informs
https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 72, No. 1, January–February 2024, pp. 151–166
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Crosscutting Areas

Persuading Risk-Conscious Agents: A Geometric Approach

Jerry Anunrojwong,[a] Krishnamurthy Iyer,[b,*] David Lingenbrink[c]

[a] Graduate School of Business, Columbia Business School, New York, New York 10027; [b] Industrial and Systems Engineering, University of Minnesota, Minneapolis, Minnesota 55455; [c] School of Operations Research and Information Engineering, Cornell University, Ithaca, New York 14853

*Corresponding author

Contact: jerryanunroj@gmail.com, https://orcid.org/0000-0001-8422-9539 (JA); kriyer@umn.edu, https://orcid.org/0000-0002-5538-1432 (KI); dal299@cornell.edu, https://orcid.org/0000-0001-6214-9941 (DL)

Received: May 10, 2020
Revised: March 12, 2021; August 7, 2022
Accepted: November 23, 2022
Published Online in Articles in Advance: March 28, 2023

Area of Review: Market Analytics and Revenue Management

https://doi.org/10.1287/opre.2023.2438

Abstract. We consider a persuasion problem between a sender and a receiver where utility may be nonlinear in the latter's belief; we call such receivers *risk conscious*. Such utility models arise when the receiver exhibits systematic biases away from expected utility maximization, such as uncertainty aversion (e.g., from sensitivity to the variance of the waiting time for a service). Because of this nonlinearity, the standard approach to finding the optimal persuasion mechanism using revelation principle fails. To overcome this difficulty, we use the underlying geometry of the problem to develop a convex optimization framework to find the optimal persuasion mechanism. We define the notion of *full persuasion* and use our framework to characterize conditions under which full persuasion can be achieved. We use our approach to study *binary persuasion*, where the receiver has two actions and the sender strictly prefers one of them at every state. Under a convexity assumption, we show that the binary persuasion problem reduces to a linear program and establish a canonical set of signals where each signal either reveals the state or induces in the receiver uncertainty between two states. Finally, we discuss the broader applicability of our methods to more general contexts, and we illustrate our methodology by studying information sharing of waiting times in service systems.

Funding: The second and the third authors gratefully acknowledge support from the Division of Civil, Mechanical and Manufacturing Innovation of the National Science Foundation [Grants CMMI-1633920 and CMMI-2002156].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2023.2438.

Keywords: Bayesian persuasion • information design • risk consciousness

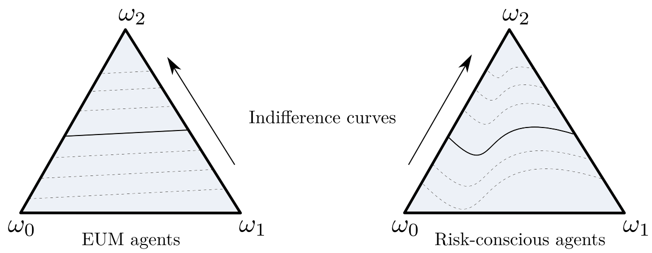
1. Introduction

Given the inherent informational asymmetries in online marketplaces between a platform and its users, information design has an important role to play in their design and operation. Building on the methodological contributions of Rayo and Segal (2010), Kamenica and Gentzkow (2011), and Bergemann and Morris (2016), information design has been applied in a number of different application contexts, such as engagement/misinformation in social networks (Candogan and Drakopoulos 2020), service systems (Lingenbrink and Iyer 2019), and online retail (Lingenbrink and Iyer 2018, Drakopoulos et al. 2021).

In much of the previous work, the standard assumption is that the agent being persuaded (the *receiver*) is an expected utility maximizer (EUM). Although this assumption is well supported theoretically via axiomatic characterizations (Savage 1954), it is empirically well documented that human behavior is inadequately explained by the central tenets of the theory (Ellsberg 1961, Allais 1979, Rabin 1998, DellaVigna 2009). In particular, there is a long line of work in economics studying the systematic biases in human behavior leading to deviations from expected utility maximization (Kahneman and Tversky 1972, 1979; Machina 1982; Tversky and Kahneman 1992). Because of these shortcomings, existing models of Bayesian persuasion may not satisfactorily apply to information design problems in online markets and other practical settings.

Motivated by this concern, our main goal in this paper is to extend the methodology of Bayesian persuasion to settings where the receiver may not be an expected utility maximizer. In our general model of utility under uncertainty, given a finite state space Ω and a set of actions A, we take as model primitive the utility function $\rho(\mu, a)$ that specifies the receiver's utility for action $a \in A$ under belief μ. For EUM receivers, this function is given by $\rho(\mu, a) \triangleq \sum_{\omega \in \Omega} \mu(\omega) u(\omega, a)$ for some function u, and thus ρ is *linear* in the belief. Our framework relaxes this linearity assumption and allows for the utility function ρ to depend nonlinearly on the belief. We refer to such general receivers as being *risk conscious*. See Figure 1 for a graphical comparison.

Figure 1. (Color online) Indifference Curves in the Belief Space



Notes. Here, the triangle denotes the simplex of beliefs over $\Omega = \{\omega_0, \omega_1, \omega_2\}$, and the lines/curves denote the beliefs with a fixed utility. The indifference curves for EUM receivers are necessarily hyperplanes, whereas they need not be so for risk-conscious receivers.

1.1. Motivations for Risk Consciousness

Although the standard model of expected utility maximization is subsumed in the risk-conscious framework, the latter allows for far more generality. In what follows, we provide three motivations for studying risk-conscious behavior.

1.1.1. Deviations from EUM. The literal interpretation of risk consciousness is that the agent's behavior deviates systematically from expected utility maximization. Such deviations are well documented in the literature; we provide a few examples in Table 1 and discuss them in Online Appendix F. For instance, an agent with uncertainty aversion may be modeled using mean-standard deviation utility. Thus, there is a need to develop a theory for the persuasion of such agents, both to better capture realistic behavior in operational settings and to gain qualitative insights into phenomena that do not arise with EUM receivers.

1.1.2. Dynamic Decision Making. A second source of risk-conscious behavior comes from dynamic settings where agents must make multiple decisions over time, with past decisions influencing the information available to the agents in the future. For any fixed action in the present, a change in the agent's belief may alter his or her actions in the future, thereby affecting future (expected) payoffs. The net effect on the agent's utility of a change in belief may then be nonlinear, thus leading to risk consciousness. (We provide details in Section 6.) Although such models arise in a number of operational settings, a concrete example that has received recent attention is the market for data and information products (Bergemann and Bonatti 2015, 2019; Bergemann et al. 2018; Zheng and Chen 2021). A buyer in such a market often uses the data to make subsequent decisions; thus, the buyer's utility for the data may depend nonlinearly in his or her belief about the data quality. A seller who wishes to persuade such a buyer into purchasing the data then faces the problem of risk-conscious persuasion.

Table 1. Examples of Risk-Conscious Utility

Risk-conscious utility	Utility representation $\rho(\mu, a)$
Expected utility	$\mathbf{E}_\mu[u(\omega, a)]$
Maximin utility	$\min_\theta \mathbf{E}_\mu[u(\omega, a; \theta)]$
Mean-standard deviation utility	$\mathbf{E}_\mu[u(\omega, a)] - \beta\sqrt{\mathrm{Var}_\mu(g(\omega, a))}$
Value-at-risk	$-\min\{t \in \mathbb{R} : \mathbf{P}_\mu(\ell(\omega, a) > t) \le 1 - \alpha\}$
Conditional value-at-risk	$-\mathbf{E}_\mu[\ell(\omega, a) \mid \ell(\omega, a) > \tau]$
Cumulative prospect theory	$\sum_\omega f_\omega(\mu) u(\omega, a)$

1.1.3. Modeling Device. Finally, risk consciousness can be a useful modeling tool to analyze more complex settings. In particular, using the risk-conscious framework, one can analyze instances of *public persuasion*, where a sender seeks to persuade a group of receivers by sending a common public signal; such instances are common in markets, public services, platforms, and social networks, where practical concerns such as information leakage and fairness preclude private persuasion (Candogan 2019, Yang et al. 2019, Anunrojwong et al. 2020, Candogan and Drakopoulos 2020). Similarly, the framework can be applied to study *robust persuasion*, where a receiver with a private type is persuaded by a sender taking a worst-case view. Robust persuasion is important in market and platform services, where participants often have idiosyncratic components to their payoffs that are unknown to the platform (Bimpikis and Papanastasiou 2019, Candogan 2020).

1.2. Challenges and Opportunities

From a theoretical perspective, optimal persuasion of risk-conscious agents presents new analytical challenges. When agents are expected utility maximizers, a revelation principle–style argument is often invoked to reduce the set of possible messages the sender might send (i.e., *signals*) to the set of actions available to the receiver. By pairing each signal directly with the action taken by the receiver, this reduction simplifies the persuasion problem substantially, and the resulting optimization problem can be written as a linear program with one *obedience* constraint for each action. By contrast, with risk-conscious agents, because of the nonlinearity of the receiver's utility, a key step of the revelation principle argument fails (as we describe in Section 3), rendering this approach to finding an optimal *signaling scheme* ineffective. The lack of a convenient set of signals, where each signal is paired with an optimal action, presents analytical and computational challenges to persuasion.

