAO, R., D. FUDENBERG, AND D. LEVINE (2022): “Adversarial forecasters, surprises and randomization,” Tech. rep., David K. Levine.

DEBREU, G. (1960): “Topological methods in cardinal utility theory, Mathematical methods in the social sciences (KJ Arrow, S. Karlin, and P. Suppes, eds.),”.

DUBÉ, J.-P., G. J. HITSCH, AND P. MANCHANDA (2005): “An empirical model of advertising dynamics,” Quantitative marketing and economics, 3, 107–144.

DWORKOZAK, P. AND A. KOLOTILIN (2022): “The persuasion duality,” arXiv preprint arXiv:1910.11392.

ELY, J., A. FRANKEL, AND E. KAMENICA (2015): “Suspense and surprise,” Journal of Political Economy, 123, 215–260.

ELY, J. C. (2017): “Beeps,” American Economic Review, 107, 31–53.

ELY, J. C. AND M. SZYDLOWSKI (2020): “Moving the goalposts,” Journal of Political Economy, 128, 468–506.

FENNIS, B. M. AND W. STROEBE (2015): The psychology of advertising, Psychology Press.

FISHBURN, P. C. AND A. RUBINSTEIN (1982): “Time preference,” International economic review, 677–694.

FREELAND, J. R. AND C. B. WEINBERG (1980): “S-Shaped response functions: Implications for decision models,” Journal of the Operational Research Society, 31, 1001–1007.

FUDENBERG, D., P. STRACK, AND T. STRZALECKI (2018): “Speed, accuracy, and the optimal timing of choices,” American Economic Review, 108, 3651–84.

GENTZKOW, M. AND E. KAMENICA (2016): “A Rothschild-Stiglitz approach to Bayesian persuasion,” American Economic Review, 106, 597–601.

GIBBARD, A. (1973): “Manipulation of voting schemes: a general result,” Econometrica: journal of the Econometric Society, 587–601.

GOSSNER, O., J. STEINER, AND C. STEWART (2021): “Attention please!” Econometrica, 89, 1717–1751.

GUO, Y. AND E. SHMAYA (2018): “The value of multistage persuasion,” Tech. rep., Working paper.

HÉBERT, B. AND W. ZHONG (2022): “Engagement Maximization,” arXiv preprint arXiv:2207.00685.

KAMENICA, E. AND M. GENTZKOW (2011): “Bayesian persuasion,” American Economic Review, 101, 2590–2615.
Knoepfle, J. (2020): “Dynamic Competition for Attention,” Tech. rep., University of Bonn and University of Mannheim, Germany.

Kolotilin, A., R. Corrao, and A. Wolitzky (2022): “Persuasion with Non-Linear Preferences.”

Liang, A., X. Mu, and V. Syrgkanis (2022): “Dynamically aggregating diverse information,” *Econometrica*, 90, 47–80.

Morris, S. and P. Strack (2019): “The wald problem and the relation of sequential sampling and ex-ante information costs,” *Available at SSRN 2991567*.

Myerson, R. B. (1982): “Optimal coordination mechanisms in generalized principal–agent problems,” *Journal of mathematical economics*, 10, 67–81.

——— (1986): “Multistage games with communication,” *Econometrica: Journal of the Econometric Society*, 323–358.

Pomatto, L., P. Strack, and O. Tamuz (2018): “The cost of information,” *arXiv preprint arXiv:1812.04211*.

Renault, J., E. Solan, and N. Vieille (2017): “Optimal dynamic information provision,” *Games and Economic Behavior*, 104, 329–349.

Sims, C. A. (2003): “Implications of rational inattention,” *Journal of monetary Economics*, 50, 665–690.

Stiglitz, J. E. and M. Rothschild (1970): “Increasing risk: I. A definition,” *Journal of Economic Theory*, 2, 225–243.

Treves, F. (1990): *Topological vector spaces, distributions and kernels*, Academic Press.

Vakratsas, D., F. M. Feinberg, F. M. Bass, and G. Kalyanaram (2004): “The shape of advertising response functions revisited: A model of dynamic probabilistic thresholds,” *Marketing Science*, 23, 109–119.

Wald, A. (1947): “Foundations of a general theory of sequential decision functions,” *Econometrica, Journal of the Econometric Society*, 279–313.

Wu, W. (2011a): “Attention as selection for action,” *Attention: Philosophical and psychological essays*, 97.

——— (2011b): “Confronting Many-Many problems: Attention and agentive control,” *Noûs*, 45, 50–76.

Zhong, W. (2022): “Optimal dynamic information acquisition,” *Econometrica*, 90, 1537–1582.
Appendix to ‘Attention Capture’
Andrew Koh and Sivakorn Sanguanmoo

Outline of Appendix. Appendix A proves the reduction principle. Appendix B collects the other proofs in Section 3 which were omitted from the main text. Appendix C collects the proofs in Section 4 which were omitted from the main text.

Appendix A: Proof of Proposition 1 (Reduction principle)

A.1 Preliminaries. Sequential formulation. We first develop a sequential formulation of the DM’s optimization problem.

Equivalently, the problem can be formulated as a sequential optimization problem: for any history \( H_t \), the DM solves the following optimization problem:

\[
U(I) := \sup_{\tau, a_\tau} \mathbb{E}^I [v(a_\tau, \theta, \tau) \mid H_t],
\]

where \( \tau \) is a stopping time, \( a_\tau \) is a stochastic action under the natural filtration, and \( \mathbb{E}^I [\cdot \mid H_t] \) is the conditional expectation under information structure \( I \in \mathcal{I} \) after history of messages \( H_t \in \mathcal{H}_t \).

Throughout, we will assume that the DM breaks indifferences in favour of not stopping. As such, given the history \( H_t \), the DM will stop paying attention if and only if the expected utility from taking action right away, \( \max_{a \in A} \mathbb{E}^I [v(a, \theta, t) \mid H_t] \), is strictly greater than the continuation payoff \( \mathbb{E}^I [U_I(H_{t+1}) \mid H_t] \). This dynamic relationship between \( U_I(H_t) \) and \( U_I(H_{t+1}) \) is captured in the following recursive equation:

\[
U_I(H_t) = \max \left\{ \max_{a \in A} \mathbb{E}^I [v(a, \theta, t) \mid H_t], \; \mathbb{E}^I [U_I(H_{t+1}) \mid H_t] \right\}.
\]

Notation for histories. It will also be helpful to develop notation to keep track of different histories.

Definition 8 (Feasible histories). \( H_t \in \mathcal{H}_t \) is a feasible history under \( I \) if

(i) \( \mathbb{P}^I (H_t) > 0 \); and

(ii) for all \( s < t \), \( \max_{a \in A} \mathbb{E}^I [v(a, \theta, s) \mid H_s] \leq U_I(H_s) \)

where \( H_s \subseteq H_t \) is the history \( H_t \) truncated at time \( s \).

Condition (i) states that in order for \( H_t \) to be feasible, the messages comprising that history must happen with positive probability; condition (ii) states that the DM must find it optimal to reach \( H_t \): at every history along the path \( (H_s)_{s=1}^{t-1} \) of histories leading up to \( H_t \), she must find it optimal to continue paying attention. We denote the set of all feasible histories under information structure \( I \) with \( \mathcal{H}^F(I) \subseteq \bigcup_{t \in \mathcal{T}} \mathcal{H}_t \). Further define the time \( t \) feasible histories with \( \mathcal{H}^F_t(I) := \mathcal{H}^F(I) \cap \mathcal{H}_t \).
Definition 9 (Optimal stopping histories). $H_t \in \mathcal{H}_t$ is an optimal stopping history under $I$ if

(i) $H_t$ is feasible under $I$; and

(ii) $\max_{a \in A} \mathbb{E}^I[v(a, \theta, t) \mid H_t] > \mathbb{E}^I[U(H_{t+1}) \mid H_t]$.

Optimal stopping histories require, in addition to feasibility, that the DM strictly prefers to stop and take action on history $H_t$ (recalling that we broke indifferences in favour of not stopping). Define $\mathcal{H}^O(I) \subseteq \mathcal{H}^F(I) \subseteq \bigcup_{t \in T} \mathcal{H}_t$ as the set of all optimal stopping histories with respect to the information structure $I$. Further define the time $t$ optimal stopping histories with $\mathcal{H}^O_t(I) := \mathcal{H}^O(I) \cap \mathcal{H}_t$.

A.2 Proof of Proposition 1

Proof of Proposition 1. Start from an arbitrary information structure $I$ defined over the message space $M$. For convenience, we will embed new messages $\{m_\theta\}_{\theta \in \Theta} \cup \{m_\phi\}$ into $M$ and work on this augmented message space $M^* := M \cup \{m_\theta\}_{\theta \in \Theta} \cup \{m_\phi\}$.

Further define $\mathcal{H}_t = \prod_t M^*$ as the set of time $t$ histories on the augmented message space. Doing so will allow us to modify $I$ while fixing the message space.

Our proof proceeds in two steps. Step 1 modifies the optimal stopping histories of $I$ and we denote the resultant dynamic information structure with $I'$. Step 2 collapses the branches of $I'$ and we denote the resultant simple dynamic information structure with $I'' \in I^*$. We will show that $d(I) = d(I'')$.

Recall that $I \in I$ is just a collection of probability distributions

$$I = \left\{ p_t^I(\cdot \mid \theta, H_t) \right\}_{t \in T, \theta \in \Theta, H_t \in M}$$

where throughout we use $p_t^I(\cdot \mid \theta, H_t)$ to denote the distribution of message $m$ under information structure $I$, at time $t$, state $\theta$, and on history $H_t$.

Step 1: Modify optimal stopping histories $\mathcal{H}^O(I)$ on the original structure $I$.

Construct a new information structure $I'$ from $I$ as follows:

(i) For all times $t$, all histories $H_t \in M'$ and all states $\theta, \theta' \in \Theta$, $\theta \neq \theta'$, set

$$p_t^{I'}(m_{\theta'} \mid \theta, H_t) = \sum_{m \in M: (H_t, m) \in \mathcal{H}_{t+1}^O(I)} p_t^I(m \mid \theta, H_t),$$

$$p_t^{I'}(m_{\theta'} \mid \theta, H_t) = 0, \quad p_t^{I'}(m \mid \theta, H_t) = 0 \quad \text{for all } m \in M \text{ if } (H_t, m) \in \mathcal{H}^O(I).$$

---

44This is without loss as long as $|M| \geq |\Theta| + 1$ since simple information structures only use at most $|\Theta| + 1$ messages and can be relabelled. Here we leave the messages $\{m_\theta\}_{\theta \in \Theta} \cup \{m_\phi\}$ abstract, and assume that they were not part of the original messages space $M$ i.e., for all $\theta \in \Theta$, $\{m_\theta\} \cap M = \emptyset$ and $\{m_\phi\} \cap M = \emptyset$.

45Note that these are histories comprised of messages in the original message space $M$. $p_t^I(m \mid \theta, H_t)$ for $H_t \in \mathcal{H}_t$, the product of our augmented message space, is not always well-defined.
where \((H_t, m) \in \mathcal{H}_{t+1}\) is the history at \(t + 1\) constructed from taking the history \(H_t\) and adding on message \(m\) at time \(t + 1\).

(ii) For all \(m \in M\),

\[
p_t^I(m \mid \theta, H_t) = p_t^I(m \mid \theta, H_t) \quad \text{if } (H_t, m) \not\in \mathcal{H}^O(I)
\]

Part (i) of the modification takes all time \(t + 1\) histories which are optimal stopping, \((H_t, m) \in \mathcal{H}^O(I)\), and combines the probabilities of each of those messages into a single message, \(m_\theta\).

Part (ii) of the modification ensures that \(I'\) is identical to \(I\) on non-optimal stopping histories. One can check that the resultant information structure \(I'\) is well-defined.

Further define \(U^I_t(H_t)\) as the expected utility of DM at history \(H_t\) under the information structure \(I \in I\) if she does not stop until she sees message \(m_\theta\) for some \(\theta \in \Theta\).

**Lemma 2.** The histories which are feasible but not optimal stopping under \(I\) are also feasible but not optimal stopping under \(I'\), i.e.,

\[
\mathcal{H}^F(I) \setminus \mathcal{H}^O(I) = \mathcal{H}^F(I') \setminus \mathcal{H}^O(I')
\]

**Proof of Lemma 2** Our goal is to show that on history \(H_t \in \mathcal{H}^F(I) \setminus \mathcal{H}^O(I)\), the value of continuing to pay attention under the modified information structure, \(\mathbb{E}^{I'}[U_t^I(H_{t+1}) \mid H_t]\) is weakly greater than the value of stopping at the history \(H_t\), \(\max_{a \in A} \mathbb{E}^I[v(a, \theta, t) \mid H_t]\).

We will first show that for any \(H_t \in \mathcal{H}^F(I) \setminus \mathcal{H}^O(I)\), we have

\[
\mathbb{E}^{I'}[U_t^I(H_{t+1}) \mid H_t] \geq \mathbb{E}^{I'}(H_t) \geq U^I(H_t).
\]

The first inequality is immediate from the fact that not stopping until DM sees message \(m_\theta\) for some \(\theta \in \Theta\) is itself a stopping time, and the left hand side the supremum over all stopping times. We now show the second inequality.
\[ U^I_t(H_t) = \sum_{\theta \in \Theta} \sum_{s=t}^{+\infty} \sum_{H_s \in M^t} \mathbb{P}^I(H_s \mid H_t) p^I_s(m_{\theta} \mid \theta, H_s, \sigma^I_t((H_s, m_{\theta}))) \]

\[
= \sum_{\theta \in \Theta} \sum_{s=t}^{+\infty} \sum_{H_s \in M^t} \mathbb{P}^I(H_s \mid H_t) p^I_s(m_{\theta} \mid \theta, H_s) \max_{a \in A} \nu(a, \theta, s + 1) \]

\[
= \sum_{\theta \in \Theta} \sum_{s=t}^{+\infty} \sum_{H_s \in M^t} \mathbb{P}^I(H_s \mid H_t) \sum_{(H_s, m) \in \mathcal{H}^O(I)} p^I_s(m \mid \theta, H_s) \max_{a \in A} \nu(a, \theta, s + 1) \]

\[
= \sum_{\theta \in \Theta} \sum_{s=t}^{+\infty} \mathbb{P}^I((H_s, m) \in \mathcal{H}^O(I) \mid H_t) \max_{a \in A} \nu(a, \theta, s + 1) \]

\[
= \sum_{s=t+1}^{+\infty} \mathbb{E}^I[\max_{a \in A} \nu(a, \theta, s) 1\{H_s \in \mathcal{H}^O(I)\} \mid H_t] \]

\[
= \sum_{s=t+1}^{+\infty} \mathbb{E}^I[1\{H_s \in \mathcal{H}^O(I)\} \mathbb{E}^I[\max_{a \in A} \nu(a, \theta, s) \mid H_s] \mid H_t] \]

\[
\geq \sum_{s=t+1}^{+\infty} \mathbb{E}^I[1\{H_s \in \mathcal{H}^O(I)\} \max_{a \in A} \mathbb{E}^I[\nu(a, \theta, s) \mid H_s] \mid H_t] \]

\[
= \sum_{s=t+1}^{+\infty} \mathbb{E}^I[U^I_t(H_s) 1\{H_s \in \mathcal{H}^O(I)\} \mid H_t], \]

\[
= \mathbb{E}^I[U^I_t(H_{t+1}) \mid H_t] \geq U^I_t(H_t), \]

where the last inequality follows from the fact that \( H_t \) is not an optimal stopping history under \( I \). Here we switched the order of summations because the sum is uniformly bounded by \( \max_{\theta, a} \nu(a, \theta, 0) \). Furthermore,

\[
U^I_t(H_t) \geq \max_{a \in A} \mathbb{E}_t[\nu(a, \theta, t \mid H_t)] = \max_{a \in A} \mathbb{E}_t^I[\nu(a, \theta, t \mid H_t)],
\]

where the inequality is from the definition of \( U^I_t \), and the equality follows because we only modified the optimal stopping histories on \( I \); non-optimal stopping histories are unchanged, so the DM’s beliefs under \( H_t \in \mathcal{H}^F(I) \setminus \mathcal{H}^O(I) \) are the same as that under \( H_t \in \mathcal{H}^F(I') \setminus \mathcal{H}^O(I') \).

Putting both sets of inequalities together, we have that for all \( H_t \in \mathcal{H}^F(I) \setminus \mathcal{H}^O(I) \),

\[
\mathbb{E}^I[U^I_t(H_{s+1}) \mid H_s] \geq \max_{a \in A} \mathbb{E}^I_t[\nu(a, \theta, s) \mid H_s]
\]

so \( H_t \) is not optimal stopping under \( I' \). Conclude by noting that for all \( H_s \subseteq H_t \) (i.e., the history \( H_t \) truncated at time \( s \leq t \)), they are also not optimal stopping under \( I' \). Then since these histories are feasible under \( I \), they must also be feasible under \( I' \). \( \square \)

The next lemma states that the distributions of DM’s stopping times under \( I \) and \( I' \) are identical.

**Lemma 3.** \( d(I') = d(I) \).
Proof. For any history $H_t \in M'$ (histories comprised of messages in the original message space $M$) which is both feasible and not optimal stopping under $I$, Lemma 2 states that it is also feasible and not optimal stopping under $I'$. As such, for every $t \in T$, we have
\[
\mathbb{P}^{I'}(\tau(I') = t | \theta) = \sum_{H_{t-1} \in M'^{t-1}} \mathbb{P}^{I'}(H_{t-1} | \theta) p_t^{I'}(m_0 | \theta, H_{t-1})
\]
\[
= \sum_{H_{t-1} \in M'^{t-1}} \mathbb{P}^{I}(H_{t-1} | \theta) p_t^{I'}(m_0 | \theta, H_{t-1})
\]
\[
= \sum_{H_{t-1} \in M'^{t-1}} \left[ \mathbb{P}^{I}(H_{t-1} | \theta) \sum_{m \in M: (H_{t-1}, m) \in \mathcal{H}_I^0(I)} p_t^{I}(m | \theta, H_{t-1}) \right]
\]
\[
= \sum_{H_t \in M': H_t \in \mathcal{H}_I^0(I)} \mathbb{P}^{I}(H_t | \theta)
\]
\[
= \mathbb{P}^{I}(\tau(I) = t | \theta),
\]
when can be unconditioned on $\theta$ as required.

\[\square\]

Step 2: Collapse non-optimal stopping histories $\mathcal{H}^F(I') \setminus \mathcal{H}^O(I')$.

Given the information structure $I'$ resulting from Step 1, we now construct a simple information structure
\[I'' = \left\{ p_t^{I''} (\cdot | \theta, H_t) \right\}_{t \in T, \theta \in \Theta, H_t \in \mathcal{H}_t}\]
as follows. Define a new static information structure at history $(m_0)^t$ (the history comprising the message $m_0$ for $t$ consecutive periods) such that for every pair of states $\theta, \theta' \in \Theta$, $\theta \neq \theta'$,

(i) $p_t^{I''}(m_0 | \theta, (m_0)^t) = \frac{\sum_{H_t \in \mathcal{H}_t^0(I)} \mathbb{P}^{I'}(H_t | \theta) p_t^{I'}(m_0 | \theta, H_t)}{\sum_{H_t \in \mathcal{H}_t^0(I)} \mathbb{P}^{I'}(H_t | \theta)}$

(ii) $p_t^{I''}(m_0 | \theta, (m_0)^t) = 0$

(iii) $p_t^{I''}(m_0 | \theta, (m_0)^t) = \frac{\sum_{H_{t+1} \in \mathcal{H}_t^0(I)} \mathbb{P}^{I'}(H_{t+1} | \theta) p_t^{I'}(m_0 | \theta, H_t)}{\sum_{H_t \in \mathcal{H}_t^0(I)} \mathbb{P}^{I'}(H_t | \theta)}$

and one can check that this dynamic information structure is well-defined i.e., $I'' \in I^*$. Part (i) takes the probability under $I'$ that the DM receives message $m_0$ at time $t+1$, conditional on not stopping before $t$, and equates that to the probability under $I''$ that she receives the message $m_0$ at time $t + 1$, conditional on receiving only the message $m_0$ before $t$. Part (ii) ensures that upon receipt of message $m_0$, the DM learns the state immediately i.e., places full probability on $\theta$. Part (iii) is the complement of part (i) and takes the probability under $I'$ that the DM does not stop at time $t + 1$ conditional on not stopping before $t$ and equates it to the probability under $I''$ that the DM receives message $m_0$ at time $t + 1$ conditional on receiving messages $m_0$ up to time $t$.

The next lemma establishes that if, under $I''$, the DM has not yet learnt the state at time $t$ her beliefs are the same as if she DM has not learnt the state at time $t$ under $I'$. We emphasise that
the latter event is over (possibly many) histories rather than any particular history.

**Lemma 4.** \( P''(\theta \mid (m_\theta)^t) = P'(\theta \mid \tau(I') > t) \) for all \( \theta \in \Theta \).

**Proof of Lemma 4.** Observe

\[
\begin{align*}
P''((m_\theta)^t \mid \theta) &= \prod_{s=0}^{t-1} P''(m_\theta \mid \theta, (m_\theta)^s) \\
&= \sum_{H_t \in H^t(I') \setminus H^t(I)} P'(H_t \mid \theta) \\
&= P'(\tau(I') > t \mid \theta)
\end{align*}
\]

so by Bayes’ rule,

\[
\begin{align*}
P''(\theta \mid (m_\theta)^t) &= \frac{\mu_0(\theta) P''((m_\theta)^t \mid \theta)}{\sum_{\theta' \in \Theta} \mu_0(\theta') P''((m_\theta)^t \mid \theta)} \\
&= \frac{\mu_0(\theta) P'(\tau(I') > t \mid \theta)}{\sum_{\theta' \in \Theta} \mu_0(\theta') P'(\tau(I') > t \mid \theta')} \\
&= P'(\theta \mid \tau(I') > t),
\end{align*}
\]

for each \( \theta \in \Theta \) as desired. \( \square \)

The next lemma establishes that if, under \( I'' \), the DM is not yet certain of the state at time \( t \), her expected utility from not stopping until seeing message \( m_\theta \) for some \( \theta \in \Theta \) is the same if she followed the same strategy under \( I' \). This follows because (i) by Lemma 4 the DM’s belief under \( I' \) and \( I'' \) at time \( t \) conditional on not yet learning the state is identical; and (ii) the probabilities of receiving message \( m_\theta \) in future periods conditional on not yet learning the state at time \( t \) is identical by the construction of \( I'' \).

**Lemma 5.** If the DM never stops until seeing the full information under \( I'' \), then his expected utility given history \( (m_\theta)^t \) is \( U''((m_\theta)^t) = U'[\phi(\theta, \tau) \mid \tau(I') > t] \).

**Proof.** For any \( \theta \in \Theta \), \( t \in T \), and \( s > t \), the probability of history \( ((m_\theta)^{s-1}, m_\theta) \) given state \( \theta \) and history \( (m_\theta)^t \) is

\[
\begin{align*}
P''((m_\theta)^{s-1}, m_\theta) \mid \theta, (m_\theta)^t) &= P''(m_\theta \mid \theta, (m_\theta)^{s-1}) \times \prod_{s'=t}^{s-2} P''(m_\theta \mid \theta, (m_\theta)^{s'}) \\
&= \frac{\sum_{H_t \in H^t(I')} P'(\theta, H_t)}{\sum_{H_t \in H^t(I') \setminus H^t(I)} P'(H_t \mid \theta)} \\
&= P'(\tau(I') = s \mid \theta, \tau(I') > t).
\end{align*}
\]

The DM’s expected utility given history \( (m_\theta)^t \) is conditioned on both when she receives mes-
sage \( m_\theta \) for some \( \theta \in \Theta \), as well as her belief \( \mu|(m_\theta)^t \in \Delta(\Theta) \):

\[
\mathbb{U}''(\mu^t) = \sum_{\theta \in \Theta} \left[ \sum_{s=t+1}^{\infty} \mathbb{P}( ((m_\theta)^{s-1}, m_\theta) \mid \theta, (m_\theta)^t ) \mathbb{E}(\mu^t(\theta^t) \mid \theta, s) \right]
\]

\[
= \sum_{\theta \in \Theta} \left[ \sum_{s=t+1}^{\infty} \mathbb{P}( \tau(I') = s \mid \theta, \tau(I') > t ) \mathbb{E}(\mu^t(\theta^t) \mid \theta, s) \right]
\]

\[
= \mathbb{E}[\mu^t(\theta, \tau(I')) \mid \tau(I') > t],
\]

as desired. The second equality used (i) the result developed above that the conditional distribution of the stopping time \( \tau(I') \) under \( I' \) is identical to that of the time at which she receives message \( m_\theta \) under \( I'' \); and (ii) Lemma 4.

The next Lemma states that this modification from \( I' \) to \( I'' \) remains obedient.

**Lemma 6.** Under \( I'' \), for any \( t \in \mathcal{T} \), the DM continues paying attention at history \( (m_\theta)^t \).

**Proof.** It will suffice to show that the DM does weakly better by continuing until seeing message \( m_\theta \) for some \( \theta \in \Theta \) since that is a stopping time the DM can choose. To this end, note that the DM’s expected utility if she stops at history \( (m_\theta)^t \) under \( I'' \) is

\[
\max_{a \in A} \mathbb{E}[\mu^t(\theta, a, t) \mid (m_\theta)^t] = \max_{a \in A} \mathbb{E}[\mu^t(\theta, a, t) \mid \tau(I') > t] \quad \text{(Lemma 4)}
\]

\[
= \max_{a \in A} \mathbb{E}[\mu^t(\theta, a, t) \mid H_i] \mid \tau(I') > t \quad \text{(Lemma 2)}
\]

\[
\leq \mathbb{E}[\tau(I') \mid H_i] \quad \text{(Jensen)}
\]

\[
\leq \mathbb{E}[\mathbb{U}(H_i) \mid \tau(I') > t] \quad \text{(Def. of } \mathbb{U}(\cdot)\text{)}
\]

\[
= \mathbb{E}[\mathbb{U}(\theta, H_i) \mid \tau(I') > t] \quad \text{(Iterated expectation)}
\]

\[
= \mathbb{U}(\theta, \tau(I')) \mid \tau(I') > t \quad \text{(Lemma 5)}
\]

which implies that the DM does weakly better by continuing to pay attention.

We are almost done. From Lemma 6, under \( I'' \) the DM does not stop unless she sees message \( m_\theta \) for some \( \theta \in \Theta \) i.e., \( \mathbb{P}(\tau(I'') = s \mid \theta) = \mathbb{P}((m_\theta)^{s-1}, m_\theta) \mid \theta \). From the first part of the proof of Lemma 3, the probability that the DM stops paying attention at time \( s > t \) given that she has not stopped up to time \( t \) is

\[
\mathbb{P}((m_\theta)^{s-1}, m_\theta) \mid \theta, (m_\theta)^t = \mathbb{P}(\tau(I') = s \mid \theta, \tau(I') > t)
\]

and pick \( t = 0 \) to yield \( d(I') = d(I'') \). But from Lemma 3, \( d(I') = d(I) \) which concludes the proof.

39
Appendix B: Remaining Omitted Proofs in Section 3

In this Appendix we collect proofs which were omitted in Section 3.

B.1 Proof that $C - FOSD$ is equivalent to decreasing ratio of CDFs.

Proof. ($\Rightarrow$) If $\tau \geq \tau'$, then, for every $t \in \mathcal{T}$, $\tau > t \geq FOSD \tau' > t$, which implies $\mathbb{P}(\tau > t + 1 | d = \tau > t) \geq \mathbb{P}(\tau' > t + 1 | \tau' > t)$.

($\Leftarrow$) If $\frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)} \geq \frac{\mathbb{P}(\tau' > t + 1)}{\mathbb{P}(\tau' > t)}$ for every $t \in \mathcal{T}$, then for every $s > t$,

$$\mathbb{P}(\tau > s | \tau > t) = \prod_{i=t}^{s-1} \frac{\mathbb{P}(\tau > i + 1)}{\mathbb{P}(\tau > i)} \geq \prod_{i=t}^{s-1} \frac{\mathbb{P}(\tau' > i + 1)}{\mathbb{P}(\tau' > i)} = \mathbb{P}(\tau' > s | d' > t),$$

which implies $\tau \geq C - FOSD \tau'$.

B.2 Proof of Proposition 2

Proof of Proposition 2. We begin by stating a helpful lemma which we prove below.

Lemma 7. For every sequence $\{\epsilon_t\}_{t=0}^{+\infty} \subset [0, 1]^{\mathcal{T} \cup \{+\infty\}}$, if $\tau$ is feasible, then so is $\tau'$, where

$$\frac{\mathbb{P}(\tau' > t + 1)}{\mathbb{P}(\tau' > t)} = (1 - \epsilon_t) \frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)},$$

for every $t \in \mathcal{T}$.

We use this to complete the proof of Proposition 2. Take any dynamic information structure $I$ with induced stopping time $\tau(I)$. Now let $\tau'$ be any stopping time such that $\tau(I) \geq C - FOSD \tau'$. Using the fact that C-FOSD is equivalent to decreasing ratio of CDFs, we must have that

$$\frac{\mathbb{P}(\tau(I) > t + 1)}{\mathbb{P}(\tau' > t + 1)} \geq \frac{\mathbb{P}(\tau(I) > t)}{\mathbb{P}(\tau' > t)}.$$

There then exists a sequence $\{\epsilon_t\}_{t=0}^{+\infty} \subset [0, 1]^{\mathcal{T} \cup \{+\infty\}}$ which makes the inequality above tight anywhere. But by Lemma 7, this implies $\tau'$ is also feasible.

Proof of Lemma 7. Suppose $\tau$ can be obtained by a simple and obedient information structure $I = \{p_i^I(\cdot | \theta, H_i)\}_{t,0,H_i}$. We define a new simple information structure $I' = \{p_i^{I'}(\cdot | \theta, H_i)\}_{t,0,H_i}$ as follows:

$$p_i^{I'}(m_\theta | \theta, (m_\theta)^I) = (1 - \epsilon_t) p_i^I(m_\theta | \theta, (m_\theta)^I),$$

for every $t \in \mathcal{T}$. Then $I'$ is also a feasible information structure.
for every \( t \in \mathcal{T} \). We will show that \( I' \) is also obedient. For any \( t \in \mathcal{T} \) and state \( \theta \in \Theta \),
\[
\mathbb{P}^{I'}((m_{t})^{t} \mid \theta) = \left( \prod_{s=1}^{t-1}(1 - e_{s}) \right) \mathbb{P}^{I}((m_{0})^{t} \mid \theta),
\]
so for every \( \theta \in \Theta \),
\[
\mathbb{P}^{I'}(\theta \mid (m_{0})^{t}) = \frac{\mathbb{P}^{I'}((m_{0})^{t} \mid \theta, I')\mu(\theta)}{\sum_{\theta' \in \Theta} \mathbb{P}^{I'}((m_{0})^{t} \mid \theta', I')\mu(\theta')} = \frac{\mathbb{P}^{I}((m_{0})^{t} \mid \theta, I)\mu(\theta)}{\sum_{\theta' \in \Theta} \mathbb{P}^{I}((m_{0})^{t} \mid \theta', I)\mu(\theta')} = \mathbb{P}^{I}(\theta \mid (m_{0})^{t})
\]
i.e., the beliefs at history \((m_{0})^{t}\) under the new and old dynamic information structures are identical.

Define \( \underline{U}^{I}((m_{0})^{t} \mid \theta) \) as the expected payoff from continuing to pay attention until receiving message \( m_{0} \) for \( \theta \in \Theta \) at which point she learns the state perfectly. We can derive dynamic recursions of both \( \underline{U}^{I}((m_{0})^{t} \mid \theta) \) and \( \overline{U}^{I}((m_{0})^{t} \mid \theta) \) as follows:
\[
\underline{U}^{I}((m_{0})^{t} \mid \theta) = (1 - p_{1}^{I}(m_{0} \mid \theta, (m_{0})^{t})v^{*}(\theta, t + 1) + p_{1}^{I}(m_{0} \mid \theta, (m_{0})^{t})\underline{U}^{I}((m_{0})^{t+1} \mid \theta),
\]
\[
\overline{U}^{I}((m_{0})^{t} \mid \theta) = (1 - p_{1}^{I'}(m_{0} \mid \theta, (m_{0})^{t})v^{*}(\theta, t + 1) + p_{1}^{I'}(m_{0} \mid \theta, (m_{0})^{t})\overline{U}^{I'}((m_{0})^{t+1} \mid \theta).
\]

Therefore taking the difference and grouping terms, we have
\[
\underline{U}^{I'}((m_{0})^{t} \mid \theta) - \underline{U}^{I}((m_{0})^{t} \mid \theta)
= \left( p_{1}^{I}(m_{0} \mid \theta, (m_{0})^{t}) - p_{1}^{I'}(m_{0} \mid \theta, (m_{0})^{t}) \right) \left( v^{*}(\theta, t + 1) - \underline{U}^{I}((m_{0})^{t+1} \mid \theta) \right) + p_{1}^{I'}(m_{0} \mid \theta, (m_{0})^{t}) \left( \overline{U}^{I'}((m_{0})^{t+1} \mid \theta) - \underline{U}^{I}((m_{0})^{t+1} \mid \theta) \right).
\]

Our goal is to show that the difference above is positive. To this end, we will recursively expand it: for every \( t' > t \), write
\[
\underline{U}^{I'}((m_{0})^{t} \mid \theta) - \underline{U}^{I}((m_{0})^{t} \mid \theta)
= \sum_{s=t}^{t'} \left( p_{1}^{I}(m_{0} \mid \theta, (m_{0})^{s}) - p_{1}^{I'}(m_{0} \mid \theta, (m_{0})^{s}) \right) \left( v^{*}(\theta, s + 1) - \underline{U}^{I}((m_{0})^{s+1} \mid \theta) \right) \prod_{s'=t}^{s-1} p_{s'}^{I'}(m_{0} \mid \theta, (m_{0})^{s'})
+ \sum_{s=t}^{t'} \left( p_{s}^{I'}(m_{0} \mid \theta, (m_{0})^{s'}) \right) \left( \overline{U}^{I'}((m_{0})^{s+1} \mid \theta) - \underline{U}^{I}((m_{0})^{s+1} \mid \theta) \right).
\]

If \( \prod_{s'=t}^{\infty} p_{s'}^{I'}(m_{0} \mid \theta, (m_{0})^{s'}) = 0 \), the second term vanishes. If \( \prod_{s'=t}^{\infty} p_{s'}^{I'}(m_{0} \mid \theta, (m_{0})^{s'}) \neq 0 \), this implies \( \mathbb{P}(\tau = +\infty) > 0 \), which in turn implies that \( \lim_{t' \to +\infty} v^{*}(\theta, t') = -\infty \). Therefore,
\[
\lim_{t' \to +\infty} \left( \underline{U}^{I'}((m_{0})^{t'+1} \mid \theta) - \underline{U}^{I}((m_{0})^{t'+1} \mid \theta) \right) = \lim_{t' \to +\infty} v^{*}(\theta, t' + 1) - \lim_{t' \to +\infty} v^{*}(\theta, t' + 1) = 0
\]
which then implies the second term also vanishes:

$$\lim_{t' \to +\infty} \left( \prod_{s'=t}^{t'} p'_{s'}(m_\theta \mid \theta, (m_\theta)^{s'}) \right) \left( U^{I'}((m_\theta)^{t'+1} \mid \theta) - U^{I'}((m_\theta)^{t' \mid \theta}) \right) = 0.$$ 

Thus,

$$U^{I'}((m_\theta)^t \mid \theta) - U^{I'}((m_\theta)^t \mid \theta)$$

$$= \sum_{s=t}^{+\infty} \left[ p^I_t(m_\theta \mid \theta, (m_\theta)^s) - p^I_t(m_\theta \mid \theta, (m_\theta)^s) \right] \left( v^*(\theta, s+1) - U^{I'}((m_\theta)^{s+1} \mid \theta) \right) \prod_{s'=t}^{s-1} p'_{s'}(m_\theta \mid \theta, (m_\theta)^{s'})$$

$$\geq 0,$$

where the last line follows from the facts that (i) by construction, for every $s \geq t$, $p^I_t(m_\theta \mid \theta, (m_\theta)^s) < p^I_t(m_\theta \mid \theta, (m_\theta)^s)$; and (ii)

$$U^{I'}((m_\theta)^{s+1} \mid \theta) \leq \mathbb{E}^{I'}[v^*(\theta, \tau) \mid \tau > s + 1] \leq v^*(\theta, s + 1).$$

This implies that, under $I'$, the DM’s utility if he continues to pay attention until seeing $m_\theta$ given the history $(m_\theta)^t$ is at least

$$\mathbb{E}^{I'}[U^{I'}((m_\theta)^t \mid \theta)] \geq \mathbb{E}^{I'}[U^{I'}((m_\theta)^t \mid \theta)]$$

$$= \mathbb{E}^{I'}((m_\theta)^t)$$

$$\geq \max_{a \in A} \mathbb{E}^{I'}[v(a, \theta, t) \mid (m_\theta)^t]$$

$$(I \text{ is obedient})$$

$$= \max_{a \in A} \mathbb{E}^{I'}[v(a, \theta, t) \mid (m_\theta)^t],$$

$$(\text{Equation 1})$$

where the last equality follows from the fact that the DM’s beliefs under the $I$ and $I'$ after each history $(m_\theta)^t$ are identical. As such, the DM does not stop paying attention given the history $(m_\theta)^t$. This implies the information structure $I'$ is obedient. Therefore,

$$\frac{P(\tau(I') > t + 1)}{P(\tau(I') > t)} = p^I_t(m_\theta \mid (m_\theta)^t)$$

$$= (1 - \epsilon_t)p^I_t(m_\theta \mid (m_\theta)^t)$$

$$= (1 - \epsilon_t)\frac{P(\tau > t + 1)}{P(\tau > t)}.$$

which implies $\tau^* \overset{d}{=} \tau(I')$, as desired.

\[\square\]

B.3 Proof of Proposition 3

Proof of Proposition 3 (1 $\Rightarrow$ 2) The obedience constraint (i) is implied by the assumption that the information structure we start with is obedient. For the second condition, the martingale
property of beliefs implies

\[ \mu_t(\theta) := \mathbb{P}^t(\theta \mid (m_0)^t) \]

\[ = p_t^f(m_0 \mid (m_0)^t) \cdot 1 + p_t^f(m_0 \mid (m_0)^t) \cdot \mathbb{P}^f(\theta \mid (m_0)^{t+1}) \]

\[ \geq p_t^f(m_0 \mid (m_0)^t) \cdot \mu_{t+1}(\theta) \]

\[ = \mathbb{P}(\tau > t + 1 \mid \tau > t) \cdot \mu_{t+1}(\theta), \]

which yields the second condition.

(2 \Rightarrow 1) Define a simple information structure \( I \) as follows:

\[ p_t^f(m_0 \mid (m_0)^t) = \mu_t(\theta) - \mathbb{P}(\tau > t + 1 \mid \tau > t) \cdot \mu_{t+1}(\theta), \]

which, from the boundary condition (ii), is weakly greater than 0. This means \( p_t^f(m_0 \mid (m_0)^t) = \mathbb{P}(\tau > t + 1 \mid \tau > t) \in [0, 1] \), so the information structure \( I \) is well-defined. Clearly, if the DM always continues paying attention until seeing \((m_0)^t\), the (random) stopping time has the same distribution as that of \( \tau \). Therefore, from the first condition, the IC constraint holds in such a way that the DM always continue paying attention after seeing \((m_0)^t\), which implies \( I \) is obedient. Hence, \( \tau \overset{d}{=} \tau(I) \).