Despite these challenges, the setting with risk-conscious agents also yields new possibilities for persuasion. We illustrate this aspect with the following example.

Example 1. Consider a sender who seeks to persuade a receiver into taking an action. Let there be four payoff-relevant states, $\omega \in \Omega = \{0,1,2,3\}$. The sender and the receiver's prior belief is given by $\mu^* = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. For any belief $\mu = (\mu_0, \mu_1, \mu_2, \mu_3)$, the receiver's utility for taking the action equals $\hat{\rho}(\mu) \triangleq \max_{\omega \in \{0,1,2\}} \left(\mu_\omega - \frac{2}{3}\right)$; the receiver takes the action if and only if the utility is nonnegative.

Because $\hat{\rho}(\mu^*) < 0$, without persuasion, the receiver will not take the action. As $\hat{\rho}(e_\omega) = \frac{1}{3} > 0$ for $\omega \in \{0,1,2\}$ and $\hat{\rho}(e_4) = -\frac{2}{3} < 0$, under the full-information scheme that reveals the state, the receiver takes the action only in states $\omega \in \{0,1,2\}$. (Here, e_ω denotes the belief that puts all its weight on $\omega \in \Omega$.)

Now, consider the following signaling scheme that uses three signals $\{s_0, s_1, s_2\}$: when the state is $\omega = i \in \{0,1,2\}$, the scheme sends the signal s_i, and when the state is $\omega = 3$, the scheme sends each signal with equal probability. One can verify that upon receiving the signal s_i the receiver's posterior belief equals $\mu^i = \frac{3}{4}e_i + \frac{1}{4}e_4$, with $\hat{\rho}(\mu^i) = \frac{1}{12} > 0$. This implies that, regardless of the signal received, the receiver finds it optimal to take the action. Thus, the preceding signaling scheme is optimal and *fully* persuades the receiver into taking the action.

Note that the optimal scheme requires three signals, which is strictly more than the number of choices available to the receiver. (It can be verified that under any signaling scheme that uses two signals, the probability of the receiver taking the action is at most $\frac{3}{8}$.) This is in contrast with expected utility–maximizing receivers, where the revelation principle states that action recommendations (and hence two signals) suffice for optimal persuasion. □

In this example, with optimal persuasion, the receiver takes the sender's most preferred action at each state. We refer to this outcome as *full persuasion,* and we study it in detail in Section 4. Observe that here, full persuasion occurs despite the fact that under his or her prior, the receiver would choose otherwise. This is due to the nonlinearity inherent in the receiver's utility, and in Theorem 2 we show that this cannot occur when the receiver is an expected utility maximizer.

1.3. Main Contributions

The main contributions of our paper are as follows: (1) we provide a convex optimization framework to overcome the analytical challenges in the persuasion of risk-conscious agents; (2) using this framework, we identify conditions under which persuasion is beneficial to the sender, and through the notion of full persuasion, we provide insights into the extent of this benefit; and (3) we demonstrate the structural properties of the resulting optimal persuasion mechanisms, by examining settings that impose additional regularity assumptions on the receiver's utility.

The convex programming framework we develop in Section 3 uses the underlying geometry of the persuasion problem to optimize directly over the joint distribution of the state and the receiver's actions. These variables are shown to lie in the convex hull of the set of beliefs for which a fixed action is optimal for the receiver. The optimal persuasion mechanism is then obtained through a convex decomposition of the optimal solution to the convex program. Although related, our approach is different from the concavification approach (Kamenica and Gentzkow 2011), yielding a different convex program; we discuss the connection in detail in Online Appendix A.

Using this framework, in Section 4, we first characterize conditions under which the sender strictly benefits from the persuasion of a risk-conscious receiver. This result is analogous to, and extends, similar characterization for expected utility maximizers (Kamenica and Gentzkow 2011). Specific to the risk-conscious framework, we formally define the notion of *full persuasion* as a measure of the extent of the benefits to persuasion and provide necessary and sufficient conditions under which full persuasion is achievable.

To obtain more insight into the structure of the optimal persuasion mechanism, we then study a specialized setting in Section 5-namely, *binary persuasion*-in which the receiver has two actions (0 and 1), and the sender always prefers that the receiver choose action 1 over action 0. Under a convexity assumption on the receiver's utility function, we show that the convex program, in fact, reduces to a linear program whose solution can be efficiently computed. By analyzing this linear program, we establish a *canonical set* of signals for optimal persuasion. In other words, we show there exists an optimal signaling scheme that always sends signals in this canonical set for any prior belief of the receiver. This canonical set of signals consists of *pure* signals, which fully reveal the state to the receiver, and *binary mixed* signals, which induce uncertainty between two states. With additional monotonicity assumptions on the utility function, we show that the optimal signaling scheme induces a threshold structure in the receiver's action.

In summary, our work provides a methodology to solve for the optimal persuasion mechanism with risk-conscious agents and demonstrates the fruitfulness of analyzing more realistic models of human behavior. In Section 6, we provide a brief discussion of the use of our approach in more general settings, beyond the persuasion of a risk-conscious receiver. Finally, as an illustration of our methodology, in Section 7 we analyze a model of a queueing system where the service provider seeks to persuade arriving risk-conscious customers to join an unobservable single-server queue, and we establish the intricate "sandwich" structure of the optimal signaling scheme.

1.4. Literature Review

Our work contributes to the literature on Bayesian persuasion (Rayo and Segal 2010; Kamenica and Gentzkow 2011; Bergemann and Morris 2016, 2019; Dughmi and Xu 2016; Kolotilin et al. 2017; Taneva 2019), where a sender *commits* to a mechanism of sharing payoff-relevant information with a receiver in order to influence the latter's actions. For a recent review of the literature, see Kamenica (2019). Our work particularly takes influence from Kamenica and Gentzkow (2011), who use a convex-analytic *concavification* approach to study the persuasion problem; we discuss the close relation to our work in Online Appendix A.

In the operations research literature, a number of authors have applied the methodology of Bayesian persuasion to study varied settings such as crowdsourced exploration (Papanastasiou et al. 2018), spatial resource competitions (Yang et al. 2019), engagement-misinformation trade-offs in online social networks (Candogan 2019, Candogan and Drakopoulos 2020), warning policies for disaster mitigation (Alizamir et al. 2020), throughput maximization in queues (Lingenbrink and Iyer 2019), inventory/demand signaling in retail (Lingenbrink and Iyer 2018, Drakopoulos et al. 2021), and quality of matches in matching markets (Romanyuk and Smolin 2019). Our work is inspired by this stream of work and seeks to broaden the domain of applicability by incorporating more general utility models for the receiver.

Our other source of inspiration is the economics literature on the theory of individual preferences toward risk. Apart from the standard expected utility hypothesis, there have been a number of theoretical frameworks proposed to model preferences under uncertainty, including the widely studied prospect theory (Kahneman and Tversky 1979) and the cumulative prospect theory (Tversky and Kahneman 1992). Machina (1982, p. 278; 1995, p. 15) singles out the "independence axiom" in the expected utility framework, which leads to the utilities being "linear in the probabilities" and study models that relax the independence axiom. Our notion of risk consciousness, on the one hand, encompasses all of these nonexpected utility models, as it only requires the utility to be continuous in beliefs. On the other hand, our notion does not capture some nonexpected utility models, such as the maxmin expected utility with multiple priors (Gilboa and Schmeidler 1989).

The notion of risk consciousness is related to the concept of *risk measures* in mathematical finance (Artzner et al. 2001, Föllmer and Schied 2016). Commonly studied risk measures, such as the variance of the portfolio return (Markowitz 1952), the value-at-risk (Jorion 2006), the expected shortfall (Acerbi and Tasche 2002), and the entropic value-at-risk (Ahmadi-Javid 2012), are all nonlinear functions of the distribution of the return. We note that to be a good measure of financial risk, a risk measure needs to satisfy a number of properties (e.g., *coherence*; Artzner et al. 2001) that make sense in the context of portfolio management but may not be relevant for capturing aspects of human decision making.

We end our discussion by mentioning two closely related recent works. Beauchêne et al. (2019) study a persuasion setting where a sender uses an *ambiguous* communication device comprising multiple ways of sending signals. The receiver is ambiguity averse and has the maxmin expected utility (Gilboa and Schmeidler 1989) over the multiple resulting posteriors. As mentioned earlier, such maxmin-utility preferences are distinct from the class of risk-conscious receivers considered in this paper. The authors analyze optimal persuasion and, parallel to our result, note that full persuasion can be achieved through the use of an ambiguous communication device. By contrast, our work demonstrates that full persuasion can be obtained from unambiguous communication, purely as a result of risk-conscious preferences.

Lipnowski and Mathevet (2018) consider persuasion of a receiver with similar nonlinear preferences as in this paper but focus on the setting where the sender's preferences are perfectly aligned with those of the receiver (for instance, the sender might be a trusted advisor). Noting the failure of the revelation principle, the authors provide a sufficient condition-namely, that the receiver's payoffs are concave in the belief for any fixed action, under which action recommendations are optimal. In our intended applications, the sender is a platform or a marketplace, and hence we model the sender as an expected utility maximizer. Because of this, apart from trivial instances of our model, the sender's preferences will not be aligned with those of the receiver.