B.4 Proof of Proposition 4 (Increasing and extremal belief paths). Notation. We first introduce some useful notation. Define the set of belief paths with \( W = (\Delta(\Theta))^T \). Fixing a stopping time \( \tau \), define \( W(\tau) \subset W \) as the set of belief paths corresponding to stopping time \( \tau \).

Definition 10 (Maximal belief path). Fix stopping time \( \tau \). A belief path \((\mu_t)_{t \in T} \in W(\tau)\) is a maximal belief path under \( \tau \) if there is no \((\mu'_t)_{t \in T} \in W(\tau)\) such that \( \phi(\mu'_t) \geq \phi(\mu_t) \) for every \( t \in T \) and the inequality is strict for some \( t \in T \).

The following proposition guarantees the existence of a maximum belief path for any feasible stopping time \( \tau \). The proof is deferred to Online Appendix[1]. This will be a useful object to prove Propositions[4] and[5].

Proposition 11. For every feasible stopping time \( \tau \), there exists a maximal belief path corresponding to \( \tau \).

Proof idea. We now outline the ideas behind the proof before presenting it formally. From Proposition[3] we know that simple information structures are characterized by two functions over \( T \), and relate to each other via the IC and Boundary constraints. Notice, however, that fixing the distribution of \( \tau \), the curve \( \mathbb{E}[\tau \mid \tau > t] \) as a function of \( t \) is completely pinned down. Our goal is to show that any maximal belief path has the property of increasing and extremal belief paths. We prove by a contradiction that if a maximal belief path \((\mu_t)_{t \in T} \) is not increasing on \( \phi \) over \( T \) (or not extremal), we can find another belief path \((\mu'_t)_{t \in T} \) fulfilling the boundary and IC constraints which is increasing in \( \phi \) over \( T \) (or extremal), and \( \phi(\mu'_t) \geq \phi(\mu_t) \) for every \( t \in T \) with at least one strict inequality. This contradicts the property of a maximal belief path, which establishes the proposition.
To establish the existence of this maximal belief path, we first suppose that a maximal belief path \((\mu_t)_{t \in \mathcal{T}}\) is not increasing on \(\phi\) over \(\mathcal{T}\). Our proof will proceed inductively. We index the sequence of beliefs created at each inductive step with superscripts (e.g., \((\mu_i^t)\) is the sequence of beliefs created at inductive step \(i\)) and as usual, use subscripts to track time. At each step \(i \in \mathcal{T} \cup \{-1\}\), assuming that the sequence of beliefs from the previous step is \((\mu_i^t)\) and that up to period \(i\), the beliefs are an increasing function of \(\phi\), i.e., \(\phi(\mu_{i-1}^t)\) is increasing over the interval \(t = i - 1\), we construct a new sequence of beliefs, \((\mu_i^t)\), by retaining the first \(i - 1\) beliefs in the previous sequence (i.e., \((\mu_i^t)_{i-1}^t = (\mu_{i-1}^t)_{i-1}^t\) and “straightening out” the decreasing portions of the sequence \((\mu_{i-1}^t)_{i-1}^t\).

More explicitly, let \(t_{i-1}\) be the first time \(\phi(\mu_{i-1}^t)\) exceeds \(\phi(\mu_{i-1}^s)\) i.e., \(\phi(\mu_{i-1}^s) \leq \phi(\mu_{i-1}^t)\) but \(\phi(\mu_{i-1}^s) > \phi(\mu_{i-1}^t)\) for all \(i-1 < s < t_{i-1}\). By choosing weights carefully, we can find an alternate sequence of beliefs \((\mu_i^t)_{i-1}^t\) which is arc convex combinations of the beliefs \(\mu_i^t\) and \(\mu_{i-1}^t\) while, at the same time, respecting both boundary and obedience constraints. Call this new sequence of beliefs \((\mu_i^t)\), where for \(t = i - 1\), the beliefs are the same as \((\mu_i^t)\), and for \(t > i - 1\), the beliefs follow the modification we performed. However, this new sequence of beliefs is not necessarily increasing in \(\phi\) over \(t = i - 1, \ldots, t_{i-1}\) so we might have, through this process, created a new decreasing region. But the sequence is (i) concave; and (ii) locally strictly increasing i.e., \(\phi(\mu_{i-1}^t) < \phi(\mu_i^t)\). The outcome of this procedure is illustrated in Figure 13 (a). The solid red line plots \(\phi(\mu_{i-1}^t)\), the value of full information under the original path of beliefs; the solid blue line plots the additional expected costs of waiting until the DM receives full information. Obedience requires that the red line is, at all times, above the blue line. Following the modification above, the dotted red line depicts the value of information under the modified path of beliefs \(\phi(\mu_i^t)\). We then iteratively perform this procedure; the outcome is depicted by the dotted red line in Figure 13 (b).

We now prove Proposition 4 formally.
Proof of Proposition (i) (increasing paths). Start from an arbitrary simple and obedient dynamic information structure \( I \in I^+ \). Denote the DM’s beliefs upon receipt of the null message up to time \( t \) under this information structure with \( \mu_t^I := \mu\left| (m_0)^I \right. \).

Consider any sequence of \( (\mu_t^I)_{t \in T} \in (\Delta(\Theta))^T \). We recursively construct another sequence of beliefs \( (\mu_t^{i+1})_{t \in T} \in (\Delta(\Theta))^T \) which satisfies the conditions in Lemma 3 for each \( i \in T \cup \{ -1 \} \) as follows:

(i) If \( i = -1 \), set \( \mu_t^I = \mu_t^I \) for every \( t \in T \).

(ii) If \( i \geq 0 \), given the sequence of \( (\mu_t^I)_{t \in T} \in (\Delta(\Theta))^T \), define

\[
t_i = \begin{cases} 
\min \{ t \in T : t > i, \phi(\mu_t^I) \leq \phi(\mu_i^I) \} & \text{if the minimum exists,} \\
+\infty & \text{otherwise.}
\end{cases}
\]

- If \( t_i = +\infty \), set \( \mu_t^{i+1} = \begin{cases} 
\mu_t^I, & \text{if } t < i \\
\mu_i^I, & \text{if } t \geq i
\end{cases} \)
i.e., the sequence \( (\mu_t^{i+1}) \) follows the sequence \( (\mu_t^I) \) up to period \( i \), and remains constant at belief \( \mu_i^I \) thereafter.

- If \( t_i < +\infty \), define \( \theta_i \in \arg\max_{\theta \in \Theta} \frac{\mu_i^I(\theta)}{\mu_{t_i}^I(\theta)} \) i.e., the state at which the ratio of the beliefs at time \( i \) and time \( t_i \) — the time at which \( \phi \) increases relative to that at time \( i \).

It is clear that \( \frac{\mu_i^I(\theta)}{\mu_{t_i}^I(\theta)} \geq 1 \). Thus, there exists a sequence \( \pi_i^1, \ldots, \pi_i^{t_i-1} \in \mathbb{R} \) such that \( \pi_i^t \in [1, \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t_i)}] \) for every \( t \in \{i, \ldots, t_i-1\} \) and \( \prod_{t=i}^{t_i-1} \pi_i^t = \frac{\mu_i^I(\theta_i)}{\mu_{t_i}^I(\theta_i)} \) because the second condition of Lemma 3 for \( \mu_I \) implies

\[
\frac{\mu_i^I(\theta_i)}{\mu_{t_i}^I(\theta_i)} \leq \frac{\mathbb{P}(\tau > i)}{\mathbb{P}(\tau > t_i)} = \prod_{t=i}^{t_i-1} \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t + 1)}.
\]

This sequence \( (\pi_i^t)_{t=i}^{t_i-1} \) simply splits up the ratio of \( \mu_i^I(\theta_i)/\mu_{t_i}^I(\theta_i) \) into \( t_i - 1 - i \) sub-ratios. We will now construct a new sequence of beliefs to bridge \( \mu_t^I \) and \( \mu_{t_i}^I \) while ensuring that \( \phi \) increases over the interval. For any \( t \in \{i, \ldots, t_i\} \), set

\[
\mu_t^{i+1} = \lambda_{i,t} \mu_t^I + (1 - \lambda_{i,t}) \mu_{t_i}^I,
\]
i.e., a linear combination of the belief at \( t_i \) and that at \( i \) where the weights are given by

\[
\lambda_{i,t} = \frac{1}{\mu_{t_i}^I(\theta_i) - \mu_i^I(\theta_i)} \left( \mu_i^I(\theta_i) - \prod_{s=i}^{t_i-1} \pi_s^t \cdot \mu_i^I(\theta_i) \right).
\]

Moreover, if \( t \notin \{i, \ldots, t_i\} \), we set \( \mu_t^{i+1} = \mu_t^I \). From the construction of \( \mu^{i+1} \), note that \( \mu_t^{i+1} = \mu_t^I \) and \( \mu_{t_i}^{i+1} = \mu_{t_i}^I \).

We will now inductively show that a sequence \( (\mu_t^I)_{t \in T} \) satisfies the conditions in Proposition 45.
for each \( i \in \mathcal{T} \cup \{-1\}:

**Basic step.** \( (\mu_i^j)_{i \in \mathcal{T}} = (\mu^j | (m_0)^i)_{i \in \mathcal{T}} \) satisfies the conditions in Proposition 3 because by \( I \) was assumed to be simple and obedient.

**Inductive step.** Suppose \( t \in \mathcal{T} \) such that \( (\mu_i^j)_{i \in \mathcal{T}} \) satisfies the conditions in Proposition 3. We consider two cases.

Case 1: \( t_i = +\infty \). We have that \( \phi(\mu_i^j) > \phi(\mu_i^j) \) for every \( t_i > j \). This implies \( \phi(\mu_i^{j+1}) = \phi(\mu_i^j) > \phi(\mu_i^j) \geq \mathbb{E}[c(\tau)|\tau > t] - c(t) \) for every \( t_i > j \). For every \( t_i \leq j, \phi(\mu_i^{j+1}) = \phi(\mu_i^j) \geq \mathbb{E}[c(\tau)|\tau > t] - c(t) \). Therefore, the boundary constraint holds for every time \( t \in \mathcal{T} \) for the sequence of beliefs \( (\mu_i^{j+1})_{i \in \mathcal{T}} \). Moreover, for every \( t_i < j, \) we know \( \frac{\mu_i^{j+1}(\theta)}{\mu_i^j(\theta)} = \frac{\mu_i^j(\theta)}{\mu_i^j(\theta)} = 1 \leq \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t + 1)} \). For every \( t_i \geq j, \) we have \( \frac{\mu_i^{j+1}(\theta)}{\mu_i^j(\theta)} = \frac{\mu_i^j(\theta)}{\mu_i^j(\theta)} = 1 \leq \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t + 1)} \). Thus, the boundary constraint holds for every time \( t \in \mathcal{T} \) for the sequence of beliefs \( (\mu_i^{j+1})_{i \in \mathcal{T}} \), as required.

Case 2: \( t_i < +\infty \). From the construction of \( (\pi_i^l)_{i \in \mathcal{T}} \), we know

\[
1 \leq \prod_{s=1}^t \pi_s \leq \pi_{t-1} \prod_{s=1}^t \pi_s = \frac{\mu_i^j(\theta_t)}{\mu_i^j(\theta_t)} \implies 0 \leq \lambda_i \leq 1,
\]

for every \( t \in \{i, \ldots, t_i - 1\} \). We now verify each of the constraints holds.

(i) (Obedience constraint) The definition of \( t_i \) implies \( \phi(\mu_i) > \phi(\mu_i) \) for every \( t_i \in \{i + 1, \ldots, t_i - 1\} \) and \( \phi(\mu_i) \leq \phi(\mu_i) \). Because \( \phi \) is concave and \( \lambda_i \leq 1 \), Jensen’s inequality implies that for every \( t_i \in \{i, \ldots, t_i - 1\} \),

\[
\phi(\mu_i^{j+1}) = \phi(\lambda_i \mu_i^j + (1 - \lambda_i) \mu_i^j) \geq \lambda_i \phi(\mu_i^j) + (1 - \lambda_i) \phi(\mu_i^j) \geq \phi(\mu_i^j) \geq \mathbb{E}[c(\tau)|\tau > t] - c(t),
\]

where the last inequality follows from the first condition in Proposition 3 of \( \mu^j \). As such, for every \( t_i \in \{i, \ldots, t_i - 1\} \), we have \( \phi(\mu_i^{j+1}) = \phi(\mu_i^j) \geq \phi(\mu_i^j) \geq \mathbb{E}[c(\tau)|\tau > t] - c(t) \). Therefore, the obedience constraint holds for every time \( t \in \mathcal{T} \) for the sequence of beliefs \( (\mu_i^{j+1})_{i \in \mathcal{T}} \).

(ii) (Boundary constraint) For every \( t_i \in \{i, \ldots, t_i - 1\} \) and \( \theta \in \Theta \), we have

\[
\frac{\mu_i^{j+1}(\theta)}{\mu_i^{j+1}(\theta)} = \frac{\lambda_i \mu_i^j(\theta) + (1 - \lambda_i) \mu_i^j(\theta)}{\lambda_i \mu_i^j(\theta) + (1 - \lambda_i) \mu_i^j(\theta)} = \frac{\lambda_i \mu_i^j(\theta) + (1 - \lambda_i) \mu_i^j(\theta)}{\lambda_i \mu_i^j(\theta) + (1 - \lambda_i) \mu_i^j(\theta)} \leq \frac{\lambda_i \mu_i^j(\theta) + (1 - \lambda_i) \mu_i^j(\theta)}{\lambda_i \mu_i^j(\theta) + (1 - \lambda_i) \mu_i^j(\theta)} \leq \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t + 1)}.
\]
Moreover, for every \( t \notin \{i, \ldots, t_i - 1\} \) and \( \theta \in \Theta \), we have \( \frac{\mu_{i+1}^t(\theta)}{\mu_i^t(\theta)} = \frac{\mu_{j+1}^t(\theta)}{\mu_j^t(\theta)} \leq \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t+1)} \) from the induction hypothesis. Therefore, the boundary constraint holds for every time \( t \in T \) and state \( \theta \in \Theta \) for the sequence of beliefs \( (\mu_i^t)_{t \in T} \).

We have shown that the sequence \( (\mu_i^t)_{t \in T} \) satisfies the conditions in Proposition 3 for each \( i \in T \cup \{-1\} \). We now complete the proof of Proposition 4. Define a sequence \( (\mu_i^t)_{t \in T} \in (\Delta(\Theta))^T \) such that \( \mu_i^t = \mu_i^t \) for every \( t \in T \) i.e., taking the 'diagonal' by choosing time \( t \)'s beliefs to be sequence \( t \)'s beliefs at time \( t \).

- For the obedience constraint, we see that \( \phi(\mu_i^*) = \phi(\mu_i^t) \geq \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t+1)} \).
- For the boundary constraint, we have \( \frac{\mu_{i+1}^t(\theta)}{\mu_i^t(\theta)} = \frac{\mu_{j+1}^t(\theta)}{\mu_j^t(\theta)} \leq \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t+1)} \), as desired.
- For the increasing property, we see that from Equation 3 that \( \phi(\mu_i^*) = \phi(\mu_i^t) \leq \phi(\mu_i^t) = \phi(\mu_i^t+1) \).

By Lemma 3 we can therefore find a simple and obedient information structure \( I^* \in I \) such that \( \mu^t | (m_0)^t = \mu_i^t \) for which (i) \( \phi(\mu^t | (m_0)^t) = \phi(\mu_i^t) \) is weakly increasing in \( t \); and (ii) preserves the same distribution of stopping times i.e., \( \tau(I) \sim \tau(I^*) \) (which was held fixed throughout the proof). From the proof of the obedience constraint, we have \( \phi(\mu_i^*) \geq \phi(\mu_i^t) \) for every \( t \in T \), and the inequality is strict for some \( t \in T \) if \( \phi(\mu_i^t) \) is not weakly increasing. This contradicts the fact that \( (\mu_i^t)_{t \in T} \) is a maximal belief path under \( \tau \).

It remains to show that a maximal belief path \( (\mu_i^t)_{t \in T} \) must have a property of extremal belief paths

**Proof of Proposition 4 (ii) (extremal paths).** Consider feasible stopping time \( \tau \). Pick a maximal belief path \( (\mu_i^t)_{t \in T} \) under \( \tau \). From the proof before, if \( \phi(\mu_i^t) \) is not weakly increasing, we can find another belief path \( (\mu_i^t)_{t \in T} \) such that \( \phi(\mu_i^t) \) is weakly increasing and \( \phi(\mu_i^t) \geq \phi(\mu_i^t) \), which contradicts the fact that \( (\mu_i^t)_{t \in T} \) is a maximal belief path under \( \tau \).

Next, we will show that \( (\mu_i^t)_{t \in T} \) has a property of extremal paths. Suppose a contradiction that there is a minimum \( t_0 \in T \) such that \( \mu_{t_0+1}^{t_0} \notin \Phi^* \) and \( \frac{\mathbb{P}(\tau > t_0 + 1)}{\mathbb{P}(\tau > t_0)} < \min_{\theta \in \Theta} \frac{\mu_{t_0}(\theta)}{\mu_{t_0+1}(\theta)} \). Pick any \( \mu^* \in \Phi^* \). We pick \( \lambda \in [0, 1] \) as following. If \( \min_{\theta \in \Theta} \frac{\mu_{t_0}(\theta)}{\mu_{t_0+1}(\theta)} \geq \frac{\mathbb{P}(\tau > t_0 + 1)}{\mathbb{P}(\tau > t_0)} \), choose \( \lambda = 0 \). Otherwise, choose \( \lambda \in [0, 1] \) such that

\[
\frac{\mathbb{P}(\tau > t_0 + 1)}{\mathbb{P}(\tau > t_0)} = \min_{\theta \in \Theta} \frac{\mu_{t_0}(\theta)}{\mu_{t_0+1}(\theta)} = \frac{\mu_{t_0}(\theta)}{\mu_{t_0+1}(\theta)} \leq \frac{\mathbb{P}(\tau > t_0 + 1)}{\mathbb{P}(\tau > t_0)}.
\]

The existence of \( \lambda \) follows from the intermediate value theorem. We define a new belief path \( (\mu_i^t)_{t \in T} \) as follows:

\[
\mu_i^t = \begin{cases} 
\mu_i^t, & \text{if } t \leq t_0 \\
\lambda \mu_i^t + (1 - \lambda) \mu^*, & \text{otherwise.}
\end{cases}
\]
We will show that a pair of a belief path and a distribution of stopping time \((\mu'_t)_{t \in T}, d(\tau)\) is feasible. The obedience constraint is still the same for \(t \in \{0, \ldots, t_0\}\). For \(t \geq t_0\), we have 
\[
\phi(\mu'_t) = \phi(\lambda \mu_t + (1 - \lambda) \mu^*) \geq \lambda \phi(\mu_t) + (1 - \lambda) \phi^* \geq \phi(\mu_t) \geq \mathbb{E}[c(\tau) | \tau > t] - c(t),
\]
so the obedience constraint for every \(t \in T\). The boundary constraint is still the same for \(t \in \{0, \ldots, t_0 - 1\}\). For \(t = t_0\), the boundary constraint holds because of the construction of \(\lambda\). For \(t > t_0\), we have 
\[
\min_{\theta \in \Theta} \frac{\mu'_t(\theta)}{\mu'_{t+1}(\theta)} = \min_{\theta \in \Theta} \frac{\lambda \mu_t(\theta) + (1 - \lambda) \mu^*(\theta)}{\lambda \mu_{t+1}(\theta) + (1 - \lambda) \mu^*(\theta)} \geq \min_{\theta \in \Theta} \frac{\mu_t(\theta)}{\mu_{t+1}(\theta)} \geq \frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)},
\]
where the first inequality follows from the fact that \(\theta\) that minimizes the LHS must satisfy \(\mu_t(\theta) \leq \mu_{t+1}(\theta)\). This concludes that a pair of a belief path and a distribution of stopping time \(((\mu'_t)_{t \in T}, d(\tau))\) is feasible.

For every \(t \leq t_0\), \(\phi(\mu'_t) = \phi(\mu_t)\). Moreover, for every \(t > t_0\), we have 
\[
\phi(\mu'_t) = \phi(\lambda \mu_t + (1 - \lambda) \mu^*) \geq \lambda \phi(\mu_t) + (1 - \lambda) \phi^* \geq \phi(\mu_t).
\]
Therefore, \(\phi(\mu'_t) \geq \phi(\mu_t)\) for every \(t \in T\). Because \((\mu_t)_{t \in T}\) is a maximal path under \(\tau\), we must have \(\phi(\mu'_t) = \phi(\mu_t)\) for every \(t \in T\), which implies that \(\phi(\mu_t) = \phi^*\) for every \(t > t_0\). This contradicts the fact that \(\mu_{t_0+1} \notin \Phi^*\). Therefore, the belief path \((\mu_t)_{t \in T}\) must have a property of extremal paths, as desired. \(\square\)

With these two parts of the proof, Proposition 4 follows by considering maximal belief paths of any feasible stopping time.

**B.5 Proof of Proposition 5**

**Proof of Proposition 5** Assume that \(\Theta = \{0, 1\}\). We abuse notation of a belief \(\mu \in \Delta(\Theta)\) by \(\mu = \mu(0) \in [0, 1]\). Pick \(\mu^* \in \Phi^*\). By the optimality of \(\phi(\mu^*)\) and the concavity of \(\phi\), we must have \(\phi(\mu)\) is increasing and decreasing if \(\mu < \mu^*\) and \(\mu > \mu^*\), respectively.

Without loss of generality, assume that \(\mu_0 < \mu^*\). Suppose a contradiction that \(\mathbb{P}(\tau > T) > 0\) for every \(T \in T\) and every belief path \(\mu_t\) corresponding to \(\tau\) must satisfy \(\mu_t \notin \Phi^*\). By Proposition 4 there exists an increasing and extremal belief path \((\mu_t)_{t \in T}\) corresponding to \(\tau\). Therefore, \(\phi(\mu_t)\) is increasing in \(t\).

**Case 1:** \(\mu_t < \mu^*\) for every \(t \in T\). Because \(\phi(\mu)\) is increasing if \(\mu < \mu^*\) and \(\phi(\mu_t) < \mu^*\), we have \(\mu_t\) is increasing in \(t\). Therefore, \(\mu : \lim_{t \to \infty} \mu_t\) exists and less than or equal to \(\mu^*\). Because \((\mu_t)_{t \in T}\) is extremal, we have 
\[
\frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)} = \min \left\{ \frac{\mu_t}{\mu_{t+1}}, \frac{1 - \mu_t}{1 - \mu_{t+1}} \right\} = \frac{\mu_t}{\mu_{t+1}}.
\]
for every $t \in \mathcal{T}$, which implies that

$$0 = \lim_{T \to \infty} \mathbb{P}(\tau > T) = \prod_{t=0}^{\infty} \frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)} = \prod_{t=0}^{\infty} \frac{\mu_t}{\bar{\mu}} = \frac{\mu_0}{\bar{\mu}} > 0,$$

which is a contradiction.

**Case 2:** there exists $t \in \mathcal{T}$ such that $\mu_t > \mu^*$. Because $\mu_0 < \mu^*$, we can find $t_0 \in \mathcal{T}$ such that $\mu_{t_0 - 1} < \mu^*$ but $\mu_{t_0} > \mu^*$. Therefore,

$$\frac{\mathbb{P}(\tau > t_0)}{\mathbb{P}(\tau > t_0 - 1)} = \min \left\{ \frac{\mu_{t_0 - 1}}{\mu_{t_0}}, \frac{1 - \mu_{t_0 - 1}}{1 - \mu_{t_0}} \right\} = \frac{\mu_{t_0 - 1}}{\mu_{t_0}} < \frac{\mu_{t_0 - 1}}{\mu^*} = \min \left\{ \frac{\mu_{t_0 - 1}}{\mu^*}, \frac{1 - \mu_{t_0 - 1}}{1 - \mu_t} \right\}.$$

We define a new belief path $(\mu'_t)_{t \in \mathcal{T}}$ such that

$$\mu'_t = \begin{cases} \mu_t, & \text{if } t < t_0 \\ \mu^*, & \text{otherwise.} \end{cases}$$

The inequality above implies that the belief path $(\mu'_t)_{t \in \mathcal{T}}$ is feasible corresponding to $\tau$. However, $\mu'_0 = \mu^* \in \Phi^*$, which is a contradiction.
Appendix C: Omitted Proofs in Section 4 (designer’s problem)

C.1 Proof of Lemma 1

Proof of Lemma 1 (1 ⇔ 2) The forward direction is clear by the obedience condition of Lemma 3. The converse is also clear because we can simply consider a sequence of beliefs \((\mu_t)_{t \in T}\) such that \(\mu_t = \mu_0\) for every \(t \in T\) and apply Lemma 3 again.

\((2 \iff 3)\) For the forward direction, define \(C_t := \mathbb{E}[c(\tau) \mid \tau > t] - c(t) \in [0, \phi^*]\), for every \(t \in T\). Then,

\[
C_t + c(t) = \mathbb{E}[c(\tau) \mid \tau > t] = \mathbb{E}[c(\tau) \mid \tau > t + 1]\mathbb{P}(\tau > t + 1 \mid \tau > t) + c(t + 1)\mathbb{P}(\tau = t + 1 \mid \tau > t)
= \mathbb{E}[c(\tau) \mid \tau > t + 1]\mathbb{P}(\tau > t + 1 \mid \tau > t) + c(t + 1)(1 - \mathbb{P}(\tau > t + 1 \mid \tau > t)),
\]

which implies that

\[
\mathbb{P}(\tau > t + 1 \mid \tau > t) = \frac{C_t + c(t) - c(t + 1)}{C_{t+1}},
\]

as desired. Note here that \(\mathbb{P}\) is with respect to the stopping time we hold fixed. For the converse, we use the same computation to obtain \(\mathbb{E}[c(\tau) \mid \tau > t] - c(t) = C_t \leq \phi^*\) for every \(t \in T\). □

C.2 Proof of Proposition 6 (i)

Proof of Proposition 6 (i). (⇒) First, we show the proposition in the case of simple and obedient information structures. Suppose a contradiction that there is a distribution \(d \in D\) on the FOSD frontier such that its corresponding simple and obedient information structure makes the DM does not stop paying attention at period 0. This implies that

\[
\phi(\mu_0) > \mathbb{E}[c(\tau)].
\]

Pick a corresponding maximal belief path \((\mu_t)_{t \in T}\) and let \(t^* \in T\) be the minimum of \(t\) such that \(\mathbb{P}(\tau > t^*) < 1\). By Proposition 3, we have

\[
\phi(\mu'_t) \geq \mathbb{E}[c(\tau') \mid \tau' > t] - c(t)
\]

\[
\iff \mathbb{P}(\tau' > t)\phi(\mu'_t) \geq \sum_{s=t+1}^{\infty} (c(s) - c(t))(\mathbb{P}(\tau' > s - 1) - \mathbb{P}(\tau' > s))
\]

\[
\iff \mathbb{P}(\tau' > t)(\phi(\mu'_t) + c(t) - c(t + 1)) \geq \sum_{s=t+1}^{\infty} (c(s + 1) - c(s))\mathbb{P}(\tau' > s). \tag{4}
\]
We set
\[
P(\tau' > t) = \begin{cases} 
\mathbb{P}(\tau > t) + \epsilon, & \text{if } t = t^* \\
\mathbb{P}(\tau > t) & \text{otherwise.}
\end{cases}
\]

for arbitrarily small \(\epsilon > 0\). We can see that \(\tau' \leq FOSD\). We construct a belief path \((\mu_t')_{t \in T}\) as follows:

\[
\mu_t' = \begin{cases} 
\delta \mu_0 + (1 - \delta) \mu_t, & \text{if } t = t^* \\
\mu_t & \text{otherwise,}
\end{cases}
\]

where \(\delta > 0\) is a unique solution of
\[
\delta + \frac{1 - \delta}{\mathbb{P}(\tau > t^*)} = \frac{1}{\mathbb{P}(\tau > t^*) + \epsilon} \implies \delta = \frac{\epsilon}{(\mathbb{P}(\tau > t^*) + \epsilon)(1 - \mathbb{P}(\tau > t^*))}
\]

First, we will show that the boundary constraint holds for a pair of a stopping time and a belief path \((\tau', (\mu_t')_{t \in T})\). Consider that

\[
\frac{\mu_{t-1}'(\theta)}{\mu_t'(\theta)} = \frac{\mu_{t-1}(\theta)}{\delta \mu_{t-1}(\theta) + (1 - \delta) \mu_t(\theta)} = \frac{\mathbb{P}(\tau > t)}{\delta \mathbb{P}(\tau > t) + (1 - \delta)} = \mathbb{P}(\tau' > t).
\]

Moreover,
\[
\frac{\mu_t'(\theta)}{\mu_{t+1}'(\theta)} = \frac{\delta \mu_{t-1}(\theta) + (1 - \delta) \mu_t(\theta)}{\mu_{t+1}(\theta)}
\leq \delta \mathbb{P}(\tau' > t^*) + (1 - \delta) \frac{\mathbb{P}(\tau' > t^* + 1)}{\mathbb{P}(\tau > t^*)}
= \mathbb{P}(\tau' > t^* + 1) \left( \delta + \frac{1 - \delta}{\mathbb{P}(\tau' > t^*)} \right) = \frac{\mathbb{P}(\tau' > t^* + 1)}{\mathbb{P}(\tau' > t)},
\]

as desired. We will show that the obedient constraint holds for a pair of a stopping time and a belief path \((\tau', (\mu_t')_{t \in T})\) for small \(\epsilon > 0\). Because \(\phi(\mu_0) > \mathbb{E}[c(\tau)]\) and \(\|\mathbb{E}[c(\tau')] - \mathbb{E}[c(\tau)]\| = \epsilon(c(t^* + 1) - c(t^*))\), we can find sufficiently small \(\epsilon > 0\) so that \(\phi(\mu_0) > \mathbb{E}[c(\tau')]\). For \(t < t^*\), because \(\mathbb{P}(\tau > t) = 1\), we have
\[
\phi(\mu_t') = \phi(\mu_0) > \mathbb{E}[c(\tau')] = \mathbb{E}[c(\tau') \mid \tau' > t] > \mathbb{E}[c(\tau') \mid \tau' > t] - c(t).
\]

For \(t > t^*\), the obedience constraints are the same for \(\tau\) and \(\tau'\). Moreover, the obedience constraint for \(t < t^*\) is slackened if the obedience constraint at time \(t^*\) holds. Thus, it is sufficient to check the obedience constraint at time \(t^*\). Consider that
\[
\mathbb{P}(\tau' > t^*) (\phi(\mu_{t^*}) + c(t^*) - c(t^* + 1))
= (\mathbb{P}(\tau > t^*) + \epsilon) (\phi(\delta \mu_0 + (1 - \delta) \mu_{t^*}) + c(t^*) - c(t^* + 1))
\geq (\mathbb{P}(\tau > t^*) + \epsilon) (\delta \phi(\mu_0) + (1 - \delta) \phi(\mu_{t^*}) + c(t^*) - c(t^* + 1))
= (\mathbb{P}(\tau > t^*) + \epsilon) (\phi(\mu_{t^*}) + c(t^*) - c(t^* + 1)) + \delta (\phi(\mu_0) - \phi(\mu_{t^*})) \mathbb{P}(\tau > t^*) + C \delta \epsilon,
\]

51
where \( C = \phi(\mu_0) - \phi(\mu_{t'}) \). The first equality comes from the construction of \( \tau' \) and \((\mu_{t'})_t\). Therefore,

\[
\begin{align*}
\mathbb{P}(\tau' > t^*)(\phi(\mu_{t'}) + c(t^*) - c(t^* + 1)) &\geq \mathbb{E}[c(\tau) | \tau > t^*] - \mathbb{E}[c(\tau) | \tau > t^*] + \mathbb{P}(\tau > t^*)c(t^*) \\
&= \phi(\mu_{t'}) - \frac{\phi(\mu_{t'}) - \phi(\mu_0)}{1 - \mathbb{P}(\tau > t^*)}c(t^*) - c(t^* + 1) + \mathbb{P}(\tau > t^*)c(t^*) + C\delta\epsilon \\
&= \phi(\mu_{t'}) - \frac{\phi(\mu_{t'}) - \phi(\mu_0)}{1 - \mathbb{P}(\tau > t^*)}c(t^*) - c(t^* + 1) + C\delta\epsilon,
\end{align*}
\]

where the second inequality employs Proposition 4 part (i) which states that it is without loss to take \( \phi(\mu_{t'}) \geq \phi(\mu_0) \). In order to maintain the sender’s optimality of \( \tau \), we require

1. \( \phi(\mu_{t'}) = \mathbb{E}[c(\tau) | \tau > t^*] - c(t^*) \) and
2. \( \frac{1}{1 - \mathbb{P}(\tau > t^*)} \left( \phi(\mu_0) - \mathbb{P}(\tau > t^*)\phi(\mu_{t'}) \right) + c(t^*) - c(t^* + 1) \leq 0 \), which imply that \( \left( \phi(\mu_0) - \mathbb{P}(\tau > t^*)\phi(\mu_{t'}) \right) + c(t^*) - c(t^* + 1) \leq 0 \).

To see this, if 1. is false, then we can choose small \( \epsilon > 0 \) in such a way that the difference of \( \phi(\mu_{t'}) \) and \( \phi(\mu_{t}) \) is small so that the new obedience constraint is not violated. Therefore, \( \tau' \) is a feasible stopping time, which contradicts the sender’s optimality of \( \tau \). On the other hand, if 2. is false, then choosing arbitrarily small \( \epsilon > 0 \) slackens the LHS of (4). This implies \( \tau' \) is a feasible stopping time, which is again a contradiction.

By these two conditions, we have

\[
\phi(\mu_0) > \mathbb{E}[c(\tau)] = \mathbb{P}(\tau > t^*)\mathbb{E}[c(\tau) | \tau > t^*] + \mathbb{P}(\tau = t^*)c(t^*) \\
= \mathbb{P}(\tau > t^*)\phi(\mu_{t'}) + c(t^*) \\
\geq \phi(\mu_0) + 2c(t^*) - c(t^* + 1),
\]

which means that \( c(t^*) < c(t^* + 1) - c(t^*) \). From the definition of smoothing out \( f \) and \( c \) given in the main text and made formal in Online Appendix IV, we have \( c_k(t_k^* / k) = c_k((t_k^* + 1) / k) - c_k(t_k^* / k) \), where \( t_k^* \) is corresponding \( t^* \) to the model indexed \( k \). By smoothing out \( f \) and \( c \), for any sufficiently large \( k \), we have \( c_k(t + 1 / k) - c(t) < c'(1) \) for every \( t \in T \). Thus, \( c_k(t_k^* / k) = c_k((t_k^* + 1) / k) - c_k(t_k^* / k) < c'(1) = c_k(1) \). Therefore, \( t_k^* < k \). This implies \( c_k \) is linear at \( [0, (t_k^* + 1) / k] \), which means that \( c((t^* + 1) / k) = c(t^* / k) + c(1 / k) \leq 2c(t^' / k) \), which is a contradiction.

To generalize this result to any arbitrary information structure, we can see that its induced simple and obedient information structure yields a weakly better DM’s surplus because its induced stopping time is the same and a belief at every optimal history is guaranteed with full information. Therefore, the DM’s surplus under any arbitrary structure is still 0 (or uniformly bounded above by \( \epsilon \) when \( k \) is sufficiently large), as desired.
C.3 Proof of Proposition 6 (ii).

Proof of Proposition 6 (ii). We begin the proof of the second half, which is $\mathcal{P} = \bar{\mathcal{P}}$ when $\mu_0 \in \Phi^*$. Suppose that $\mu_0 \in \Phi^*$. Because the set of feasible stopping times is convex, so is the set of pairs of platform’s expected value and DM’s induced by feasible stopping times. Therefore, every point on the concave envelope of $(0, f^*)$ and the curve $F$ is induced by some dynamic information structure. It suffices to show that there is no information structure $I$ such that $S(I)$ Pareto dominates some point on the convex envelope of $(0, f^*)$ and the curve $F$. Suppose a contradiction that there is an information structure $I \in I^*$ that violates the previous statement. Define

$$\text{conv}(F \cup \{(0, f^*)\}) := \{s \in \mathbb{R}_+^2 : s \leq s^*, \exists s^* \in \text{co}(F \cup \{(0, f^*)\})\}$$

as the set of all points on co$(F \cup \{(0, f^*)\})$ and their free-disposal points. It is clear that conv$(F \cup \{(0, f^*)\})$ is a convex set.

The assumption of $I$ implies that $S(I) \not\in \text{conv}(F \cup \{(0, f^*)\})$. From Proposition 8, it is sufficient to consider the case that $I$ has a block structure generated by an increasing sequence $(t_1, \ldots, t_n)$ with terminal stopping times $(t_{n+1}, \bar{t})$ with the initial term $C_0 < \phi^*$. Consider the following two cases:

Case 1: $n = 0$. This means the information structure induces only two possible stopping times $t_{n+1}$ and $\bar{t}$. Thus, $S(I)$ is a convex combination of $F(t_{n+1})$ and $F(\bar{t})$, where $F(t)$ is defined as a pair of platform’s expected value and DM’s if the DM stops at time $t$ with probability 1 i.e.,

$$F(t) := (\phi^* - c(t), f(t)).$$

The obedient constraint a time 0 implies that $c(t_{n+1}) \leq \phi^*$ and $c(\bar{t}) - c(t_{n+1}) \leq \phi^*$. If $c(\bar{t}) \leq \phi^*$, then both $F(t_{n+1})$ and $F(\bar{t})$ are in $F$. This implies that $S(I) \in \text{co}(F)$, which is a contradiction. Thus, $c(\bar{t}) > \phi^*$ and $c(t_{n+1}) \leq \phi^*$. Then there exists $p \in [0, 1]$ such that $\phi^* = pc(t_{n+1}) + (1 - p)c(\bar{t})$. Consider information structure $I'$ that reveals full information at time $t_{n+1}$ with probability $p$ and at time $\bar{t}$ with probability $p$. $I'$ satisfies the obedient constraint at time 0 and time $t_{n+1}$ because $\phi^* = pc(\bar{t}) + (1 - p)c(t_{n+1})$ and $c(t_{n+1}) - c(\bar{t}) \leq \phi^*$.

Note that

$$\phi^* \geq \mathbb{E}^I[c(\tau(I))] = \mathbb{P}(\tau = t_{n+1})c(t_{n+1}) + (1 - \mathbb{P}(\tau = t_{n+1}))c(\bar{t}),$$

which means $\mathbb{P}(\tau(I) = t_{n+1}) \geq p = \mathbb{P}(\tau(I') = t_{n+1})$. This implies there is $\lambda \in [0, 1]$ such that $d(I) = \lambda d(I') + (1 - \lambda)\delta_{t_{n+1}}$, which means $S(I) = \lambda S(I') + (1 - \lambda)F(t_{n+1})$. However, $I'$ yields zero DM’s surplus, which means $(0, f^*)$ Pareto dominates $S(I')$. Therefore, $S(I')$ and $F(t_{n+1})$ are in $\text{conv}(F \cup \{(0, f^*)\})$. This implies $S(I) \in \text{conv}(F \cup \{(0, f^*)\})$, which is a contradiction.

Case 2: $n > 0$. Suppose $I''$ be an information structure generated by an increasing sequence $(t_1, \ldots, t_n)$ with terminal stopping times $(t_{n+1}, \bar{t})$ but with the initial term $C_0'' = \phi^*$. This is well defined because the obedience constraint at time 0 of the information structure $I$ implies $\phi^* \geq c(t_1)$. Therefore, $\tau(I') < \tau(I'') < t_1$ and $\tau(I) < \tau(I'') < t_1$ have the same distribution.