2. Model

In the following, we present the model of Bayesian persuasion with risk-conscious agents. Our development of the model follows closely to that of the standard Bayesian persuasion setting (Kamenica and Gentzkow 2011, Kamenica 2019).

2.1. Setup

We consider a persuasion problem with one *sender* and one *receiver*. Let X be a payoff-relevant random variable with support on a known set $\mathcal{X}$. We assume that neither the receiver nor the sender observes X. However, as we describe in the following, the sender has more information about X than the receiver and seeks to use this information to influence the receiver's actions.

Formally, we assume that the distribution of X depends on the *state* of the world $\bar{\omega}$, which takes values in a finite set Ω and is observed by the sender but not the receiver. We denote the distribution of X, conditional on $\bar{\omega} = \omega$, by F_ω. The distributions $\{F_\omega : \omega \in \Omega\}$ are commonly known between the sender and the receiver, and both share a common prior $\mu^* \in \Delta(\Omega)$ about the state of the world $\bar{\omega}$. (Throughout, for any set S, we let $\Delta(S)$ denote the set of probability measures over S. When S is finite, we consider $\Delta(S)$ a subset of $\mathbb{R}^{|S|}$, endowed with the Euclidean topology.) For each $\mu \in \Delta(\Omega)$, we let F_μ be the distribution of X when $\bar{\omega}$ is distributed as μ: we have $F_\mu = \sum_\omega \mu(\omega) F_\omega$. Finally, we let X_ω denote an independent random variable distributed as F_ω.

As in the standard Bayesian persuasion setting, we assume that the receiver is Bayesian and that the sender can commit to a *signaling scheme* to influence receiver's choice of an action (which is described later in detail). A signaling scheme (S, π) consists of a signal space S and a joint distribution $\pi \in \Delta(\Omega \times S)$ such that the marginal of π over Ω equals μ^*: for each $\omega \in \Omega$, $\pi(\omega, S) = \mu^*(\omega)$. Specifically, under the signaling scheme (S, π), if the realized state is $\bar{\omega} = \omega$, the sender draws a signal $\bar{s} \in S$ according to the conditional distribution $\pi(\cdot | \bar{\omega} = \omega)$ and conveys it to the receiver. For simplicity of notation, we denote a signaling scheme (S, π) by the joint distribution π. Throughout, we assume that the sender commits to a signaling scheme prior to observing the state $\bar{\omega}$ and that the sender's choice of the signaling scheme π is common knowledge between the sender and the receiver.

As mentioned previously, we assume that the receiver is Bayesian. Given the sender's signaling scheme π, upon observing the signal $\bar{s} = s$, the receiver uses Bayes' rule to update his or her belief from the prior μ^* to the posterior $\mu_s \in \Delta(\Omega)$. In particular, we have for all $\omega \in \Omega$

$$\mu_s(\omega) = \frac{\pi(\omega, s)}{\sum_{\omega' \in \Omega} \pi(\omega', s)}$$

whenever the denominator on the right-hand side is positive. (We let $\mu_s \in \Delta(\Omega)$ be arbitrary if the denominator is 0.) This implies that upon receiving the signal $\bar{s} = s$, the receiver believes that the payoff-relevant variable X is distributed as F_{μ_s}.

2.2. Actions, Strategy, and Utility

Upon observing the signal $\bar{s}$, the receiver chooses an action a from a finite set A of actions. Given a signaling scheme π, the receiver's strategy $a(\cdot)$ specifies an action $a(s) \in A$ for each realization $s \in S$ of the signal $\bar{s}$. (Although our definition implies a pure strategy, we can easily incorporate mixed strategies where the receiver chooses an action at random. We suppress this technicality for the sake of readability.)

We let $v(\omega, a)$ denote the sender's utility in state $\bar{\omega} = \omega$ when the receiver chooses the action $a \in A$. Furthermore, we assume that the sender is an expected utility maximizer. (One can equivalently represent the utility function $v(\omega, a)$ as an expectation of a utility function $\hat{v}(X, a)$ over the payoff-relevant variable X and the action a, conditional on $\bar{\omega} = \omega$; we suppress the details for brevity.)

Our point of departure from the standard persuasion framework is in the definition of the receiver's utility. Specifically, we relax the assumption that the receiver is an expected utility maximizer; as we describe next, our setup allows for more general models of the receiver's utility over the uncertain outcome X. We refer to such receivers as being *risk conscious*.

Formally, for any belief $\mu \in \Delta$ of the receiver, we assume that the receiver's *utility* upon taking an action $a \in A$ is given by $\breve{\rho}(F_\mu, a) \in \mathbb{R}$. For notational simplicity, we define the *utility function* $\rho : \Delta(\Omega) \times A \to \mathbb{R}$ as $\rho(\mu, a) \triangleq \breve{\rho}(F_\mu, a)$. Given a belief $\mu \in \Delta(\Omega)$, we assume that the receiver chooses an action $a \in A$ that achieves the highest utility $\rho(\mu, a)$.

Observe that a receiver is an expected utility maximizer if and only if, for each $a \in A$, the utility function $\rho(\mu, a)$ is linear in μ. (Note that if $\rho(\mu, a)$ is a utility function of an agent, then so is $g(\rho(\mu, a))$ for any increasing function $g : \mathbb{R} \to \mathbb{R}$. Thus, this linearity holds only up to an increasing transformation. We suppress such transformations for the sake of clarity.) In particular, there exists a function $u : \mathcal{X} \times \Omega \times A \to \mathbb{R}$ such that $\rho(\mu, a) = \mathbf{E}[u(X, \bar{\omega}, a)] = \sum_{\omega \in \Omega} \mu(\omega) \mathbf{E}[u(X_\omega, \omega, a)]$ for all $\mu \in \Delta(\Omega)$ and $a \in A$, if and only if the receiver is an expected utility maximizer. Our setup therefore includes as a special case the standard Bayesian persuasion framework with an expected utility–maximizing receiver. However, the generality of our setting allows us to capture a much wider range of receiver behavior. (We emphasize that the notion of risk consciousness is different from, and much more general than, the conventional notion of risk aversion, where utility is modeled as the expectation of a function that is *concave in the payoffs*. A risk-averse agent is still an expected utility maximizer, and this expected utility is necessarily *linear in the agent's belief*.)

From the sender's perspective, the only relevant aspect of the receiver's belief is the corresponding receiver's action induced by that belief. Thus, rather than modeling receiver's utilities directly, we could model the receiver as being characterized by the sets of beliefs for which the receiver chooses each particular action. In other words, instead of the utility function ρ, our approach could equivalently take as model primitives the sets $\{\mathcal{P}_a : a \in A\}$, where $\mathcal{P}_a$ is the set of posterior beliefs for which action $a \in A$ is optimal for the receiver:

$$\mathcal{P}_a \triangleq \left\{ \mu \in \Delta(\Omega) : a \in \arg\max_{a' \in A} \rho(\mu, a') \right\}, \quad \text{for each } a \in A. \quad (1)$$

As we discuss in Section 6, this perspective allows us to apply our methods to more general settings beyond the context of risk-conscious receivers.

As an illustration, consider the setting of a customer deciding whether to wait for service in an unobservable queue. The receiver's utility depends on unknown waiting time X, and suppose the queue operator observes some correlated feature ω (queue length, congestion, server availability, etc.). A natural risk-conscious customer model posits that customers only join the queue and wait for service if, given their beliefs, the means of their waiting time plus a multiple of its standard deviation are below a threshold (Nikolova and Stier-Moses 2014, Cominetti and Torrico 2016, Lianeas et al. 2019). Such a behavioral model may arise from the customer's requirement for service reliability or from an aversion to uncertainty as a result of the customer's desire to plan for the day subsequent to service completion. This model can be captured in our setting by letting $A = \{\mathsf{join}, \mathsf{leave}\}$ and assuming, for example, that $\rho(\mu, \mathsf{join}) = \tau - (\mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu(X)})$ for some $\beta, \tau > 0$, and $\rho(\mu, \mathsf{leave}) = 0$, implying $\mathcal{P}_{\mathsf{join}} = \{\mu \in \Delta(\Omega) : \tau - (\mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu(X)}) \geq 0\}$ and $\mathcal{P}_{\mathsf{leave}} = \{\mu \in \Delta(\Omega) : \tau - (\mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu(X)}) \leq 0\}$. (We use the notation $\mathbf{E}_\lambda$ to denote expectation with respect to a distribution λ.) It is straightforward to check that $\rho(\mu, \mathsf{join})$ is not linear in μ.

Throughout this paper, we make the following assumption.

Assumption 1. *For each $a \in A$, the set $\mathcal{P}_a$ is closed.*

We remark that Assumption 1 holds if, for each $a \in A$, the utility function $\rho(\mu, a)$ is continuous in μ.