---

Footnote: Note that Proposition 8 requires that $C_0 = \phi^*$. However, we can impose another constraint $C_0 \leq \hat{C}$ for some constant $\hat{C} \leq \phi^*$ to the designer’s optimization problem. Then the result of Proposition 8 is still the same but changes from $C_0 = \phi^*$ to $C_0 = \hat{C}$.
because they share the same block structure and start with \( C_{t_1} = C'_{t_1} = \phi^* \). Consider that
\[
\mathbb{P}(\tau(I) > t_1) = \frac{C_0 - c(t_1)}{\phi^*} \leq \frac{\phi^* - c(t_1)}{\phi^*} \leq \mathbb{P}(\tau(I'') > t_1).
\]

This means there exists \( \lambda \in [0, 1] \) such that \( d(I) = \lambda d(I'') + (1 - \lambda)\delta_{t_1} \). However, \( I'' \) yields zero DM’s surplus, which means \((0, f^*)\) Pareto dominates \( S(I'') \). Therefore, \( S(I) = \lambda S(I'') + (1 - \lambda) F(t_1) \in \text{conv}(F \cup \{(0, f^*)\}) \), which is a contradiction.

From these two cases, we can imply that for every \( I \in I \), we must have \( S(I) \in \text{conv}(F \cup \{(0, f^*)\}) \), which implies the second half of the proposition statement.

Next, we will show that \( \mathcal{P} \leq \overline{\mathcal{P}} \) for any belief \( \mu_0 \). Proposition 8 implies that \((c_0, f_0) \in \mathcal{P}\) if and only if
\[
c_0 = \max_{(d(\tau), \phi(\tau)) \in \Lambda(\tau) \times \Lambda(\Theta)^T} \mathbb{E}[\phi(\mu_0) - c(\tau)]
\]
\[
\text{s.t. } \begin{align*}
\phi(\mu_t) &\geq \mathbb{E}[c(\tau) | \tau > t] - c(t) \quad \forall t \in T \quad \text{(Obedience)} \\
\mathbb{P}(\tau > t + 1) &\mu_{t+1} \leq \mathbb{P}(\tau > t) \mu_t \quad \text{(Boundary)}
\end{align*}
\]
\[
f_0 = \mathbb{E}[f(\tau)]. \quad \text{(Fix the the designer’s value)}
\]

We relax the obedience constraint to \( \phi^* \geq \mathbb{E}[c(\tau) | \tau > t] - c(t) \) for every \( t \in T \) and eliminate the boundary constraint. Applying \( \phi(\mu_0) \leq \phi^* \), we obtain
\[
c_0 \leq \max_{d(\tau) \in \Lambda(T)} \mathbb{E}[\phi^* - c(\tau)]
\]
\[
\text{s.t. } \begin{align*}
\phi^* &\geq \mathbb{E}[c(\tau) | \tau > t] - c(t) \quad \forall t \in T \quad \text{(Obedience)} \\
f_0 = \mathbb{E}[f(\tau)]. \quad \text{(Fix the the designer’s value)}
\end{align*}
\]
which is exactly the optimization problem of solving points on the Pareto frontier when \( \mu_0 \in \Phi^* \). The fact that \( \mathcal{P} = \overline{\mathcal{P}} \) when \( \mu_0 \in \Phi^* \) implies that \( c_0 \leq c^* \) where \((c^*, f_0) \in \overline{\mathcal{P}} \). Therefore, \((c_0, f_0) \leq (c^*, f_0) \), which implies \( \mathcal{P} \leq \overline{\mathcal{P}} \), as desired. \( \square \)

### C.4 Proof of Proposition 8

Before we prove Proposition 8, we first state and prove a helpful lemma.

**Lemma 8.** For every increasing function \( f : T \rightarrow \mathbb{R} \), there is a feasible stopping time \( \tau \) corresponding to an obedient sequence \((C_t)_{t \in T}\) such that
\[
\text{(i)} \quad \tau \in \arg\max_{\tau', d(\tau') \in \Omega} \mathbb{E}[f(\tau')].
\]
\[
\text{(ii)} \quad \text{For every } r < s \in T, \text{ the following statement is true:}
\]
\[
\text{If all of the following holds:}
\]
\[
\text{(ii.a) } C_{t-1} - C_t = c(t) - c(t - 1) \text{ for every } t \in \{r + 1, \ldots, s\};
\]
\[
\text{(ii.b) } C_{r-1} - C_r < c(r) - c(r - 1); \text{ and}
\]
\[
54
\]
Before we prove Lemma \[\text{8}\], let us briefly interpret it. The lemma states that fixing our value function \(f\), we can find an optimal stopping distribution corresponding to a obedience sequence \((C_t^*)_{t \in \mathcal{T}}\) which fulfills property (ii). Condition (ii.a) states that between periods \(r\) and \(s\), the DM obtains no information with full probability. Condition (ii.b) and (ii.c) states that the DM receives full information with some strictly positive probability at time \(r\) and \(s\) respectively. If these conditions are fulfilled, then either the DM is indifferent at time \(r\) \((C_r = \phi^r)\) or the DM must receive full information with probability one at some time between \(r + 1\) and \(s\).

**Proof of Lemma \[\text{8}\]** Fixing \(r < s\), we denote the if-then condition of Lemma \[\text{8}\](ii) as a statement whose truth value depends on \(r, s\) \((\text{the boundary times})\) and \(\tau\) \((\text{the stopping time in question})\). Denote this statement with \(S(r, s, \tau)\).

For any \(\tau \in \arg \max_{t' \leq \tau} \mathbb{E}[f(t')]\), define \(r(\tau)\) as, if it exists, the smallest number \(r \in \mathcal{T}\) such that there exists \(s > r\) which violates \(S(r, s, \tau)\). If no such number exists, define \(r(\tau) = +\infty\).

Pick any \(\tau \in \arg \max_{t' \leq \tau} \mathbb{E}[f(t')]\). If \(r(\tau) = +\infty\), there is nothing to show. Next suppose that \(r(\tau) < +\infty\). Denote \(r := r(\tau)\), and let \(s > r\) be such that the statement \(S(r, s, \tau)\) is false. Thus, \(C_r < \phi^r\) and \(C_t > c(t + 1) - c(t)\) for every \(t \in \{r + 1, \ldots, s\}\). Thus, we can modify \((C_r, \ldots, C_s)\) to either \((C_r + \epsilon, \ldots, C_s + \epsilon)\) or \((C_r - \epsilon, \ldots, C_s - \epsilon)\) without violating the obedience constraint for sufficiently small \(\epsilon > 0\).

Now we rewrite \(\mathbb{E}[f(\tau)]\) as a function of \((C_t)_{t \in \mathcal{T}}\) as follows:

\[
F\left((C_t)_{t \in \mathcal{T}}\right) := \mathbb{E}[f(\tau)] - f(0) = \sum_{t=0}^{\infty} \mathbb{P}(\tau > t) \left( f(t + 1) - f(t) \right)
= \sum_{t=0}^{\infty} \left( \prod_{u=0}^{t-1} \frac{C_u + c(u) - c(u + 1)}{C_{u+1}} \right) \left( f(t + 1) - f(t) \right).
\]

where the second equality follows by noting that we can rewrite conditional probabilities of \(\tau\) as a fraction of \(C_t\)'s, as in part (iii) of Lemma \[\text{1}\].

We define \(G(\epsilon) := F((C_t^\epsilon)_{t \in \mathcal{T}})\), where \(C_t^\epsilon = \begin{cases} C_t + \epsilon, & \text{for } t \in \{r, \ldots, s\} \\ C_t, & \text{otherwise.} \end{cases}\)

Denote \(\tau_{\text{ef}}\) as a feasible stopping time corresponding to \((C_t^\epsilon)_{t \in \mathcal{T}}\) whenever \((C_t^\epsilon)_{t \in \mathcal{T}}\) is obedient. Then

\[
(C_r + \epsilon)^2 \frac{dG}{d\epsilon} = -\left( \prod_{t=0}^{r-1} \frac{C_t + c(t) - c(t + 1)}{C_{t+1}} \right) \left( C_{r-1} + c(r - 1) - c(r) \right) \left( f(s) - f(r) \right)
+ \sum_{t'=s}^{\infty} \left( \prod_{t' \leq t' - 1, t' \notin \{r-1, \ldots, s\}} \frac{C_t + c(t) - c(t + 1)}{C_{t+1}} \right) \left( c(s + 1) - c(r) \right) \left( f(t' + 1) - f(t') \right)
=: H((C_t)_{t \in \mathcal{T}})
\]
which is not a function of \( \epsilon \). Therefore, \( \frac{d\phi}{d\epsilon} \geq 0 \) if and only if \( H((C_t)_{t \in T}) \geq 0 \). Because we can modify \((C_t)_{t \in T}\) to \((C_t^\epsilon)_{t \in T}\) and \((C_t^{-\epsilon})_{t \in T}\) without violating the obedience constraint for small \( \epsilon > 0 \), \( \frac{d\phi}{d\epsilon} \bigg|_{\epsilon=0} = 0 \), which, in turn, implies that \( H((C_t)_{t \in T}) = 0 \). This implies \( F((C_t)_{t \in T}) = F((C_t^\epsilon)_{t \in T}) \) for every \( \epsilon \in \mathbb{R} \). Therefore, \( \tau^{\epsilon,r} \in \text{argmax}_{\tau':d(\tau') \in D} \mathbb{E}[f(\tau')] \) as long as the sequence \((C_t^\epsilon)_{t \in T}\) is obedient.

We can then find \( \epsilon > 0 \) so that the statement \( S(r,s,\tau^{\epsilon,r}) \) is true and the sequence \((C_t^\epsilon)_{t \in T}\) is obedient by increasing \( \epsilon > 0 \) until the constraint \( C_r \leq \phi^* \) binds i.e., set \( \epsilon^* = \phi^* - C_r \). This implies \( r(\tau^{\epsilon,r}) > r = r(\tau) \). Note that the sequence \((C_t^\epsilon)_{t \in T}\) is weakly greater than \((C_t)_{t \in T}\) pointwise. We can iteratively apply this procedure to the stopping time \( \tau \) i.e., define a sequence of stopping times \( \{\tau_i\}_{i=0}^{\infty} \) such that

(i) \( \tau_0 = \tau \).

(ii) Given \( \tau_i \), we set \( \tau_{i+1} = \tau_i^{\epsilon_i,r(\tau_i)} \), where \( \epsilon_i = \phi^* - C_r(\tau(i)) \).

Therefore, the sequence \((r(\tau_i))_{i=0}^{\infty}\) is strictly increasing, which means that \( \lim_{i \to +\infty} r(\tau_i) = +\infty \). Moreover, \( \tau_i \in \text{argmax}_{\tau':d(\tau') \in D} \mathbb{E}[f(\tau')] \) for every \( i \in \mathbb{N} \). Suppose an obedient sequence \((C_t^\epsilon)_{t \in T}\) correspond to the feasible stopping time \( \tau_i \). For every \( t \in T \), the sequence \((C_t^\epsilon)_{i=0}^{\infty}\) is increasing and bounded above by \( \phi^* \). Thus, we can define \( C_r^* := \lim_{i \to +\infty} C_t^\epsilon \). Note that \((C_t^\epsilon)_{t \in T}\) is still obedient by putting limit to the obedient constraint. Suppose \((C_t^\epsilon)_{t \in T}\) induce a stopping time \( \tau^* \). Because \( \lim_{i \to +\infty} r(\tau_i) = +\infty \), we obtain \( r(\tau^*) = +\infty \) by putting limit to conditions (ii.a)-(ii.c). Moreover,

\[
\mathbb{E}[f(\tau^*)] - f(0) = \sum_{t=0}^{+\infty} \mathbb{P}(\tau^* > t) \left( f(t+1) - f(t) \right) \\
= \sum_{t=0}^{+\infty} \lim_{i \to +\infty} \mathbb{P}(\tau_i > t) \left( f(t+1) - f(t) \right) \\
= \lim_{i \to +\infty} \sum_{t=0}^{+\infty} \mathbb{P}(\tau_i > t) \left( f(t+1) - f(t) \right) \\
= \max_{\tau':d(\tau') \in D} \mathbb{E}[f(\tau')] - f(0),
\]

where the limit interchange follows from the dominated convergence theorem and

\[
\lim_{i \to +\infty} \sum_{t=0}^{+\infty} \mathbb{P}(\tau_i > t) \left| f(t+1) - f(t) \right| = \lim_{i \to +\infty} \sum_{t=0}^{+\infty} \mathbb{P}(\tau_i > t) \left( f(t+1) - f(t) \right) \\
= \max_{\tau':d(\tau') \in D} \mathbb{E}[f(\tau')] - f(0) < +\infty.
\]

Therefore, \( \tau = \tau^* \) makes Lemma 8 hold.

**Proof of Proposition 8** Fix an increasing function \( f : T \to \mathbb{R} \), we can find a feasible stopping time \( \tau \) corresponding to an obedient sequence \((C_t)_{t \in T}\) as shown in Lemma 8. Define

\[
T_{c_{t-1}} := \{ t \in T : C_{t-1} - C_t < c(t) - c(t-1) \}.
\]
Then we arrange all members of the set \( \mathcal{T}_{\text{no}} \) as an increasing sequence \( (t_i)_{i=1}^{n+1} \). Note that \( n \) can be \( +\infty \). Lemma \( \text{[5]} \) implies that \( C_i = \phi^* \) for every \( i \in \{1, \ldots, n\} \), and Proposition \( \text{[6]} \) implies \( C_0 = \phi^* \), which corresponds to indifferent blocks. Suppose that \( n < +\infty \). If \( t > t_{n+1} \), we have

\[
C_t - C_{t_{n+1}} = \sum_{t' = t_{n+1}+1}^{t} (C_{t'} - C_{t'-1}) = \sum_{t' = t_{n+1}}^{t-1} (c(t') - c(t' + 1)) = c(t_{n+1}) - c(t),
\]

which converges to \(-\infty\) as \( t \to +\infty \). To make the constraint \( C_t \geq c(t + 1) - c(t) > 0 \), the DM must receive full information at some deterministic time, say \( \bar{t} \), with probability one. Therefore, \( C_{\bar{t}-1} = c(\bar{t}) - c(\bar{t} - 1) \). By \( \text{[5]} \), we obtain \( C_{t_{n+1}} = c(\bar{t}) - c(t_{n+1}) \), which concludes a terminal block.

\[ \square \]

C.5 Proof of Proposition \( \text{[9]} \)

\textbf{Proof of Proposition} \( \text{[9]} \) Consider a stopping time \( \tau \) generated by an increasing sequence \( (t_i)_{i=1}^{n+1} \) and a terminal time \( \bar{t} > t_{n+1} \). If \( t_1 > 1 \), take a new increasing sequence

\[
(t'_i)_i = (1, 2, \ldots, t_1 - 1, t_1, t_2, \ldots, t_{n+1})
\]

and the same terminal time \( \bar{t} \) to construct a stopping time \( \tau' \) with our algorithm. We see that

\[
\mathbb{E}[c(\tau)] = \mathbb{P}(\tau = t_1)c(t_1) + \mathbb{P}(\tau > t_1)\mathbb{E}[c(\tau) \mid \tau > t_1]
\]

and

\[
\mathbb{E}[c(\tau')] = \sum_{t=1}^{t_1} \mathbb{P}(\tau' = t)c(t) + \mathbb{P}(\tau' > t_1)\mathbb{E}[c(\tau') \mid \tau' > t_1]
\]

\[
= \sum_{t=1}^{t_1} \mathbb{P}(\tau' = t)c(t) + \mathbb{P}(\tau' > t_1)\mathbb{E}[c(\tau) \mid \tau > t_1]
\]

because \( \tau|\tau > t_1 \) and \( \tau'|\tau > t_1 \) have the same distribution. Consider that

\[
\mathbb{P}(\tau' > t_1) = \prod_{t=1}^{t_1} \left(1 - \frac{c(t) - c(t - 1)}{\phi^*}\right) \geq 1 - \frac{\sum_{t=1}^{t_1} (c(t) - c(t - 1))}{\phi^*}
\]

\[
= 1 - \frac{c(t_1)}{\phi^*} = \mathbb{P}(\tau > t_1),
\]

where the inequality follows from the fact that \( (1 - x)(1 - y) \geq 1 - x - y \) for every \( x, y \geq 0 \). We know that \( \mathbb{E}[c(\tau')] = \mathbb{E}[c(\tau)] = \phi^* \). Thus,

\[
0 = \mathbb{E}[c(\tau')] - \mathbb{E}[c(\tau)]
\]

\[
= \sum_{t=1}^{t_1} \mathbb{P}(\tau' = t)c(t) + \left(\mathbb{P}(\tau' > t_1) - \mathbb{P}(\tau > t_1)\right)\mathbb{E}[c(\tau) \mid \tau > t_1] - \mathbb{P}(\tau = t_1)c(t_1),
\]
and rearranging, we have
\[
\left( \sum_{t=1}^{t_1} \frac{\mathbb{P}(\tau' = t)}{\mathbb{P}(\tau = t_1)} c(t) \right) + \left( \frac{\mathbb{P}(\tau > t_1) - \mathbb{P}(\tau > t_1)}{\mathbb{P}(\tau = t_1)} \cdot \mathbb{E}[c(\tau) \mid \tau > t_1] \right) = c(t_1).
\]
Because \( \frac{\mathbb{P}(\tau > t_1) - \mathbb{P}(\tau > t_1)}{\mathbb{P}(\tau = t_1)} \geq 0 \) and \( f \) is more convex than \( c, f \circ c^{-1} \) is convex and applying Jensen’s inequality, we have
\[
\left( \sum_{t=1}^{t_1} \frac{\mathbb{P}(\tau' = t)}{\mathbb{P}(\tau = t_1)} f(t) \right) + \left( \frac{\mathbb{P}(\tau > t_1) - \mathbb{P}(\tau > t_1)}{\mathbb{P}(\tau = t_1)} \cdot \mathbb{E}[f(\tau) \mid \tau > t_1] \right) \geq f(t_1).
\]
This implies \( \mathbb{E}[f(\tau')] \geq \mathbb{E}[f(\tau)] \). Iterating on the above argument, we see that to maximize \( \mathbb{E}[f(\tau)] \), it weakly better to choose the increasing sequence \((t_i)_{i=1}^{n+1}\) such that \( t_i = i \) for every \( i \in \{1, \ldots, n\} \).