2.3. Persuasion of Risk-Conscious Agents

We are now ready to describe the sender's persuasion problem. First, we require that for any choice of the signaling scheme π, the receiver's strategy maximizes his or her utility with respect to his or her posterior beliefs: for each $s \in S$, we have

$$a(s) \in \arg\max_{a \in A} \rho(\mu_s, a). \quad (2)$$

We call any strategy that satisfies (2) an optimal strategy for the receiver. Given an optimal strategy $a(\cdot)$, the sender's expected utility for choosing a signaling scheme π is given by $\mathbf{E}_\pi[v(\bar{\omega}, a(\bar{s}))]$, where $\mathbf{E}_\pi$ denotes the expectation over $(\bar{\omega}, \bar{s})$ with respect to π. The sender seeks to choose a signaling scheme π that maximizes his or her expected utility, assuming that the receiver responds with an optimal strategy. (When the receiver has multiple optimal strategies, we assume that the sender chooses the most preferred one; the literature refers to this as the sender-preferred subgame-perfect equilibrium. See Kamenica and Gentzkow (2011).) Thus, the sender's problem can be posed as

$$\begin{aligned} \max_{\pi \in \Delta(\Omega \times S)} \quad & \mathbf{E}_\pi[v(\bar{\omega}, a(\bar{s}))] \\ \text{subject to} \quad & a(s) \in \arg\max_{a \in A} \rho(\mu_s, a), \quad \text{for all } s \in S, \\ & \pi(\omega, S) = \mu^*(\omega), \quad \text{for all } \omega \in \Omega. \end{aligned} \quad (3)$$

Our main goal in this paper is to find and characterize the sender's optimal signaling scheme to the persuasion problem (3). Note that the problem as posed is computationally challenging, as it requires first choosing an optimal set of signals S and then a joint distribution π over $\Omega \times S$. Without an explicit handle on the set S and the resulting receiver actions, the persuasion problem seems intractable. In the next section, we reframe the problem to obtain a tractable formulation.

3. Toward a Tractable Formulation

When the receiver is treated as an expected utility maximizer, a revelation principle–style argument is typically invoked (Bergemann and Morris 2016) to restrict attention to signaling schemes that use the set of actions A as the signaling space S, such that the receiver upon seeing a signal $a \in A$ finds it optimal to take action a. Before we discuss our approach for general risk-conscious agents, we provide a more detailed discussion of this argument and discuss why it fails in our setting.

3.1. Failure of the Revelation Principle

The revelation principle–style argument rests on the following observation: when the receiver is an expected utility maximizer, if two signals s_1 and s_2 both lead to the same optimal action $a(s_1) = a(s_2) = a$, then a is still an optimal action for the receiver if the signaling scheme reveals only that $\bar{s} \in \{s_1, s_2\}$ whenever it was supposed to reveal s_1 or s_2. This property is straightforward to show using the linearity of the utility functions $\rho(\cdot, a)$ for an expected utility maximizer. One can then use this property to coalesce all signals that lead to the same optimal action for the receiver into a single signal. Such a coalesced signaling scheme has at most one signal per action, which, after identifying the signal with the corresponding action, can be turned into an *action recommendation*. Moreover, for such a signaling scheme, the agent's optimal strategy is *obedient*; that is, it is optimal for the agent to follow the action recommendation.

However, when the receiver is risk conscious, the preceding argument may no longer hold. This is because when signals with the same optimal action are coalesced, it may alter the posterior of the receiver on the coalesced signal, and without linearity of $\rho(\cdot, a)$, the receiver's optimal action may change. (To see this, consider Example 1 in reverse: under each signal s_i it is optimal for the receiver to take the action. However, coalescing the three signals is equivalent to providing no information, and not taking the action is uniquely optimal for the receiver under his or her prior belief.) Thus, it no longer suffices to consider only those signaling schemes with action recommendations.

3.2. A Convex Programming Formulation

Despite this difficulty, a version of the preceding argument, which we term *coalescence*, continues to hold with a risk-conscious receiver. To see this, observe that if two signals $\bar{s} = s_1$ and $\bar{s} = s_2$ lead to the same posterior $\mu \in \Delta(\Omega)$ for the receiver, then the receiver's posterior is still μ if the signaling scheme reveals only that $\bar{s} \in \{s_1, s_2\}$ whenever it was supposed to reveal s_1 or s_2. This coalescence property follows immediately from the fact that the receiver's posterior belief, given $\bar{s} \in \{s_1, s_2\}$, is a convex combination of beliefs under $\bar{s} = s_i$ for $i = 1, 2$. Thus, using the same argument as before, the coalescence property allows us to coalesce all signals that lead to the same posterior belief of the receiver into a *belief recommendation*. In such a coalesced signaling scheme, we can take the signal space S to be $\Delta(\Omega)$, the set of posteriors. Furthermore, in such a scheme, a property akin to obedience holds: if the receiver is recommended a belief μ, then the posterior belief is indeed μ.

Summarizing the preceding discussion, we can write the sender's persuasion problem (3) as

$$\max_{\pi \in \Delta(\Omega \times \Delta(\Omega))} \mathbf{E}_\pi[v(\bar{\omega}, a(\bar{s}))]$$

$$\text{subject to} \quad a(s) \in \arg\max_{a \in A} \rho(\mu_s, a), \quad \text{for all } s \in \Delta(\Omega),$$
$$\pi(\omega, \Delta(\Omega)) = \mu^*(\omega), \quad \text{for each } \omega \in \Omega,$$
$$\mu_s = s, \quad \text{for all } s \in \Delta(\Omega). \tag{4}$$

Although we have characterized the set of signals, this is still a challenging problem because of the complexity of the set $\Delta(\Omega \times \Delta(\Omega))$. To make further progress, we state the following lemma (Aumann and Maschler 1995, Kamenica and Gentzkow 2011), which gives an equivalent formulation using the notion of *Bayes-plausible* measures, which are probability measures $\eta \in \Delta(\Delta(\Omega))$ over the set of beliefs $\Delta(\Omega)$ with the property that their expectation equals the prior belief μ^*. (The proof of the results in this section are in Online Appendix B.)

Lemma 1 (Aumann and Maschler 1995, Kamenica and Gentzkow 2011). *A signaling scheme $\pi \in \Delta(\Omega \times \Delta(\Omega))$ satisfies the condition $\mu_s = s$ for almost all $s \in \Delta(\Omega)$ only if the measure $\eta(\cdot) \triangleq \pi(\Omega, \cdot) \in \Delta(\Delta(\Omega))$ is Bayes plausible. Conversely, for any Bayes-plausible measure η, the signaling scheme defined as $\pi(\omega, ds) = s(\omega)\eta(ds)$ satisfies $\mu_s = s$ for all $s \in \Delta(\Omega)$.*

The preceding lemma allows us to reformulate (4) as an optimization over the space of Bayes-plausible measures $\eta \in \Delta(\Delta(\Omega))$ with objective $\sum_{\omega \in \Omega} \mathbf{E}_\eta[\bar{s}(\omega) v(\omega, a(\bar{s}))]$. (See Lemma EC.1 in Online Appendix B for the details.) This reformulation is still challenging, as it involves optimizing over a set of probability measures on $\Delta(\Omega) \subseteq \mathbb{R}^{|\Omega|}$. Our first result allows us to overcome this difficulty by establishing that one can instead optimize over a much simpler space. To state our result, we need some notation: for any set $H \subseteq \mathbb{R}^m$, let $\mathsf{Conv}(H)$ denote the convex hull of H, defined as

$$\mathsf{Conv}(H) = \left\{ y : y = \sum_{i=1}^{j} \lambda_i x_i, \text{for some } j \geq 1,\ \lambda_i \geq 0, \right.$$
$$\left. x_i \in H \text{ for all } 1 \leq i \leq j \text{ and } \sum_{i=1}^{j} \lambda_i = 1 \right\}.$$

In other words, $\mathsf{Conv}(H)$ is the set of all finite convex combinations of elements in H. We have the following lemma that states that corresponding to each Bayes-plausible measure η there exists $\{b_a \geq 0\}_{a \in A}$ and $\{m_a \in \Delta(\Omega)\}_{a \in A}$ such that the sender's expected utility under η (i.e., $\sum_{\omega \in \Omega} \mathbf{E}_\eta[\bar{s}(\omega) v(\omega, a(\bar{s}))]$) can be written as a bilinear function of m_a and b_a. Thus, the lemma allows us to directly optimize over m_a and b_a instead of over Bayes-plausible measures η.

Lemma 2. *For any Bayes-plausible measure $\eta \in \Delta(\Delta(\Omega))$ and optimal receiver strategy $a(\cdot)$, there exists $\{(b_a, m_a)\}_{a \in A}$, with $b_a \in [0, 1]$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ for each $a \in A$, such that*

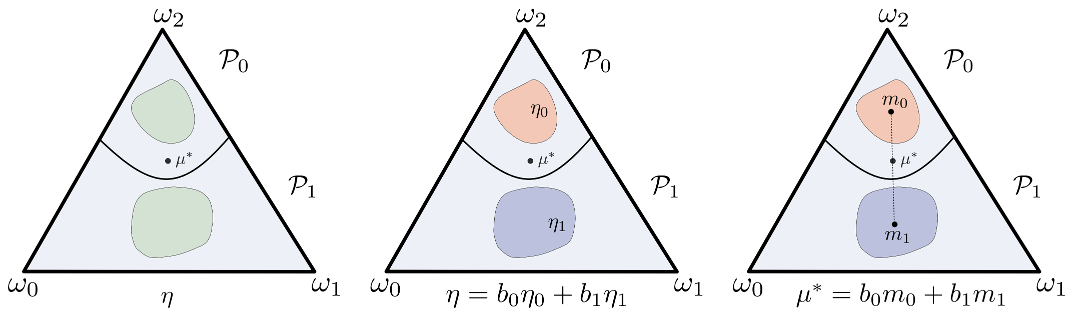
$$\sum_{a \in A} b_a m_a = \mu^*, \tag{5}$$

$$\mathbf{E}_\eta[\bar{s}(\omega) v(\omega, a(\bar{s}))] = \sum_{a \in A} b_a m_a(\omega) v(\omega, a) \quad \text{for each } \omega \in \Omega. \tag{6}$$

Conversely, for any $\{(b_a, m_a) : b_a \in [0, 1]$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ for each $a \in A\}$ *satisfying (5), there exists a Bayes-plausible measure η and an optimal receiver strategy $a(\cdot)$ such that (6) holds.*

The interpretation of the quantities m_a and b_a is as follows. Given a Bayes-plausible measure η, the quantity b_a denotes the probability that the receiver plays action $a \in A$ under the optimal strategy $a(\cdot)$ when the sender uses the signaling scheme corresponding to η; in other words, $b_a = \mathbf{P}_\eta(a(\bar{s}) = a)$. Similarly, m_a denotes the distribution of the state $\bar{\omega}$, conditioned on the receiver choosing action a. By iterated expectation, we obtain $m_a(\omega) = \mathbf{E}_\eta[\bar{s}(\omega) | a(\bar{s}) = a]$. Thus, for any $a \in A$, the quantity m_a denotes the mean of all posterior beliefs the receiver holds, conditioned on choosing action a. For this reason, we refer to m_a as the *mean posterior* of the receiver corresponding to action $a \in A$. Note that m_a may not correspond to any actual posterior that the receiver holds when choosing action $a \in A$; in fact, the mean posterior $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ may not even lie in the set $\mathcal{P}_a$. Figure 2 gives some geometric intuition for

Figure 2. (Color online) Geometric Interpretation of the Quantities m_a and b_a



Notes. In the left panel, $\eta \in \Delta(\Delta(\Omega))$ has support in the green region. Because η is Bayes plausible, $\mathbf{E}_\eta[\bar{s}(\omega)] = \mu^*(\omega)$ for each $\omega \in \Omega$. In the middle panel, we separate η into η_0 and η_1, where η_a has support in $\mathcal{P}_a$. In the right panel, we depict m_0 and m_1, where $\mathbf{E}_{\eta_a}[\bar{s}(\omega)] = m_a^*(\omega)$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$. Finally, b_a is defined as $\mathbf{P}_\eta(a(\bar{s}) = a)$.

these quantities as well as for the distributions η_a introduced in the proof.

Although Equations (5) and (6) are bilinear in m_a and b_a, we can make them linear by substituting $t_a(\omega) = b_a m_a(\omega)$ for each $\omega \in \Omega$. Notice that $b_a \in [0,1]$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ imply $t_a \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$, where $\mathbf{0} \in \mathbb{R}^{|\Omega|}$ is the vector of all zeros. With this substitution, we obtain our main theorem.

Theorem 1. *The sender's persuasion problem* (3) *can be optimized by solving the following convex optimization problem:*

$$\begin{aligned} &\max_{\{t_a : a \in A\}} \sum_{\omega \in \Omega} \sum_{a \in A} t_a(\omega) v(\omega, a) \\ &\textit{subject to} \quad \sum_{a \in A} t_a(\omega) = \mu^*(\omega), \quad \textit{for each } \omega \in \Omega, \\ &\qquad\qquad t_a \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\}), \ \textit{for each } a \in A. \end{aligned} \tag{7}$$

As m_a denotes the receiver's mean posterior conditional on choosing action a, and b_a denotes the probability the receiver chooses action a, we obtain that $t_a(\omega) = b_a m_a(\omega)$ denotes the joint probability that the receiver takes action a and the realized state is $\bar{\omega} = \omega$. Thus, the reformulation (7) directly optimizes over the joint probability distribution of the state and the receiver's actions.

We next briefly remark on the complexity of solving the convex program (7). First, note that relative to optimizing over Bayes-plausible measures $\eta \in \Delta(\Delta(\Omega))$, the convex program is extremely simple: the optimization is over $|A|$ variables t_a belonging to a convex set $\mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$ in $\mathbb{R}^{|\Omega|}$. Thus, its computational complexity rests on whether there exists an efficient characterization of the set $\mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$ for each $a \in A$, which, in turn, depends solely on the properties of the utility functions $\rho(\cdot, a)$. Although obtaining such an efficient characterization can be hard in general, in certain settings with additional structure on ρ, one can replace the convex sets $\mathsf{Conv}(\mathcal{P}_a \cup \{0\})$ with a convex polytope, yielding a linear program. We present and analyze one such setting in Section 5.

3.3. Optimal Signaling Schemes

To conclude this section, we describe how to get an optimal signaling scheme π from the optimal solution $t = \{t_a\}_{a \in A}$ to the problem (7). Because $t_a \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$ for each $a \in A$, let $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ and $b_a \in [0,1]$ be such that $t_a = b_a m_a$. (Note that for $t_a \neq \mathbf{0}$, the corresponding m_a is uniquely defined.) Because $m_a \in \mathsf{Conv}(\mathcal{P}_a)$, there exists a finite convex decomposition of m_a in terms of the elements of $\mathcal{P}_a$. That is, there exists $\{(\mu_i^a, \lambda_i^a) : i = 1, \ldots, j_a\}$ for some $j_a \geq 1$ with $\mu_i^a \in \mathcal{P}_a$, $\lambda_i^a \geq 0$, and $\sum_i \lambda_i^a = 1$, such that $m_a = \sum_{i=1}^{j_a} \lambda_i^a \mu_i^a$. An optimal signaling scheme π is then given by the (discrete) distribution over $\Delta(\Omega \times \Delta(\Omega))$ that chooses $(\bar{\omega}, \bar{s}) = (\omega, \mu_i^a)$ with probability $b_a \lambda_i^a \mu_i^a(\omega)$. Observe that, conditional on $\bar{\omega} = \omega$, the optimal signaling scheme π makes the belief recommendation μ_i^a to the receiver with probability

$$\pi(\bar{s} = \mu_i^a \mid \bar{\omega} = \omega) = \frac{b_a \lambda_i^a \mu_i^a(\omega)}{\sum_{a' \in A} b_{a'} \lambda_i^{a'} \mu_i^{a'}(\omega)}. \tag{8}$$

We note that, in general, the representation of m_a as a convex combination of $\mu_i^a \in \mathcal{P}_a$ need not be unique. Because the preceding construction works for any convex decomposition of m_a, we conclude that there may exist multiple optimal signaling schemes for the receiver.

For any set $H \subseteq \mathbb{R}^d$, let $\mathsf{Cara}(H)$ denote the minimum value of j such that any point $x \in \mathsf{Conv}(H)$ can be written as a convex combination of at most j points in H. Carathéodory's theorem (Bárány and Onn 1995) states that $\mathsf{Cara}(H) \leq \dim(H) + 1$, where $\dim(H)$ is the dimension of the smallest affine space containing H. Thus, we obtain the following bound on the size of the set of signals the sender needs to use to optimally persuade the receiver.

Proposition 1. *There exists an optimal signaling scheme $\pi \in \Delta(\Omega \times S)$, where the set of signals S satisfies $|S| \leq \sum_{a \in A} \mathsf{Cara}(\mathcal{P}_a)$. Specifically, for any $a \in A$, the signaling scheme sends at most $\mathsf{Cara}(\mathcal{P}_a) \leq |\Omega|$ signals for which the receiver's optimal action is a.*

In the case of expected utility maximizing agents, each set $\mathcal{P}_a$ is convex, and hence the preceding proposition implies that at most $\mathsf{Cara}(\mathcal{P}_a) = 1$ signals per action suffices for optimal persuasion. This matches with the bound using the revelation principle, which implies the sufficiency of action recommendations. For risk-conscious receivers, the proposition states that, in general, the sender must share more information than just action recommendations, and the additional information required to optimally induce any action may take up to $|\Omega|$ values.

The bound we obtain here is related to the bound obtained by Kamenica and Gentzkow (2011), who also use convex analytic arguments to show that at most $|\Omega|$ signals overall suffice for an optimal signaling scheme. In Online Appendix A, we discuss this connection in detail and use their bound to provide an alternative approach to arrive at the convex problem (7).