We will now show that having choosing a terminal time \( \text{i.e.}, n < +\infty \) is dominated. It is without loss to consider the case with \( n = 0 \) i.e., \( t_n = 0 \). Suppose that the sequence \((t_1, \tilde{t})\) generates a feasible stopping time \( \tau \). Now consider an alternate feasible stopping time \( \tau'' \) generated the sequence \((t_i'')_{i \in \mathcal{T}} = (t_1, t_1 + 1, \ldots) \). Observe that
\[
\mathbb{P}(\tau'' = t_1) = \frac{c(t_1)}{\phi^*} \geq \frac{(c(\tilde{t}) - \phi^*)}{c(\tilde{t}) - c(t_1)} = \mathbb{P}(\tau = t_1)
\]
Since \( \mathbb{E}[c(\tau'')] = \mathbb{E}[c(\tau)] = \phi^* \),
\[
0 = \mathbb{E}[c(\tau'')] - \mathbb{E}[c(\tau)] = (\mathbb{P}(\tau'' = t_1) - \mathbb{P}(\tau = t_1))c(t_1) + \mathbb{E}[c(\tau'') \mid \tau'' > t_1] \mathbb{P}(\tau'' > t_1) - \mathbb{P}(\tau = t_1) c(t_1),
\]
which then implies
\[
\left( \frac{\mathbb{P}(\tau'' = t_1) - \mathbb{P}(\tau = t_1)}{\mathbb{P}(\tau = \tilde{t})} \cdot c(t_1) \right) + \left( \frac{\mathbb{P}(\tau'' > t_1)}{\mathbb{P}(\tau = \tilde{t})} \cdot \mathbb{E}[c(\tau'') \mid \tau'' > t_1] \right) = c(\tilde{t}).
\]
Because \( \frac{\mathbb{P}(\tau' = t_1) - \mathbb{P}(\tau = t_1)}{\mathbb{P}(\tau = t_1)} \geq 0 \) and \( f \) is more convex than \( c, f \circ c^{-1} \) is convex and applying Jensen’s inequality,
\[
\left( \frac{\mathbb{P}(\tau'' = t_1) - \mathbb{P}(\tau = t_1)}{\mathbb{P}(\tau = \tilde{t})} \cdot f(t_1) \right) + \left( \frac{\mathbb{P}(\tau'' > t_1)}{\mathbb{P}(\tau = \tilde{t})} \cdot \mathbb{E}[f(\tau'') \mid \tau'' > t_1] \right) \geq f(\tilde{t}).
\]
Therefore, \( \mathbb{E}[f(\tau'')] \geq \mathbb{E}[f(\tau)] \). We have shown that choosing any terminal time is always weakly dominated. But we have also shown that it is always weakly better to choose sequences which are consecutive in \( \mathcal{T} \). This implies that every feasible stopping time \( \tau'' \) must satisfy
\[
\mathbb{E}[f(\tau'')] \leq \mathbb{E}^{R(\mathcal{T})}[f(\tau)]
\]
which concludes the proof. \( \square \)
Online Appendix to ‘Attention Capture’
Andrew Koh and Sivakorn Sanguanmoo†
FOR ONLINE PUBLICATION ONLY

Online Appendix I: Maximal belief paths and attainment of sup

In this online appendix we introduce a topology on the space of belief paths and feasible stopping times. This will establish the existence of a maximal belief path used in the Proof of Proposition 4 in Appendix B. It will also ensure that the supremum of the sender’s problem can be attained, as claimed in the main text.

I.1 Topological space of belief paths and the existence of a maximal belief path. Our first step is, fixing a distribution of stopping time $d(\tau)$, to study the set of belief paths that satisfy Proposition 3 in order to verify that $\tau$ is a feasible stopping time. Recall that the set of belief paths $\mathcal{W} = (\Delta(\Theta))^T$. We endow $\mathcal{W}$ with the product topology of the weak topology on $\Delta(\Theta)$. Because $\Delta(\Theta)$ endowed with the weak topology is compact and metrizable, so is $\mathcal{W}$ by Tychonoff theorem. Recall that we defined $\mathcal{W}(\tau) \subset \mathcal{W}$ as the set of belief paths corresponding to stopping time $\tau$.

Lemma 9. $\mathcal{W}(\tau)$ is compact for every feasible stopping time $\tau$.

Proof. We show that $\mathcal{W}(\tau)$ is sequentially compact instead. Consider $((\mu_t^i)_{t \in T})_{i=0}^\infty \subset \mathcal{W}(\tau)$. Because $\mathcal{W}$ is compact, there exists a convergent subsequence $((\mu_t^{i_n})_{t \in T})_{n=0}^\infty$ as $n \to \infty$. Suppose that $(\mu_t^{i_n})_{t \in T} \to (\mu_t^*)_{t \in T} \in \mathcal{W}$ as $n \to \infty$. This implies $\mu_t^{i_n} \to \mu_t^*$ as $n \to \infty$ under the weak topology on $\Delta(\Theta)$ for every $t \in T$. It suffices to show that $(\mu_t^*)_{t \in T} \in \mathcal{W}(\tau)$. Because $\phi$ is continuous under the weak topology on $\Delta(\Theta)$, for every $t \in T$, we have

$$\phi(\mu_t^*) = \lim_{n \to \infty} \phi(\mu_t^{i_n}) \geq \mathbb{E}[\epsilon(\tau) \mid \tau > t] - c(t),$$

where the inequality follows by $(\mu_t^{i_n})_{t \in T} \in \mathcal{W}(\tau)$. This implies the obedience constraint holds for the belief path $(\mu_t^*)_{t \in T}$. Moreover, we have $\mathbb{P}(\tau > t + 1)\mu_t^{i_n}(\theta) \leq \mathbb{P}(\tau > t)\mu_t^{i_n}(\theta)$ for every $t \in T$. Therefore,

$$\mathbb{P}(\tau > t + 1)\mu_{t+1}^*(\theta) = \lim_{n \to \infty} \mathbb{P}(\tau > t + 1)\mu_t^{i_n}(\theta) \leq \mathbb{P}(\tau > t)\mu_t^{i_n}(\theta) = \mathbb{P}(\tau > t)\mu_t^*(\theta),$$

which implies the boundary constraint. This implies $(\mu_t^*)_{t \in T} \in \mathcal{W}(\tau)$, as desired. □

We now prove Proposition 11 that maximal belief paths exist.

Proof of Proposition 11. Define $F : \mathcal{W} \to \mathbb{R}$ such that $F((\mu_t)_{t \in T}) = \sum_{t=0}^\infty \frac{1}{2^t} \phi(\mu_t)$. $F$ is well-defined because $|F((\mu_t)_{t \in T})| \leq \sum_{t=0}^\infty \frac{1}{2^t} \phi^* \leq 2\phi^*$. Moreover, for every $\epsilon > 0$ and every $T > 1 + \log(\phi^*/\epsilon) / \log 2$, we have $\sum_{t=T}^\infty \frac{1}{2^t} \phi(\mu_t) \leq \frac{1}{2^{T-1}}\phi^* < \epsilon$. This implies $\sum_{t=0}^\infty \frac{1}{2^t} \phi(\mu_t)$ is uniformly

---

†MIT Department of Economics; email: akoh@mit.edu and sanguanm@mit.edu
convergent. Therefore, $F$ is continuous under $\mathcal{W}$. By Lemma 9 and $\mathcal{W}(\tau)$ is nonempty, we can find

$$(\mu^*_t)_{t \in \mathcal{T}} \in \arg\max_{(\mu_t)_{t \in \mathcal{T}} \in \mathcal{W}(\tau)} F((\mu_t)_{t \in \mathcal{T}}).$$

Suppose that there exists $(\mu_t)_{t \in \mathcal{T}} \in \mathcal{W}(\tau)$ such that $\phi(\mu_t) \geq \phi(\mu^*_t)$ for every $t \in \mathcal{T}$ and the inequality is strict for some $t \in \mathcal{T}$. This directly implies $F((\mu_t)_{t \in \mathcal{T}}) > F((\mu^*_t)_{t \in \mathcal{T}})$, which contradicts the optimality of $(\mu^*_t)_{t \in \mathcal{T}}$. Therefore, $(\mu^*_t)_{t \in \mathcal{T}}$ is a maximal belief path under $\tau$. \hspace{1cm} \Box

I.2 Topological space of feasible stopping times and the existence of a solution to the sender’s optimization problem. We will create a topology of feasible stopping times by using properties of the set $\mathcal{W}$. However, we need the regularity assumption of $c$ being smooth. We start with the following lemma.

**Lemma 10.** Suppose $c$ be a smooth function. For any $\epsilon > 0$, there exists $T \in \mathcal{T}$ such that for every $t > T$, we have $\mathbb{E}[c(\tau)1\{\tau > t\}] < \epsilon$ for every feasible stopping time $\tau$.

**Proof of Lemma 10.** We know from Lemma 1 that given feasible stopping time $\tau$ we can find a sequence $(C_t)_{t \in \mathcal{T}}$ such that

$$\mathbb{P}(\tau > t + 1 \mid \tau > t) = \frac{C_t + c(t) - c(t + 1)}{C_{t+1}}$$

for every $t \in \mathcal{T}$, which means $C_t \geq c(t + 1) - c(t) \geq \bar{c}$. Thus,

$$\mathbb{P}(\tau = t) \leq \mathbb{P}(\tau > t - 1) = \prod_{s=0}^{t-2} \mathbb{P}(\tau > s + 1 \mid \tau > s)$$

$$= \prod_{s=0}^{t-2} \frac{C_s + c(s) - c(s + 1)}{C_{s+1}}$$

$$= \frac{C_0 - c(1)}{C_1} \prod_{s=1}^{t-2} \left(1 - \frac{c(s + 1) - c(s)}{\phi^*}\right)$$

$$\leq \frac{\phi^* - c(1)}{\bar{c}} \sum_{s=t}^{t-2} \left(1 - \frac{c}{\bar{c}}\right)^{s-2}.$$ 

Note that $c(s) \leq \bar{c} \cdot s$. Therefore,

$$\mathbb{E}[c(\tau)1\{\tau > t\}] \geq \sum_{s \geq t} c(s) \mathbb{P}(\tau = s) \leq \frac{\bar{c} (\phi^* - c(1))}{\bar{c}} \sum_{s \geq t} s \left(1 - \frac{c}{\phi^*}\right)^{s-2}.$$ 

We know that $\sum_{s=0}^{\infty} s \left(1 - \frac{c}{\phi^*}\right)^{s-2}$ converges. For any $\epsilon > 0$, there exists $T \in \mathcal{T}$ such that, for
every $t > T$, we have

$$
\mathbb{E}[c(\tau)1\{\tau > t\}] \leq \frac{\epsilon(\phi^* - c(1))}{c} \sum_{s > t} s \left(1 - \frac{c}{\phi^*}\right)^{s-2} \leq \epsilon,
$$

for every feasible stopping time $\tau$, as desired. □

We introduce a metric $\Delta$ of $\mathcal{D}$ as follows: for any probability measures $d_1, d_2 \in \mathcal{D}$,

$$
\Delta(d_1, d_2) := \sum_{t=0}^{\infty} |d_1(t) - d_2(t)|c(t).
$$

This metric is well-defined because the obedience constraint at $t = 0$ implies

$$
\sum_{t=0}^{\infty} |d_1(t) - d_2(t)|c(t) \leq \sum_{t=0}^{\infty} d_1(t)c(t) + \sum_{t=0}^{\infty} d_2(t)c(t) \leq 2\phi^* < \infty.
$$

It is easy to verify that this is indeed a metric. We use the metric $\Delta$ to construct a topological space of $\mathcal{D}$ and obtain the following proposition

**Proposition 12.** $\mathcal{D}$ is compact under the metric $\Delta$.

**Proof of Proposition 12.** Define $\Psi : \mathcal{D} \rightarrow l^1$ such that $(\Psi(d))_t = c(t)d(t)$ for every $d \in \mathcal{D}$ and $t \in \mathcal{T}$. It is easy to see that $\Psi(d) \in l^1$ and $\|\Psi(d)\|_1 \leq \phi^*$ by the obedience constraint at $t = 0$. Thus, $\Psi(\mathcal{D})$ is bounded.

We show that $\Psi(\mathcal{D})$ is closed. Consider any convergent sequence $(\Psi(d_i))_{i \in \mathbb{N}}$ under $l^1$ norm where $d_i \in \mathcal{D}$ for every $i \in \mathbb{N}$. Suppose that $\Psi(d_i) \rightarrow (s^*_t)_{t \in \mathcal{T}}$ under $l^1$ norm. The function $(s_t)_{t \in \mathbb{N}} \in l^1 \mapsto \sum_{t=0}^{\infty} s_t/c(t)$ is continuous because, for any $\epsilon > 0$ and $(s_t)_{t \in \mathbb{N}}$, $(s_t')_{t \in \mathbb{N}} \in l^1$, if $\|s' - s\|_1 \leq \epsilon c(1)$, then

$$
\left|\sum_{t=0}^{\infty} \frac{s'_t - s_t}{c(t)}\right| \leq \frac{1}{c(1)}\|s' - s\|_1 \leq \epsilon.
$$

Consider that $\sum_{t=0}^{\infty} (\Psi(d_i))_t/c(t) = \sum_{t=0}^{\infty} d_i(t) = 1$ for every $i \in \mathbb{N}$. This implies $\sum_{t=0}^{\infty} s^*_t/c(t) = 1$. Therefore, we can construct a distribution $d^* \in \Delta(\mathcal{T})$ in such a way that $d^*_t = s^*_t/c(t)$ for every $t \in \mathcal{T}$, which means $\Psi(d^*) = (s^*_t)_{t \in \mathcal{T}}$. Suppose the corresponding belief path to the distribution of stopping time $d_i$ is $(\mu^*_i)_{t \in \mathcal{T}} \in \mathcal{W}$. Because $\mathcal{W}$ is sequentially compact, there exists a subsequence $(\mu^*_i)_{t \in \mathcal{T}}$ converging to some belief path $(\mu^*_t)_{t \in \mathcal{T}}$. We will show that $(d^*, (\mu^*_t)_{t \in \mathcal{T}}) \in \mathcal{V}$. Let $\tau_n$ be the stopping time corresponding to $(d_n, (\mu^*_t)_{t \in \mathcal{T}})$.
1. Obedience constraint: for any \( n \in \mathbb{N} \) and \( t \in \mathcal{T} \),

\[
\mathbb{P}(\tau_{i_n} > t) \phi(\mu_{t_n}^n) \geq \mathbb{E}[c(\tau_{i_n})1\{\tau_{i_n} > t\}] - \mathbb{P}(\tau_{i_n} > t)c(t)
\]

\[
\implies \sum_{s > t} d_{i_n}(s) \phi(\mu_{t_n}^n) \geq \sum_{s > t} d_{i_n}(s)(c(s) - c(t))
\]

\[
\lim_{n \to \infty} \sum_{s > t} d_{i_n}(s) \phi(\mu_{t_n}^n) \geq \lim_{n \to \infty} \sum_{s > t} d_{i_n}(s)(c(s) - c(t))
\]

\[
\implies \sum_{s > t} d^*(s) \phi(\mu_{t}^n) \geq \sum_{s > t} d^*(s)(c(s) - c(t)),
\]

which implies the obedient constraint for \((d^*, \mu_t^*)_{t \in \mathcal{T}}\). The summation interchange follows by the fact that \( \phi(\mu) \leq \mu^* \) and Lemma 10 which implies both summations are uniformly convergent.

2. Boundary constraint: for any \( n \in \mathbb{N}, t \in \mathcal{T}, \) and \( \theta \in \Theta \),

\[
\mathbb{P}(\tau_{i_n} > t) \mu_{t+1}^n(\theta) \geq \mathbb{P}(\tau_{i_n} > t + 1)\mu_{t+1}^n(\theta)
\]

\[
\implies \sum_{s > t} d_{i_n}(s) \mu_{t}^n(\theta) \geq \sum_{s > t+1} d_{i_n}(s) \mu_{t+1}^n(\theta).
\]

With the same argument, applying limit \( n \) to \( \infty \) implies the boundary constraint for \((d^*, (\mu_t^*)_{t \in \mathcal{T}})\).

Therefore, \((d^*, (\mu_t^*)_{t \in \mathcal{T}}) \in \mathcal{V}\), which means \( d^* \in \mathcal{D}\). Thus, \((s_t^*)_{t \in \mathcal{T}} = \Psi(d^*) \in \Psi(\mathcal{D})\), so \(\Psi(\mathcal{D})\) is closed.

Next, We show that \(\Psi(\mathcal{D})\) is equismall at infinity. From Lemma 10 for any \( \epsilon > 0 \), there exists \( T \in \mathcal{T} \) such that for every \( t > T \) and every \( I \in I^* \), we have

\[
\epsilon > \mathbb{E}[c(\tau)1\{\tau(I) > t\}] = \sum_{s > t} c(s)d(I)(s) = \sum_{s > t} \left(\Psi(d(I))\right)_s,
\]

which implies that \(\Psi(\mathcal{D})\) is equismall at infinity. Therefore, by Theorem 44.2 from [Treves (1990)] we have \(\Psi(\mathcal{D})\) is compact under \(l^1\) norm. Note that, for any \( d_1, d_2 \in \mathcal{D}\), \(\Delta(d_1, d_2) = \|\Psi(d_1) - \Psi(d_2)\|_1\), which means that \(\Psi\) is continuous. Therefore, because \(\Psi(\mathcal{D})\) is compact, \(\mathcal{D}\) is compact under the metric \(\Delta\).

This proposition directly implies the existence of the sender’s optimization problem under the regularity condition of \(f\). To see this, the function \(d(\tau) \mapsto \sum_{t \in \mathcal{T}} d(t)f(t)\) is continuous under the topology of \(\mathcal{D}\) because, for any \( \epsilon > 0 \) and \( d_1, d_2 \in \mathcal{D}\), if \(\Delta(d_1, d_2) < \epsilon / L\), then

\[
\left| \sum_{t \in \mathcal{T}} d_1(t)f(t) - \sum_{t \in \mathcal{T}} d_2(t)f(t) \right| \leq \sum_{t \in \mathcal{T}} |d_1(t) - d_2(t)|f(t) \leq L\Delta(d_1, d_2) = \epsilon.
\]

This implies there is a solution to the optimization problem of the sender.
Online Appendix II: Proof of Proposition 10

Before we prove Proposition 10 we first develop several lemmas establishing properties of \( \text{cconv}_t(f) \) and \( s^* \).

**Lemma 11.** For every \( t \in T, s^*(t) > t^* \), \( \text{cconv}_t(f) \) is concave, and \( \text{cconv}_t(f) \geq f \) over the domain of \( \text{cconv}_t(f) \).

**Proof of Lemma 11.** We will show that \( \text{cconv}_t(f)(s) \) has decreasing differences. Within this proof, we abuse notation \( s^*(t) \) by \( s^* \). If \( f(s^* + 1) - f(s^*) \geq f(s^*) - f(s^* - 1) \), then

\[
(s^* - 1 - t)(f(s^*) - f(s^* - 1)) - f(s^* - 1) = (s^* - t)(f(s^*) - f(s^* - 1)) - f(s^*) \\
\leq (s^* - t)(f(s^* + 1) - f(s^*)) - f(s^*) \\
< -f(t),
\]

which implies that \( f(s^*) - f(s^* - 1) < \frac{f(s^* - 1 - t)}{s^* - t} \), which contradicts the definition of \( s^* \).

Thus, \( f(s^* + 1) - f(s^*) < f(s^*) - f(s^* - 1) \), which implies for every \( s > s^* \), we have \( f(s + 1) - f(s) < f(s + 1) - f(s) \). The definition of \( \text{cconv}_t(f) \) implies \( \text{cconv}_t(f)(s + 1) - \text{cconv}_t(f)(s) < \text{cconv}_t(f)(s) - \text{cconv}_t(f)(s - 1) \) for every \( s > s^* \). The last inequality is also true for every \( s < s^* \) because \( \text{cconv}_t(f) \) is linear over the domain \( \{t, \ldots, s^*\} \).

It suffices to show that \( \text{cconv}_t(f)(s^* + 1) - \text{cconv}_t(f)(s^*) < \text{cconv}_t(f)(s^*) - \text{cconv}_t(f)(s^* - 1) \), which is followed by

\[
\text{cconv}_t(f)(s^*) - \text{cconv}_t(f)(s^* - 1) = \frac{f(s^*) - f(t)}{s^* - t} \\
> f(s^* + 1) - f(s^*) \\
= \text{cconv}_t(f)(s^* + 1) - \text{cconv}_t(f)(s^*),
\]

as desired. To see that \( \text{cconv}_t(f) \geq f \) over the domain of \( \text{cconv}_t(f) \), it suffices to consider the domain of \( \{t, \ldots, s^*\} \). We will show by induction on \( \Delta \in \{0, \ldots, s^* - t\} \) that \( \text{cconv}_t(f)(s^* - \Delta) \geq f(s^* - \Delta) \). It is clear that the inequality is true when \( \Delta = 0 \). Assume that the inequality is true for some \( \Delta \). By the definition of \( s^* \), we must have

\[
f(s^* - \Delta) - f(s^* - \Delta - 1) \geq \frac{f(s^* - \Delta) - f(t)}{s^* - \Delta - t},
\]

which implies

\[
f(s^* - \Delta - 1) \leq \left(1 - \frac{1}{s^* - \Delta - t}\right)f(s^* - \Delta) + \frac{1}{s^* - \Delta - t}f(t) \\
\leq \left(1 - \frac{1}{s^* - \Delta - t}\right)\text{cconv}_t(f)(s^* - \Delta) + \frac{1}{s^* - \Delta - t}\text{cconv}_t(f)(t) \\
= \text{cconv}_t(f)\left(1 - \frac{1}{s^* - \Delta - t}\right)(s^* - \Delta) + \frac{t}{s^* - \Delta - t}\right) \\
= \text{cconv}_t(f)(s^* - \Delta - 1),
\]

as desired. \( \square \)
We introduce properties of \( s^* \) that will be useful in the proof of Proposition 10.