4. Benefits from Persuasion

Using Theorem 1, we now turn to the question of determining the sender's benefits from persuasion. Let $V(\mu^*)$ denote the optimal value of the program (7) as a function of the prior μ^*. Similarly, let $\hat{v}(\mu) \triangleq \mathbf{E}_\mu[v(\bar{\omega}, a_\mu)]$ denote the sender's expected payoff without persuasion as a function of the prior μ, where $a_\mu \in \{a \in A : \mu \in \mathcal{P}_a\}$ denotes the receiver's optimal action under the prior.

Note that $\hat{v}(\mu^*)$ is the objective value of (7) for the feasible solution $t = \{t_a\}_{a \in A}$ with $t_a = \mu^*$ for $a = a_{\mu^*}$ and $\mathbf{0}$ otherwise. Because $V(\mu^*)$ is its optimal value, it is immediate that the sender (strictly) benefits from persuasion if and only if $V(\mu^*) > \hat{v}(\mu^*)$. The following result characterizes when the latter condition holds; the result and its proof in Online Appendix C are analogous to the setting with expected utility–maximizing receivers (Kamenica and Gentzkow 2011). Using the terminology therein, we say *there is information the sender would share* if there exists a belief μ with $\hat{v}(\mu) > \mathbf{E}_\mu[v(\bar{\omega}, a_{\mu^*})]$. For ease of notation, let $\mathcal{P}_\mu$ denote the set $\mathcal{P}_a$ with $a = a_\mu$.

Proposition 2. *We have the following:*

(1) *If there is no information the sender would share, then the sender does not benefit from persuasion. If there is information the sender would share, and μ^* lies in the interior of $\mathcal{P}_{\mu^*}$, then the sender benefits from persuasion.*

(2) *If for each $a \in A$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ we have $\sum_{\omega \in \Omega} m_a(\omega)(v(\omega, a) - v(\omega, a_{\mu^*})) \le 0$, then the sender does not benefit from persuasion. If μ^* is in the interior of $\mathcal{P}_{\mu^*}$ and there exists an $a \in A$ and an $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ such that $\sum_{\omega \in \Omega} m_a(\omega)(v(\omega, a) - v(\omega, a_{\mu^*})) > 0$, then the sender benefits from persuasion.*

The preceding result states that if the sender benefits from persuasion, then there is an action a and a belief $\mu \in \mathsf{Conv}(\mathcal{P}_a)$ such that under prior μ the sender would strictly prefer that the receiver take action a over the action a_{μ^*}. The action a need not be optimal for the receiver with belief μ; this is indeed the case if $\mu \in \mathsf{Conv}(\mathcal{P}_a) \cap \mathcal{P}_a^c$. This latter scenario is impossible for an expected utility–maximizing receiver, for whom the sets $\mathcal{P}_a$ are convex, and hence $\mathsf{Conv}(\mathcal{P}_a) \cap \mathcal{P}_a^c = \mathcal{P}_a \cap \mathcal{P}_a^c = \emptyset$.

Having obtained the conditions under which the sender benefits from persuasion, we now turn to describing the extent of this benefit. To do this, we define the notion of *full persuasion* that we illustrated in the introduction. Using Theorem 1, we will then obtain a simple characterization of when full persuasion is possible with risk-conscious receivers.

To formally define full persuasion, we start with some notations. For each action $a \in A$, let $\Upsilon_a \subseteq \Omega$ denote the set of pure states for which action a is sender optimal:

$$\Upsilon_a = \{\omega \in \Omega : v(\omega, a) \ge v(\omega, a') \text{ for all } a' \in A\}.$$

We say the sender *fully persuades* the receiver if there exists a signaling scheme such that at each state $\omega \in \Omega$, the receiver chooses a sender-optimal action. Formally, we have the following definition.

Definition 1 (Full Persuasion). A signaling scheme $\pi \in \Delta(\Omega \times S)$ fully persuades the receiver if for each $\omega \in \Omega$ and $s \in S$ with $\pi(\omega, s) > 0$, if $a(s) = a$, then $\omega \in \Upsilon_a$. We say full persuasion is possible if there exists a signaling scheme that fully persuades the receiver.

To state the main result of this section, we make the following simplifying assumption: at each state $\omega \in \Omega$ there exists a unique action that is sender optimal. In other words, we assume that $\{\Upsilon_a : a \in A\}$ forms a partition of the state space Ω. We have the following theorem.

Theorem 2. *Suppose at each state there is a unique action that is sender optimal. Then, the sender can fully persuade the receiver if and only if $t^* = (t_a^* : a \in A)$, with $t_a^*(\omega) \triangleq \mu^*(\omega)\mathbf{I}\{\omega \in \Upsilon_a\}$ for $\omega \in \Omega$, is feasible for the convex program (7); that is, for each $a \in A$, we have $t_a^* \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$.*

Proof. Let $\pi \in \Delta(\Omega \times S)$ be a signaling scheme that fully persuades the receiver. Then, for any $s \in S$ with $a(s) = a$, the induced belief μ_s lies in the set $\mathcal{P}_a$ almost surely (with respect to π). Thus, we obtain $\mathbf{E}_\pi[\mu_{\bar{s}}\mathbf{I}\{a(\bar{s}) = a\}] \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$. Using the fact that $\mu_s(\omega) = \mathbf{P}_\pi(\bar{\omega} = \omega \mid \bar{s} = s)$, we obtain, using the tower property of conditional expectation, that

$$\begin{aligned}\mathbf{E}_\pi[\mu_{\bar{s}}(\omega)\mathbf{I}\{a(\bar{s}) = a\}] &= \mathbf{E}_\pi[\mathbf{I}\{\bar{\omega} = \omega, a(\bar{s}) = a\}] \\ &= \mathbf{E}_\pi[\mathbf{I}\{\bar{\omega} = \omega, \bar{\omega} \in \Upsilon_a\}] \\ &= \mu^*(\omega)\mathbf{I}\{\omega \in \Upsilon_a\} = t_a^*(\omega),\end{aligned}$$

where the second equality follows because, seeing as how π fully persuades the receiver and $\{\Upsilon_a : a \in A\}$ is a partition, we have $a(\bar{s}) = a$ if and only if $\bar{\omega} \in \Upsilon_a$ (π-almost surely). Thus, we have $t_a^* \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$.

Moreover, as $\{\Upsilon_a : a \in A\}$ is a partition, we have $\sum_{a\in A} t_a^* = \mu^*$. Thus, t^* is feasible for (7).

Conversely, suppose t^* is feasible for (7). From $t_a^* \in \mathrm{Conv}(\mathcal{P}_a \cup \{0\})$, we obtain that there exists $\{\mu_i^a \in \mathcal{P}_a : i = 1, \dots, j_a\}$ and $\{\lambda_i^a \geq 0 : i = 1, \dots, j_a\}$ such that $\sum_{i=1}^{j_a} \lambda_i^a \leq 1$ and $t_a^* = \sum_{i=1}^{j_a} \lambda_i^a \mu_i^a$. Because $\sum_{\omega\in\Omega}\sum_{a\in A} t_a^*(\omega) = \sum_{\omega\in\Omega}\mu^*(\omega) = 1$, we obtain that $\sum_{a\in A}\sum_{i=1}^{j_a} \lambda_i^a = 1$. Thus, we have $\mu^* = \sum_{a\in A}\sum_{i=1}^{j_a} \lambda_i^a \mu_i^a$ with $\sum_{a\in A}\sum_{i=1}^{j_a} \lambda_i^a = 1$. This implies that there exists a valid signaling scheme π that induces beliefs $\{\mu_i^a : i = 1, \dots, j_a, a \in A\}$. Furthermore, because $t_a^*(\omega) = \mu^*(\omega)\,\mathbf{I}\{\omega \in \Upsilon_a\}$ for each $\omega \in \Omega$, it must be the case that $\mu_i^a \in \Delta(\Upsilon_a)$. Thus, under π, if the receiver chooses an action $a \in A$, then the realized state is in the set Υ_a. This implies that π fully persuades the receiver. □

To build intuition about the theorem, consider the scenario where there is a fixed action a that is uniquely sender optimal at all states (i.e., $\Upsilon_a = \Omega$), and hence, $t_a^* = \mu^*$. Full persuasion in this context requires the sender to persuade the receiver to always take action a, irrespective of the realized state. The preceding theorem implies that this is possible if and only if $\mu^* \in \mathrm{Conv}(\mathcal{P}_a)$. Clearly, if $\mu^* \in \mathcal{P}_a$, the action a is optimal for the receiver under the prior belief, and the receiver needs no persuasion. In other words, if $\mu^* \in \mathcal{P}_a$, the no-information scheme already fully persuades the receiver. The theorem implies that if $\mu^* \in \mathrm{Conv}(\mathcal{P}_a) \cap \mathcal{P}_a^c$ there exists a signaling scheme that shares some state information and fully persuades the receiver. Once again, we note that this latter scenario cannot occur for an expected utility–maximizing receiver for whom we have $\mathrm{Conv}(\mathcal{P}_a) \cap \mathcal{P}_a^c = \emptyset$.

More generally, Theorem 2 implies that an expected utility–maximizing receiver can be fully persuaded if and only if t_a^*, appropriately scaled, lies in the set $\mathcal{P}_a$. In this case, the signaling scheme that "reveals the partition" (i.e., the scheme that at each state ω reveals the set Υ_a containing ω) is sufficient to fully persuade the receiver. By contrast, for a risk-conscious receiver, the signaling scheme that fully persuades the receiver may need to reveal more information than just revealing the partition. We explore this point in more detail in the following section.

5. Binary Persuasion

We now focus on a setting of practical importance that we refer to as *binary persuasion*. In this setting, the receiver's actions are binary (i.e., $A = \{0, 1\}$), and the sender's utility is always weakly higher under action 1 (i.e., $v(\omega, 1) \geq v(\omega, 0)$ for all $\omega \in \Omega$). This model matches settings where, independent of the state, the sender seeks to persuade the receiver to take an action, such as engaging with social media platforms (Candogan and Drakopoulos 2020), waiting in a queue (Lingenbrink and Iyer 2019), or purchasing a product (Lingenbrink and Iyer 2018, Drakopoulos et al. 2021).

To aid our discussion, we define the receiver's *differential utility* $\hat{\rho}(\cdot)$ as the difference in the utility between choosing action 1 and action 0: $\hat{\rho}(\mu) \triangleq \rho(\mu, 1) - \rho(\mu, 0)$ for all $\mu \in \Delta(\Omega)$. Note that action $a = 1$ is optimal for the receiver at belief μ if and only if $\hat{\rho}(\mu) \geq 0$.

5.1. Geometry of the Convex Program

The convex program (7) has variables defined over the domain $\mathrm{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$ for each action a. In this section, we show that the domain can be further simplified under the following assumption.

Assumption 2. *The set $\mathcal{P}_1^c = \{\mu \in \Delta(\Omega) : \hat{\rho}(\mu) < 0\}$ is convex.*

Intuitively, the assumption implies that the receiver is averse to uncertainty when choosing action 1: if action 1 is not optimal under beliefs μ and μ', then it cannot be optimal under a belief $\gamma\mu + (1-\gamma)\mu'$ that is obtained by inducing uncertainty between μ and μ'. Furthermore, the assumption holds for a wide class of utility functions, as the following lemma establishes. (The proof of the results in this section is in Online Appendix D.)

Lemma 3. *Suppose the differential utility $\hat{\rho}(\cdot)$ is quasiconvex. Then Assumption 2 holds.*

Specifically, Assumption 2 holds when $\rho(\mu, 1)$ is convex and $\rho(\mu, 0)$ is concave. At the same time, note that Assumption 2 is substantially weaker than requiring quasiconvexity of $\hat{\rho}$, because the latter implies that every level set $\{\mu : \hat{\rho}(\mu) < c\}$ is convex.

Using Assumption 2, we now show that the convex program (7) can be simplified to a linear program. Recall that e_ω denotes the belief that assigns all its weight to $\omega \in \Omega$. By a slight abuse of notation, we identify $\omega \in \Omega$ with e_ω and consider Ω as a subset of $\Delta(\Omega)$.

Let $K_1 \triangleq \mathcal{P}_1 \cap \Omega$ denote the set of states where action 1 is optimal for the receiver under full information. Similarly, let $K_0 \triangleq \mathcal{P}_0 \cap \Omega$ be the set of states where action 0 is optimal for the receiver under full information. Note that $K_0 \cap K_1$ may be nonempty if the receiver finds both actions optimal at some state. We let $L_0 \triangleq K_0 \cap K_1^c$ denote the set of states for which action 0 is uniquely optimal for the receiver.

Next, for $\omega_0 \in L_0$ and $\omega_1 \in K_1$, consider the set of beliefs obtained as the convex combination of ω_0 and ω_1. In this set, we let $\chi(\omega_0, \omega_1)$ denote the belief that puts the largest weight on ω_0 while still preserving the optimality of action 1 for the receiver. (Because $\mathcal{P}_1$ is closed, such a maximal convex combination exists.) Formally, for $\omega_0 \in L_0$ and $\omega_1 \in K_1$, we define $\gamma(\omega_0, \omega_1) = \sup_{\gamma\in[0,1]}\{\gamma : \gamma\omega_0 + (1-\gamma)\omega_1 \in \mathcal{P}_1\}$, and let $\chi(\omega_0, \omega_1) = \gamma(\omega_0, \omega_1)\omega_0 + (1 - \gamma(\omega_0, \omega_1))\omega_1$. Because $\mathcal{P}_0$ and $\mathcal{P}_1$ are closed, we obtain that $\chi(\omega_0, \omega_1) \in \mathcal{P}_0 \cap \mathcal{P}_1$, and hence the receiver is indifferent between actions 0 and 1 at belief $\chi(\omega_0, \omega_1)$. Finally, we define the set K_{01} to be a subset of such maximal convex combinations $\chi(\omega_0, \omega_1)$:

$$K_{01} = \{\chi(\omega_0, \omega_1) \in \Delta(\Omega) \text{ for some } \omega_0 \in L_0 \text{ and } \omega_1 \in K_1\}. \tag{9}$$

We note that whereas $K_{01} \cap L_0$ must be empty, the sets $K_{01} \cap K_0$ and $K_{01} \cap K_1$ can be nonempty. We illustrate these sets pictorially in Figure 3 (and in Figure EC.1 in Online Appendix D).

With these definitions, we can state the main theorem of this section.

Theorem 3. *Under Assumption 2, the sender's persuasion problem (7) can be optimized by solving the following linear program*:

$$\begin{aligned} \max_{t_0, t_1} \sum_{\omega \in \Omega} & v(\omega, 1) t_1(\omega) + v(\omega, 0) t_0(\omega) \\ \textit{subject to} \quad & t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\}), \\ & t_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\}), \\ & t_0(\omega) + t_1(\omega) = \mu^*(\omega) \quad \textit{for each } \omega \in \Omega. \end{aligned} \tag{10}$$

The proof's intuition is as follows. First, under Assumption 2, we prove in Lemma EC.3 that the set $\mathsf{Conv}(\mathcal{P}_1)$ is a convex polytope with extreme points in $K_1 \cup K_{01}$, and hence the constraint $t_1 \in \mathsf{Conv}(\mathcal{P}_1 \cup \{0\})$ in (7) is equivalent to the constraint $t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{0\})$. Second, we establish that any $t_0 \in \mathsf{Conv}(\mathcal{P}_0 \cup \{0\})$ that is not in $\mathsf{Conv}(L_0 \cup \{0\})$ cannot be part of an optimal solution by improving on any such feasible solution. See Figure 4 (and Figure EC.2 in Online Appendix D) for some geometric intuition.

5.2. Structural Characterizations

The preceding theorem implies several results about the structure of the optimal signaling scheme.

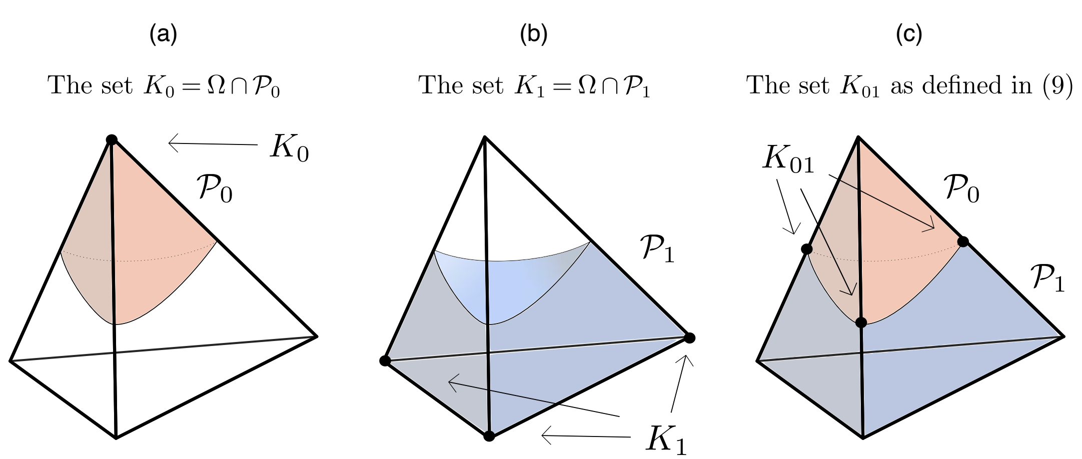
First, Theorem 3 establishes a *canonical set* of signals for an optimal signaling scheme-namely, the set $\Omega \cup K_{01}$. In other words, for a given binary persuasion setting, irrespective of the prior belief μ^*, it suffices to only send signals in the set $\Omega \cup K_{01}$. This follows from the fact that Theorem 3 implies that the mean posteriors m_0 and m_1 satisfy $m_0 \in \mathsf{Conv}(L_0)$ and $m_1 \in \mathsf{Conv}(K_1 \cup K_{01})$. Hence, m_0 can be expressed as a convex combination of beliefs in L_0, and m_1 can be expressed as a convex combination of beliefs in $K_1 \cup K_{01}$. Thus, it follows that inducing beliefs in the set $L_0 \cup K_1 \cup K_{01} = \Omega \cup K_{01}$ suffices for optimal persuasion.