**Lemma 12 (Properties of \( s^* \)).**

(i) For every \( t, t' \in \mathcal{T} \) such that \( t \leq t' < s^*(t) \), we have \( s^*(t') \leq s^*(t) \).

(ii) For every \( t < t' \in \mathcal{T} \), we have \( s^*(t') - t' \leq s^*(t) - t \). This becomes a strict inequality if \( t < t^* \).

(iii) For every \( t, t_1, t_2 \in \mathcal{T} \) such that \( t \leq t_1 < t_2 \leq s^*(t) \), we have

\[
f(t_1) < \frac{t_2 - t_1}{t_2 - t} f(t) + \frac{t_1 - t}{t_2 - t} f(t_2).
\]

**Proof.** (i) From Lemma 11, we know \( \text{cconv}_t(f)(t') \geq f(t') \). Thus,

\[
f(s^*(t) + 1) - f(s^*(t)) \leq \frac{f(s^*(t)) - f(t)}{s^*(t) - t} = \frac{\text{cconv}_t(f)(s^*(t)) - \text{cconv}_t(f)(t')}{s^*(t) - t'} 
\]

which implies that \( s^*(t') \leq s^*(t) \), as desired.

(ii) It is sufficient to show the statement in the case of \( t' = t + 1 \). If \( s^*(t) - t > 1 \), then \( t < t + 1 < s^*(t) \). From (i), we have \( s^*(t + 1) \leq s^*(t) \), which implies that \( s^*(t + 1) - (t + 1) < s^*(t) - t \). On the other hand, if \( s^*(t) - t = 1 \), then \( f(t + 2) - f(t + 1) < f(t + 1) - f(t) \), which means \( t \geq t^* \). This implies that \( f(t + 3) - f(t + 2) < f(t + 2) - f(t + 1) \), so \( s^*(t + 1) = t + 2 \). Therefore, \( s^*(t + 1) - (t + 1) = s^*(t) - t \), as desired. From the proof, we get the strict inequality when \( t < t^* \).

(iii) Consider any \( t' \in \{t + 1, \ldots, s^*(t)\} \). The definition of \( s^*(\cdot) \) implies that \( f(t') - f(t' - 1) > \frac{f(t' - 2) - f(t' - 1)}{t' - t - 1} \). Therefore,

\[(t' - t - 1)f(t' - 1) - (t' - t - 2)f(t') < f(t + 1).\]

If \( t' < s^*(t) \), we also have \( (t' - t)f(t') - (t' - t - 1)f(t' + 1) < f(t + 1) \), which implies

\[(t' - t)f(t' - 1) - (t' - t - 2)f(t' + 1) < 2f(t + 1).\]

With a simple induction, we obtain the equation in (iii). \( \square \)

We are now ready to prove Proposition 10.

**Proof of Proposition 10.** Choose a feasible stopping time \( \tau \) that maximizes \( \mathbb{E}[f(\tau)] \). Suppose that \( \tau \) is generated by an increasing sequence \( t_1, \ldots, t_n \) with a pair of terminal times \( (t_{n+1}, t) \). Without loss of generality, assume \( t_{n+1} > t_n \). If \( \bar{t} - t_{n+1} = \phi^* / c \), then the DM is indifferent between continuing paying attention and stopping at time \( t_{n+1} \). Thus, an increasing sequence \( t_1, \ldots, t_{n+1} \) with a pair of terminal times \( (\bar{t}, \bar{t}) \) also generates \( \tau \). It is sufficient to consider the
case that \( t - t_{n+1} < \phi^*/c \). We will show that \( t_i = i \) for every \( i \in \{1, \ldots, n\} \) and \( t_{n+1} = t = n + \phi^*/c \).

The proof consists of two following steps:

**Step 1: Terminal Stage.** We will show that \( t_{n+1} = i \). Suppose a contradiction that \( t_{n+1} < i \). Define stopping times \( \tau_1, \tau_2 \) generated by the same increasing sequence \( t_1, \ldots, t_n \) but with different terminal stopping times \( (t_{n+1}, i) \) and \( (t_{n+1} + 1, i) \). The stopping time \( \tau_1 \) is well-defined because \( t_{n+1} - 1 \geq t_n \) and \( t - (t_{n+1} - 1) \leq \phi^*/c \). The stopping time \( \tau_2 \) is also well-defined because \( i > t_{n+1} \). We obtain the following equations:

\[
\begin{align*}
\mathbb{E}[f(\tau) \mid \tau > t_n] &= \frac{i - \phi^*/c - t_n}{t - t_{n+1}} f(t_{n+1}) + \frac{\phi^*/c + t_n - t_{n+1}}{t - t_{n+1}} f(i) \\
\mathbb{E}[f(\tau_1) \mid \tau_1 > t_n] &= \frac{i - \phi^*/c - t_n}{t - t_{n+1} + 1} f(t_{n+1} - 1) + \frac{\phi^*/c + t_n - t_{n+1} + 1}{t - t_{n+1} + 1} f(i), \\
\mathbb{E}[f(\tau_2) \mid \tau_2 > t_n] &= \frac{i - \phi^*/c - t_n}{t - t_{n+1} - 1} f(t_{n+1} + 1) + \frac{\phi^*/c + t_n - t_{n+1} - 1}{t - t_{n+1} - 1} f(i).
\end{align*}
\]

Note that, if \( t_{n+1} = t_n + 1 \), the second equation still holds. To see this, we have \( \phi^*/c + t_{n+1} > i \geq \phi^*/c + t_n \), so \( i = \phi^*/c + t_n \), which implies that the coefficient of \( f(t_{n+1} - 1) \) is equal to 0. Because \( \tau, \tau_1, \) and \( \tau_2 \) are identical until time \( t_n \), the optimality of \( \tau \) implies

\[
\mathbb{E}[f(\tau_1)] \leq \mathbb{E}[f(\tau)] \implies \mathbb{E}[f(\tau_1) \mid \tau > t_n] \leq \mathbb{E}[f(\tau) \mid \tau > t_n] \\
\implies (i - t_{n+1}) f(t_{n+1} - 1) + f(i) \leq (i - t_{n+1} + 1) f(t_{n+1})
\]

and

\[
\mathbb{E}[f(\tau_2)] \leq \mathbb{E}[f(\tau)] \implies \mathbb{E}[f(\tau_2) \mid \tau > t_n] \leq \mathbb{E}[f(\tau) \mid \tau > t_n] \\
\implies (i - t_{n+1}) f(t_{n+1} + 1) \leq (i - t_{n+1} - 1) f(t_n) + f(i).
\]

Because \( i - t_{n+1} > 0 \), these two inequalities altogether imply \( f(t_{n+1} + 1) - f(t_{n+1}) < f(t_{n+1}) - f(t_{n+1} - 1) \). By the definition of S-shaped, we obtain \( \tau^* < t_{n+1} < t \), so \( f \) is concave in the interval \([t_{n+1}, i]\). Consider a stopping time \( \tau' \) generated by the same increasing sequence \( t_1, \ldots, t_n \) but with a pair of terminal stopping times \( (t_n + \phi^*/c, t_n + \phi^*/c) \). We obtain \( \mathbb{E}[f(\tau') \mid \tau' > t_n] = f(t_n + \phi^*/c) \), which implies that

\[
\mathbb{E}[f(\tau') \mid \tau' > t_n] - \mathbb{E}[f(\tau) \mid \tau' > t_n] \\
= f(t_n + \phi^*/c) - \left( \frac{i - \phi^*/c - t_n}{t - t_{n+1}} f(t_{n+1}) + \frac{\phi^*/c + t_n - t_{n+1}}{t - t_{n+1}} f(i) \right) > 0
\]

by the concavity of \( f \) in the interval \([t_{n+1}, i]\). Because \( \tau \) and \( \tau' \) are identical up until time \( t_n \), we have \( \mathbb{E}[f(\tau')] > \mathbb{E}[f(\tau)] \), which contradicts the optimality of \( \tau \). Therefore, we must have \( t_{n+1} = i \), and this implies \( t_{n+1} = i = t_n + \phi/c \).

**Step 2: Indifferent Stage.** Consider any \( i \in \{0, \ldots, n - 1\} \). We will show that \( t_{i+1} - t_i = 1 \), where \( t_0 \) is treated as 0. Suppose a contradiction that \( t_{i+1} - t_i > 1 \). We consider the following three cases:

**Case 1:** \( t_i \geq i \). We will show that it is sufficient to ignore this case. Consider a stopping time \( \tau_1'' \)
generated by an increasing sequence \( t_1, \ldots, t_i \) with a pair of terminal times \((t_i + \phi^*/c, t_i + \phi^*/c)\). We can see that the DM is indifferent at time \( t_i \) for both information structures corresponding to stopping times \( \tau \) and \( \tau''_i \). Thus,

\[
\mathbb{E}[\tau \mid \tau > t_i] = \mathbb{E}[\tau'' \mid \tau > t_i] = t_i + \frac{\phi^*}{c}.
\]

Therefore, by Jensen’s inequality, we have

\[
\mathbb{E}[f(\tau) \mid \tau > t_i] \leq \mathbb{E}[\text{conv}_{t_i}(f)(\tau) \mid \tau > t_i] \leq \text{conv}_{t_i}(f)\left(t_i + \frac{\phi^*}{c}\right) = \mathbb{E}[f(\tau''_i) \mid \tau > t_i],
\]

where the equality follows from the fact that \( t_i \geq \tilde{i} \), which implies \( s^*(t_i) - t_i \leq \frac{\phi^*}{c} \). Because \( \tau \) and \( \tau''_i \) are identical up until time \( t_i \), we can conclude that \( \mathbb{E}[f(\tau)] \leq \mathbb{E}[f(\tau''_i)] \). Consider that all indifferent blocks generated by \( \tau'' \) start before time \( \tilde{i} \). This implies we only need to consider that \( t_i < \tilde{i} \) for every \( i \in \{0, \ldots, n - 1\} \).

**Case 2:** \( t_i < \tilde{i} < t_{i+1} \). By Step 1 and the previous case, we can modify the stopping time so that all consequent indifference stages start before time \( \tilde{i} \), and the terminal stage has a single terminal time. This implies the stage \([t_i, t_{i+1}]\) is the last indifferent stage followed by the terminal stage \([t_{i+1}, t_{i+1} + \phi^*/c]\). Note that \( t_{i+1} < t_i + \phi^*/c \). Therefore,

\[
\mathbb{E}[f(\tau) \mid \tau > t_i] = \frac{(t_{i+1} - t_i)c}{\phi^*} f(t_{i+1}) + \frac{\phi^* - (t_{i+1} - t_i)c}{\phi^*} f\left(t_{i+1} + \frac{\phi^*}{c}\right) 
\leq \frac{(t_{i+1} - t_i)c}{\phi^*} \text{conv}_{t_{i+1}}(f)(t_{i+1}) + \frac{\phi^* - (t_{i+1} - t_i)c}{\phi^*} \text{conv}_{t_{i+1}}(f)\left(t_{i+1} + \frac{\phi^*}{c}\right).
\]

Consider a stopping time \( \tau''_{2} \) generated by an increasing sequence \( t_1, \ldots, t_i \) with a pair of terminal times \((t_{i+1}, t_{i+1} + \phi^*/c - 1)\). This satisfies the conditions of a terminal stage because \( t_{i+1} + \phi^*/c - t_i > \phi^*/c \). This implies

\[
\mathbb{E}[f(\tau''_2) \mid \tau > t_i] = \frac{c(t_{i+1} - t_i - 1)}{\phi^* - c} f(t_{i+1}) + \frac{\phi^* - (t_{i+1} - t_i)c}{\phi^* - c} f\left(t_{i+1} + \frac{\phi^*}{c} - 1\right).
\]

From Lemma 12 (ii), we have \( s^*(t_{i+1}) - t_{i+1} < s^*(\tilde{i}) - \tilde{i} \leq \phi^*/c \), where the strict inequality follows by the fact that \( \tilde{i} < t^* \). Thus, \( s^*(t_{i+1}) \leq t_{i+1} + \phi^*/c - 1 \). Therefore,

\[
\mathbb{E}[f(\tau''_2) \mid \tau > t_i] = \frac{c(t_{i+1} - t_i - 1)}{\phi^* - c} \text{conv}_{t_{i+1}}(f)(t_{i+1}) + \frac{\phi^* - (t_{i+1} - t_i)c}{\phi^* - c} \text{conv}_{t_{i+1}}(f)\left(t_{i+1} + \frac{\phi^*}{c} - 1\right).
\]

By Jensen’s inequality, we have

\[
\text{conv}_{t_{i+1}}(f)\left(t_{i+1} + \frac{\phi^*}{c} - 1\right) \geq \frac{c}{\phi^*} \text{conv}_{t_{i+1}}(f)(t_{i+1}) + \frac{\phi^* - c}{\phi^*} \text{conv}_{t_{i+1}}(f)\left(t_{i+1} + \frac{\phi^*}{c}\right).
\]

The above equalities and inequality imply that \( \mathbb{E}[f(\tau''_2) \mid \tau > t_i] \geq \mathbb{E}[f(\tau) \mid \tau > t_i] \). However, this is contradiction because \( \tau''_2 \) has two different terminal times and \( t_{i+1} - t_i < \phi^*/c \) and Step 1 says it is impossible to have sender’s optimal stopping time induced by two different terminal times.
We will show that $\tau$ is generated by an increasing sequence $1, \ldots, s$; therefore, by Lemma 12 (iii), we have $E(\tau | \tau > t_1) = E(\tau' | \tau' > t_1)$.

Consider that
\[
0 = P(\tau'' > t_i) (E(\tau'' | \tau'' > t_i) - E(\tau | \tau > t_1))
\]
\[
= P(\tau'' = t_i + 1) \cdot (t_i + 1) + (P(\tau'' = t_i + 1) - P(\tau = t_i + 1)) \cdot t_{i+1}
\]
\[
+ (P(\tau'' > t_i + 1) - P(\tau > t_i + 1)) E(\tau | \tau > t_i + 1).
\]

Using the same argument shown in Proposition 9, we have $P(\tau'' > t_i + 1) > P(\tau > t_i + 1)$. We can rewrite the above equation as
\[
P(\tau = t_i + 1) \cdot t_{i+1} = P(\tau'' = t_i + 1) \cdot (t_i + 1) + P(\tau'' = t_i + 1) \cdot t_{i+1}
\]
\[
+ (P(\tau'' > t_i + 1) - P(\tau > t_i + 1)) E(\tau | \tau > t_i + 1).
\]

Consider any $t \in \{t_i + 1, \ldots, s^* (\tilde{t} - 1)\}$. Because $t_i + 1 \leq t_i + 1 - 1 \leq \tilde{t} - 1$, we have
\[
t \leq s^* (\tilde{t} - 1) \leq s^* (t_i + 1).
\]

Therefore, by Lemma 12 (iii), we have
\[
(t - t_i - 1) f(t_{i+1}) < (t - t_i + 1) f(t) + (t_{i+1} - t_i - 1) f(t),
\]
for every $t \in \{t_i + 1, \ldots, s^* (\tilde{t} - 1)\}$. Because the support of $\tau | \tau > t_i + 1$ is a subset of $\{t_i + 1, \ldots, s^* (\tilde{t} - 1)\}$, a linear combination of the above inequality for $t \in \{t_i + 1, \ldots, s^* (\tilde{t} - 1)\}$ implies
\[
P(\tau = t_i + 1) \cdot f(t_{i+1}) < P(\tau'' = t_i + 1) \cdot f(t_i + 1) + P(\tau'' = t_i + 1) \cdot f(t_{i+1})
\]
\[
+ (P(\tau'' > t_i + 1) - P(\tau > t_i + 1)) E(f(\tau) | \tau > t_{i+1}).
\]

This implies $E[f(\tau'') | \tau'' > t_i] > E[f(\tau) | \tau > t_i]$, which contradicts the optimality of $\tau$.

These two steps imply that we can choose a sender's optimal stopping time such that it is generated by an increasing sequence $1, \ldots, s$; with a terminal time $\tau_s + \phi/c$, where $\tau_s \leq \tilde{t}$.

We will show that $\tau_s \geq \tilde{t} - 1$. Suppose a contradiction that $\tau_s < \tilde{t} - 1$ Consider a stopping time $\tau'''$ generated by an increasing sequence $1, 2, \ldots, t_s + 1$ with a terminal time $t_s + \phi^* / c$. 

Case 3: $t_i < t_{i+1} \leq \tilde{t}$. The first two cases imply that it is sufficient to consider the case that all indifference blocks end weakly before time $\tilde{t}$.
Therefore,

\[
0 = \mathbb{E}[\tau'' | \tau > t_*] - \mathbb{E}[\tau | \tau > t_*] \\
= \frac{c}{\phi^*}(t_* + 1) + \frac{\phi^* - c}{\phi^*}(t_* + 1 + \frac{\phi^*}{c}) - (t^* + \frac{\phi^*}{c})
\]

Because \(t_* + 1 < \tilde{t}\), we have \(t_* + 1, t_* + 1 + \phi^*/c \in [t_* + 1, s^*(t_* + 1)]\). Therefore,

\[
\mathbb{E}[f(\tau) | \tau > t_*] = f\left(t_* + \frac{\phi^*}{c}\right) < \text{cconv}_{t_*+1}(f)\left(t_* + \frac{\phi^*}{c}\right)
\]

\[
= \frac{c}{\phi^*}\text{cconv}_{t_*+1}(f)(t_* + 1) + \frac{\phi^* - c}{\phi^*}\text{cconv}_{t_*+1}(f)\left(t_* + 1 + \frac{\phi^*}{c}\right)
\]

\[
= \frac{c}{\phi^*}f(t_* + 1) + \frac{\phi^* - c}{\phi^*}f\left(t_* + 1 + \frac{\phi^*}{c}\right)
\]

\[
= \mathbb{E}[f(\tau'') | \tau' > t_*].
\]

This implies \(\mathbb{E}[f(\tau)] < \mathbb{E}[f(\tau'')]\), which contradicts the optimality of \(\tau\). This concludes that the sender’s feasible stopping time is induced by an increasing sequence \(1, 2, \ldots, t_*\) with a terminal time \(t_* + \phi/c\) such that \(t_* \in \{\tilde{t} - 1, \tilde{t}\}\), as desired. \(\square\)
Online Appendix III: Convergence to $\Phi^*$ for $|\Theta| > 2$

Here we develop an analog of Proposition 5 for multiple states.

**Proposition 13.** Let $|\Theta| = n > 2$. Suppose that

$$
\sum_{t=0}^{\infty} \left( \log \mathbb{P}(\tau > t) - \log \mathbb{P}(\tau > t + 1) \right)^{n-1} = +\infty.
$$

Moreover, assume that $\{\mu \in \Delta(\Theta) : \phi(\mu) \geq \phi(\mu_0)\} \subset \text{int} \Delta(\Theta)$. Then, there exists $T \in \mathcal{T}$ such that $\mu_t \in \Phi^*$ for every $t > T$.

Additional assumptions that we need for this proposition is (i) tail of the probability measure of stopping time does not decay too fast (but still allows decay with a constant exponential rate) and (ii) the prior belief does not stay too close to the boundary of $\Delta(\Theta)$ so that the value of full information at $\mu_0$ is greater than that at any belief on the boundary of $\Delta(\Theta)$.

Before proving Proposition 13, we introduce a definition of a feasible set of $\mu_{t+1}$ given $\mu_t$ and stopping time $\tau$ derived by the boundary constraint.

**Definition 11.** For every nonnegative number $r \leq 1$ and belief $\mu \in \Delta(\Theta)$, define $F(\mu, r) \subset \Delta(\Theta)$ such that

$$
F(\mu, r) = \{\mu' \in \Delta(\Theta) \mid r \mu'(\theta) \leq \mu(\theta) \quad \forall \theta \in \Theta\}.
$$

With this definition, the boundary constraint is equivalent to $\mu_{t+1} \in F(\mu_t, \frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)})$ for every $t \in \mathcal{T}$. We now begin the proof of Proposition 13.

**Proof of Proposition 13.** Consider any feasible stopping time $\tau$.

Choose a maximal belief path and a distribution of stopping time $(\mu_t)_{t \in \mathcal{T}}$. From the proof of Proposition 4, we must have $\phi(\mu_t)$ is increasing in $t \in \mathcal{T}$ and the boundary constraint binds. Consider any $t_0 \in \mathcal{T}$. We will show that, for every $t \in \mathcal{T}$ and $\lambda \in [0, 1)$ such that $t > t_0 + 1$, we have $\lambda \mu_{t_0} + (1 - \lambda) \mu_t \notin \text{int} F(\mu_{t_0}, \frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)})$.

Assume a contradiction that there are $t_1 \in \mathcal{T}$ and $\lambda \in [0, 1)$ such that $t_1 > t_0 + 1$ and $\lambda \mu_{t_0} + (1 - \lambda) \mu_{t_1} \in \text{int} F(\mu_{t_0}, \frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)})$. This means $\frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)} < 1$, so the property of an extremal path implies $\mu_{t_0} \in \text{Bd} F\left(\mu_{t_0}, \frac{\mathbb{P}(\tau > t + 1)}{\mathbb{P}(\tau > t)}\right)$, so $\mu_{t_0} \neq \mu_{t_1}$.

We define a new belief path $(\mu'_t)_{t \in \mathcal{T}}$ as follows:

$$
\mu'_t = \begin{cases} 
\lambda \mu_t + (1 - \lambda) \mu_{t_1}, & \text{if } t_0 < t \leq t_1 \\
\mu_t, & \text{otherwise.}
\end{cases}
$$

We will show that a pair of a belief path and a distribution of stopping time $((\mu'_t)_{t \in \mathcal{T}}, d(\tau))$ is feasible. The obedience constraint is still the same for $t \notin \{t_0 + 1, \ldots, t_1\}$. If $t \in \{t_0 + 1, \ldots, t_1\}$,
we have

\[ \phi(\mu_t') \geq \lambda \phi(\mu_t) + (1 - \lambda) \phi(\mu_{t_1}) \geq \phi(\mu_t), \]

where the inequality follows by the fact that \( \phi(\mu_t) \) is increasing in \( t \in T \). This directly implies the obedience constraint for \( t \in \{t_0 + 1, \ldots, t_1\} \). The boundary constraint is still the same for \( t \in \{0, \ldots, t_0 - 1\} \cup \{t_1, \ldots\} \). The boundary constraint holds when \( t = t_0 \) by the construction of \( \lambda \). For any \( t \in \{t_0 + 1, \ldots, t_1 - 1\} \), we have

\[
\min_{\theta \in \Theta} \mu_t'(\theta) = \min_{\theta \in \Theta} \frac{\lambda \mu_t(\theta) + (1 - \lambda) \mu_{t_1}(\theta)}{\mu_{t_{+1}}(\theta)} \geq \min_{\theta \in \Theta} \frac{\mu_t(\theta)}{\mu_{t_{+1}}(\theta)} \geq \frac{P(\tau > t + 1)}{P(\tau > t)}.
\]

This concludes that the belief path \((\mu_t')_{t \in T}\) is also a maximal belief path corresponding to the stopping time \( \tau \) because \( \phi(\mu_t') \geq \phi(\mu_t) \) for every \( t \in T \). However, the above inequality is strict for \( t = t_0 \) because \( \mu_{t_0+1} \neq \mu_{t_0} \), which implies that \((\mu_t')_{t \in T}\) does not satisfy the property of an extremal path. This contradicts with the fact that \((\mu_t')_{t \in T}\) is a maximal belief path. Therefore, for every \( t_0, t_1 \in T \) such that \( t_1 > t_0 + 1 \), we must have \( \lambda \mu_{t_0+1} + (1 - \lambda) \mu_{t_1} \notin \text{int} F\left(\mu_{t_0}, \frac{P(\tau > t + 1)}{P(\tau > t)}\right) \).

For any \( \Theta' \subset \Theta \), define

\[ T_{\Theta'} = \left\{ t \in T \mid \Theta' = \left\{ \theta \in \Theta \mid \mu_{t+1}(\theta) = \frac{P(\tau > t)}{P(\tau > t + 1)} \mu_t(\theta) \right\} \right\}. \]

Note that \( T_\emptyset = \emptyset \) because the boundary constraint must be satisfied. Moreover, for every \( t \in T_\emptyset \), we must have \( P(\tau > t) = P(\tau > t + 1) \) and \( \mu_t = \mu_{t+1} \). Therefore,

\[
\sum_{\Theta' \subset \Theta} \sum_{t \in T_{\Theta'}} (\log P(\tau > t) - \log P(\tau > t + 1))^{n-1}
= \sum_{t = 0}^{\infty} (\log P(\tau > t) - \log P(\tau > t + 1))^{n-1} = +\infty.
\]

Because \( 2^{\Theta} \) is finite, there is a nonempty set \( \Theta' \subset \Theta \) such that \( \sum_{t \in T_{\Theta'}} (\log P(\tau > t) - \log P(\tau > t + 1))^{n-1} = +\infty \). Consider any \( t_0 < t_1 \in T_{\Theta'} \), we will show that there exists \( \theta \in \Theta' \) such that \( \mu_{t_1}(\theta) > \frac{P(\tau > t_0)}{P(\tau > t_0 + 1)} \mu_{t_0}(\theta) \). Suppose a contradiction that \( \mu_{t_1}(\theta) < \frac{P(\tau > t_0)}{P(\tau > t_0 + 1)} \mu_{t_0}(\theta) \) for every \( \theta \in \Theta' \).

Because \( \mu_{t_0+1}(\theta) < \frac{P(\tau > t_0)}{P(\tau > t_0 + 1)} \mu_{t_0}(\theta) \) for every \( \theta \notin \Theta' \) we can find a sufficiently small \( 1 - \lambda > 0 \) such that \( \lambda \mu_{t_0+1}(\theta) + (1 - \lambda) \mu_{t_1}(\theta) < \frac{P(\tau > t_0)}{P(\tau > t_0 + 1)} \mu_{t_0}(\theta) \) for every \( \theta \notin \Theta' \). For \( \theta \in \Theta \), we have

\[
\frac{P(\tau > t_0)}{P(\tau > t_0 + 1)} \mu_{t_0}(\theta) > \lambda \mu_{t_0+1}(\theta) + (1 - \lambda) \mu_{t_1}(\theta).
\]

This implies \( \lambda \mu_{t_0+1} + (1 - \lambda) \mu_{t_1} \notin \text{int} F\left(\mu_{t_0}, \frac{P(\tau > t + 1)}{P(\tau > t)}\right) \), which is a contradiction. Thus, there exists \( \theta \in \Theta' \) such that \( 1 \geq \mu_{t_1}(\theta) \geq \frac{P(\tau > t_0)}{P(\tau > t_0 + 1)} \mu_{t_0}(\theta) \). This implies

\[
\bigcup_{\theta \in \Theta'} \left( \log \mu_{t_0}(\theta), \log \left( \frac{P(\tau > t_0)}{P(\tau > t_0 + 1)} \mu_{t_0}(\theta) \right) \right) \cap \bigcup_{\theta \in \Theta'} \left( \log \mu_{t_1}(\theta), \log \left( \frac{P(\tau > t_1)}{P(\tau > t_1 + 1)} \mu_{t_1}(\theta) \right) \right) = \emptyset.
\]
for every $t_0 < t_1 \in \mathcal{T}_{\Theta'}$.

Because $\{ \mu \in \Delta(\Theta) \mid \phi(\mu) \geq \phi(\mu_0) \} \subset \text{int} \Delta(\Theta)$ and $\phi(\mu_t)$ is increasing in $t$, we have $\mu_t \in \{ \mu \in \Delta(\Theta) \mid \phi(\mu) \geq \phi(\mu_0) \} \subset \text{int} \Delta(\Theta)$. Therefore, for each $\theta \in \Theta$, there is $\mu_{\theta} > 0$ such that $\mu_t(\theta) \geq \mu_{\theta}$ for every $t \in \mathcal{T}$. This implies that for every $t \in \mathcal{T}_{\Theta'}$ we have

$$\bigcap_{\theta \in \Theta'} \left( \log \mu_t(\theta), \log \left( \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t + 1) \mu_t(\theta)} \right) \right) \subset \bigcap_{\theta \in \Theta'} [\log \mu_{\theta}, 0].$$

This implies

$$\bigcup_{t \in \mathcal{T}_{\Theta'} \bigcap_{\theta \in \Theta'} \left( \log \mu_t(\theta), \log \left( \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t + 1) \mu_t(\theta)} \right) \right) \subset \bigcap_{\theta \in \Theta'} [\log \mu_{\theta}, 0].$$

We showed that two different boxes inside the union have disjoint interiors. This implies

$$\text{Vol} \left( \bigcap_{\theta \in \Theta'} [\log \mu_{\theta}, 0] \right) \geq \sum_{t \in \mathcal{T}_{\Theta'}} \text{Vol} \left( \bigcap_{\theta \in \Theta'} \left( \log \mu_t(\theta), \log \left( \frac{\mathbb{P}(\tau > t)}{\mathbb{P}(\tau > t + 1) \mu_t(\theta)} \right) \right) \right)$$

$$= \sum_{t \in \mathcal{T}_{\Theta'}} (\log \mathbb{P}(\tau > t) - \log \mathbb{P}(\tau > t + 1))^{[\Theta']_1}$$

$$= \infty,$$

where the last equality follows from the fact that $|\Theta'| \leq n - 1$. This is a contradiction because the LHS is finite. Therefore, a maximal belief path must stay at $\Phi^*$ for large enough $t \in \mathcal{T}$, as desired. \qed
Online Appendix IV: Details of smoothing out $f$ and $c$

In this appendix we explicitly define a sequence of models as follows.

**Definition 12.** The model indexed by $k \in \mathbb{N}_{>0}$ has times $\mathcal{T}_k = \left\{ \frac{1}{k}, \frac{2}{k}, \ldots \right\}$. Furthermore, define

(i) $c_k : \mathcal{T}_k \to \mathbb{R}$ such that for all $t \in \mathcal{T}$, (i) $c_k(t) = c'(t)$; and (ii) $c_k$ is linear between $t$ and $t + 1$.

(ii) $f_k : \mathcal{T}_k \to \mathbb{R}$ such that for all $t \in \mathcal{T}$, (i) $f_k(t) = f'(t)$; and (ii) $f_k$ is linear between $t$ and $t + 1$.

This simply ‘fills in’ the original functions through linear interpolation. In particular, $c_k (f_k)$ and $c' (f')$ agree on $\mathcal{T}$. This is illustrated in Figure 14 below.

![Figure 14: Illustration of $f_{k=3}$](image)

Now define the deterministic time change $s_k : \mathcal{T}_k \to \mathcal{T}$ such that $s(t) = t \cdot k$ and define $f$, $c$ such that for all $t \in \mathcal{T}_k$,

$$f(s_k(t)) = f_k(t) \quad \text{and} \quad c(s_k(t)) = c_k(t),$$

noting that the domain of $f$ and $c$ is $\mathcal{T}$. Finally, observe that by taking $k$ large, $|f(t + 1) - f(t)|$ and $|c(t + 1) - c(t)|$ can be made arbitrarily small.
Online Appendix V: Robustness to multiple senders

In this appendix, we outline a simple setting with multiple (possibly competing) senders and show that equilibrium distributions of stopping times must remain within $\mathcal{D}$ which we characterize in the main text.

Index the set of information providers with $\mathcal{P} := \{1, 2, \ldots, n\}$ and suppose that each provider $i \in \mathcal{P}$ has some value function $v_i : A \times T \to \mathbb{R}$ i.e., they care jointly about what action the DM ends up taking (“persuasion”) and how long the DM pays attention for (“attention”). Information provider $i$ has finite message space $M_i$ and denote $\mathcal{H}_i^t = \prod^t M_i$ as the space of possible time $t$ histories for provider $i$. Suppose that all information providers have (costless) access to the full set of dynamic information structures defined on their respective message spaces i.e., each provider $i \in \mathcal{P}$ chooses $I_i \in \mathcal{I}(M_i)$ where $\mathcal{I}(M_i)$ is the set of all dynamic information structures defined on $M_i$.

At each period $t \in T$ the DM observes the vector $(m_i^n)_{i=1}^n \in \prod_{i=1}^n M_i$ generated by $(I_i)_{i \in \mathcal{P}} \in \prod_{i=1}^n \mathcal{I}(M_i)$, the profile of dynamic information structures each provider chooses. The DM then decides whether to stop and take action, or continue paying attention for another period.

An equilibrium of this game is a profile of dynamic information structures $(I_i)_{i \in \mathcal{P}} \in \prod_{i=1}^n \mathcal{I}(M_i)$ such that for each provider $i \in \mathcal{P}$ and all $I_0_i \in \mathcal{I}(M_i)$,

$$E^{(I_i \cdot I_0_i)}[v_i(a, \tau)] \geq E^{(I_i \cdot I_0_i)}[v_i(a, \tau)]$$

where we denote $L_{-i} := (I_j)_{j \neq i}$ and the expectation operator is taken with respect to the all providers’ information structures under the assumption that for every time $t \in T$ the DM observes all providers’ time $t$ messages.

Now notice that this is simply another dynamic information structure for which messages are vector valued, and the history dependence of the $i$-th message at time $t$ only enters through the path of $i$-th messages at times $s < t$. More formally, define $M := \prod_{i \in \mathcal{P}} M_i$ and $\mathcal{H}_t = \prod_{i=1}^t \left( \prod_{i=1}^n M_i \right)$. Then construct a single dynamic information structure $I \in \mathcal{I}(M)$ by choosing

$$p^I((m_i^n)_{i=1}^n | \theta, H_t \in \mathcal{H}_t) = \prod_{i=1}^n p_i^I(m_i^n | \theta, H_t \in \mathcal{H}_t^i).$$

for all $\theta \in \Theta$. This replicates the the distribution of each path of messages the DM received under $(I_1, \ldots, I_n)$. As such, if an equilibrium exists in this game, the corresponding equilibrium distribution of stopping times must remain within $\mathcal{D}$. In fact, observe that because each provider can only condition their information on the realization of their own messages realizations (i.e., the distribution of $m_i^t$ can only depend on $H_{t-1}^i \in \mathcal{H}_{t-1}^i$), the set of distributions of stopping times might be strictly smaller than $\mathcal{D}$.

There are, of course, richer or more realistic settings to which our characterization result applies. Our goal is not to offer an exhaustive account, but to highlight that set or feasible stopping times we study holds more broadly than the case with a single designer with full
commitment. For instance, we might imagine that at time $t$, the DM receives message $m^X_t$ where $X_t \in P$ is the provider whose message the DM observes at time $t$. $X_t$ might be chosen by a separate player, and can depend on the history of all past message realizations from each provider. This might correspond to the case in which a news feed aggregator (e.g., Facebook) can choose the information provider (news outlet) whose message the DM observes, but not the message itself. It should be evident that the distributions of equilibrium stopping times remain within $\mathcal{D}$. 
Online Appendix VI: Supporting calculations for Example 1

In this Appendix we supply the supporting calculations omitted in Example 1.

VI.1 Calculation verifying the stopping nodes on the original structure $I$. Observe that for any $c > 0$, the DM stops after the histories $(m_1 = 1, m_2 = 1)$ and $(m_1 = 0, m_2 = 0)$ since on those histories, she becomes so convinced that the state is either 1 or 0 that no matter what signal she receives in $t = 3$, she will take the same action and no information is provided after $t = 4$. On the other hand, DM continues after the histories $(m_1 = 1, m_2 = 0)$ and $(m_1 = 0, m_2 = 1)$ since her belief after those histories is 1/2 and information provided at $t = 3$ remains valuable. It is then easy to see that there must exist $c > 0$ such that for all $c > c$, the DM optimally stops at the nodes depicted in Figure 1(a) in the main text. (Indeed, the reader can verify that $c = 2/39$.)

VI.2 Calculation that full information can be chosen to replicate non-stopping beliefs. We illustrate this for the history $(m_1 = 1)$ by setting

\[
\begin{align*}
\mathbb{P}(m_2 = 1 | \theta = 1) &= p_1 \\
\mathbb{P}(m_2 = 0 | \theta = 1) &= 1 - p_1 \\
\mathbb{P}(m_2 = 0 | \theta = 0) &= p_0 \\
\mathbb{P}(m_2 = 0 | \theta = 0) &= 1 - p_0.
\end{align*}
\]

We want to set the posterior belief conditional on not learning about the state perfectly to 1/2, the same belief as the corresponding node in $I$:

\[
\mathbb{P}(\theta = 1 | m_2 = 0, m_1 = 1) = \frac{(1 - p_1) \cdot (2/3)}{(2/3)(1 - p_1) + (1/3)(1 - p_0)} = 1/2
\]

on the other hand, we want to ensure that on history $(m_1 = 1)$, DM learns the state fully with probability 5/9 to preserve incentives to continue paying attention:

\[
2p_1/3 + p_0/3 = 5/9
\]

and we have two equations and two unknowns to pin down the information structure after history $(m_1 = 1)$.

VI.3 Calculation that the DM prefers to continue paying attention upon receipt of the null message. Let $V$ denote the DM’s continuation value from paying attention after histories $\{m_1 = 1\}$ or $\{m_1 = 0\}$ on information structure $I'$:

\[
V = (\underbrace{5/9 \cdot (1 - c)}_{\text{learn state perfectly at } t = 2}) + (\underbrace{4/9 \cdot (1 - 2c)}_{\text{learn state perfectly at } t = 3}),
\]

noting here that because the modification from $I'$ to $I''$ preserves incentives from time $t = 2$ onward, this also corresponds to the continuation value after history $\{m_1 = 0\}$ on the infor-
mation structure $I''$. Now let $u^*(\mu) := \max_{a \in [0,1]} \mathbb{E}_{\mu}[-(a - \theta)^2]$ be the expected payoff from stopping and taking the optimal action under belief $\mu$. We have

$$u^*(1/2) = u^* \left( (2/3) \cdot (1/2) + (1/3) \cdot (1/2) \right)$$
$$\leq (1/2) \left[ u^*(2/3) + u^*(1/3) \right]$$
$$\leq V/2 + V/2 = V$$

where the first inequality is from the convexity of $u^*$ and the second inequality is because the DM did indeed find it optimal to continue paying attention after histories $(m_1 = 1)$ and $(m_1 = 0)$ on information structure $I'$. We have thus verified that under the modification $I''$, the DM’s incentive to continue paying attention at non-terminal nodes must once again weakly increase.

**VI.4 Calculation verifying properties of $I_1$ and $I_2$. Obedience.** We argued in the main text that the DM facing $I_1$ is indifferent at $t = 0$. This implies she must strictly prefer to pay attention at $t = 1$.

For $I_2$, observe that the DM’s value of information is $1/2$. We conjecture that it is optimal for her to continue paying attention whenever she does not learn the state and verify that this is in fact optimal. Under the conjectured stopping rule, the DM is only uncertain as to when she learns the state. At $t = 0$, we have $1/2 = c \cdot \mathbb{P}(\tau = 1) + 3c\mathbb{P}(\tau > 1)$ since $c = 1/4$. At $t = 1$ and $t = 2$, the argument is identical to that of $I_1$, but conditioning on the event that $\{\tau > 1\}$.

**Optimality.** Apply Proposition 10.