Second, observe that the canonical set consists of *pure* signals $\omega \in \Omega$, which fully reveal the state ω, and *binary mixed* signals $\chi(\omega_0, \omega_1) \in K_{01}$, which induce a belief over two states $\omega_0 \in L_0$ and $\omega_1 \in K_1$. Thus, the optimal persuasion can always be achieved by either fully revealing the state or making the receiver uncertain about two states. Of course, because of the convexity of $\mathcal{P}_1^c$, there also exists an optimal signaling scheme where all posteriors that lead to the receiver taking action 0 are replaced with a single action recommendation, "take action 0."

Note that because $t_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$, if the optimal signaling scheme induces the sender's least preferred action 0, then the receiver's belief only puts weight on states in the set L_0. At each of these states, the action 0 is uniquely optimal for the receiver. This is analogous to a similar structural result for expected utility–maximizing receivers (Kamenica and Gentzkow 2011, proposition 4).

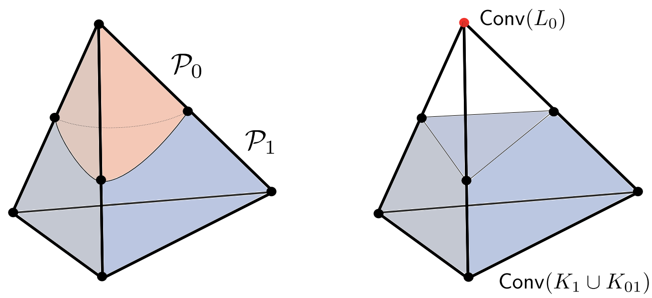
In the setting of binary persuasion, our earlier results regarding full persuasion simplify and yield a simple condition to determine whether the receiver can be fully persuaded. Suppose at all states the sender strictly prefers action 1. Then, using the notation of Section 4, we have $\Upsilon_1 = \Omega$, and thus Theorem 2 implies that the receiver can be fully persuaded if and only if $\mu^* \in \mathsf{Conv}(\mathcal{P}_1)$. As we prove in Lemma EC.3, under Assumption 2, we have $\mathsf{Conv}(\mathcal{P}_1) = \mathsf{Conv}(K_1 \cup K_{01})$. Thus we obtain a simple criterion-namely, whether $\mu^* \in \mathsf{Conv}(K_1 \cup K_{01})$-to determine whether full persuasion is possible.

Figure 3. (Color online) The Sets K_0, K_1, and K_{01}



Notes. Here, the red region is the set $\mathcal{P}_0$, and the blue region is $\mathcal{P}_1$. Furthermore, in this example, $L_0 = K_0$.

Figure 4. (Color online) Geometry of $\mathsf{Conv}(L_0)$ and $\mathsf{Conv}(K_1 \cup K_{01})$



Note. Here, $\mathsf{Conv}(L_0)$ is the red vertex of the simplex.

Finally, the linear programming formulation allows us to further characterize the structure of the optimal signaling scheme when the differential utility function $\hat{\rho}$ satisfies a monotonicity condition. For simplicity, we focus on the case where $v(\omega, 1) - v(\omega, 0) = v > 0$ for all $\omega \in \Omega$. We consider the following condition on the differential utility function.

Assumption 3 (Monotonicity). *There exists a strict total order $\prec$ on Ω such that if $\omega \prec \omega'$, then either* (1) *$\omega \in K_1$ or* (2) *$\omega, \omega' \in L_0$ and $\gamma(\omega, \hat{\omega}) > \gamma(\omega', \hat{\omega})$ for all $\hat{\omega} \in K_1$.*

Recall that for $\omega_0 \in L_0$ and $\omega_1 \in K_1$, $\gamma(\omega_0, \omega_1)$ is the largest value of $\gamma \in [0, 1]$ such that the belief $\gamma\omega_0 + (1 - \gamma)\omega_1$ lies in $\mathcal{P}_1$. Thus, the preceding monotonicity condition implies that for any $\omega, \omega' \in L_0$ with $\omega \prec \omega'$ and for any $\omega_1 \in K_1$, if $\gamma\omega' + (1 - \gamma)\omega_1$ lies in $\mathcal{P}_1$, then so does $\gamma\omega + (1 - \gamma)\omega_1$.

A wide class of differential utility functions satisfies monotonicity, including the class of expected utility functions. In particular, given a strict total order on Ω, the condition holds for any differential utility function $\hat{\rho}(\cdot)$ that is strictly decreasing under the first-order stochastic ordering on μ.

We have the following result.

Proposition 3. *Suppose $v(\omega, 1) - v(\omega, 0) = v > 0$ for all $\omega \in \Omega$, and the differential utility function $\hat{\rho}$ satisfies Assumptions 2 and 3. Then the optimal signaling scheme induces a threshold structure in the receiver's actions: there exists an $\omega_0 \in L_0$ such that*

$$t_1(\omega) = \begin{cases} \mu^*(\omega) & \text{if } \omega \prec \omega_0, \\ 0 & \text{if } \omega_0 \prec \omega. \end{cases}$$

In particular, under the optimal signaling scheme, the receiver chooses action 1 in states $\omega \prec \omega_0$ and chooses action 0 in states ω with $\omega_0 \prec \omega$.

Finally, we emphasize that it is the induced action of the receiver that has a threshold structure; the optimal signaling scheme typically possesses a more intricate structure in order to induce beliefs in $\Omega \cup K_{01}$. (We illustrate this structure in more detail in an application in Section 7.)

6. Broader Applicability to Other Contexts

Next, we discuss the broader applicability of our approach to more general contexts beyond the literal interpretation of persuading risk-conscious receivers.

As mentioned in the introduction, risk consciousness arises naturally in dynamic decision-making settings even with expected utility maximizers. To illustrate, consider a two-period setting where the receiver, after choosing an action $a \in A$ in the first period, must choose an action $d \in D$ in the second period. After taking the first-period action a, but prior to the choice of d, the receiver obtains additional state information whose quality depends on the choice of the first-period action. Formally, the receiver observes the realization of a random variable Z whose distribution depends on (ω, a). A natural model of utility that captures this setting is given by

$$\rho(\mu, a) \triangleq \mathbf{E}_\mu \left[u(\omega, a) + \max_{d \in D} \ \mathbf{E}_\mu[r_a(\omega, d) | Z, a] \right],$$

where $u(\omega, a)$ is the receiver's first-period utility, and $r_a(\omega, d)$ denotes the receiver's utility in the second period. Letting $d_\mu(Z, a)$ denote the decision rule that attains the inner maximization, the receiver's utility can be written as $\rho(\mu, a) = \mathbf{E}_\mu[u(\omega, a) + r_a(\omega, d_\mu(Z, a))]$. It follows that, in general, $\rho(\mu, a)$ is nonlinear in μ. A special case of this general model arises in data markets, where a buyer must choose whether to purchase relevant data ($a = 1$) or not ($a = 0$) before making a decision (Bergemann and Bonatti 2015, 2019; Bergemann et al. 2018; Zheng and Chen 2021). Letting Z denote the information content of the data, we have $\rho(\mu, 0) = \max_{d \in D} \mathbf{E}_\mu[r_0(\omega, d)]$ and $\rho(\mu, 1) = \mathbf{E}_\mu[\max_{d \in D} \mathbf{E}_\mu[r_1(\omega, d) | Z]]$. In such a setting, a seller who wishes to persuade the buyer to purchase the data faces the problem of persuading a risk-conscious receiver.

Our methods can be used to study the *public persuasion* (Das et al. 2017, Arieli and Babichenko 2019, Bimpikis et al. 2019, Yang et al. 2019, Candogan and Drakopoulos 2020) of a group of interacting agents, where the sender shares information publicly with all the agents. For any public signal, the agents share a *common posterior* and subsequently play an equilibrium of an incomplete information game. The sender seeks to publicly share payoff-relevant information to influence the agents' choice of the equilibrium. To apply our methods, we view the group of agents as a single risk-conscious receiver and the equilibrium profile as the action chosen by the receiver. Then, for any equilibrium a, the set $\mathcal{P}_a$ describes the set of common posteriors for which the agents play the same equilibrium profile a. With this mapping, our results can be used to find the optimal public signaling scheme, as long as the set of (relevant) equilibria over all common posteriors is finite (see, e.g., Yang et al. (2019)).

Another application of our methods is to *robust persuasion* (Ziegler 2020, Hu and Weng 2021, Pavan and Inostroza 2021), where a sender persuades a single receiver with a private type $\theta \in \Theta$. The sender takes a worst-case view and seeks to persuade the receiver irrespective of his or her type. Formally, suppose the utility of the type θ receiver with belief μ for action a is given by $\rho_\theta(\mu, a)$, and let $a_\theta(\mu) \in \arg\max_a \rho_\theta(\mu, a)$ denote the optimal action chosen by the receiver. Let $v(a)$ denote the sender's (state-independent) utility when the receiver chooses action a. Because the receiver's type θ is unknown to the sender, the receiver maximizes the (expectation) of the minimum of his or her utility across all receiver types: $\mathbf{E}[\min_\theta v(a_\theta(\mu_s))]$. Here, μ_s is the receiver's posterior belief subsequent to persuasion. Such a setting of robust persuasion maps to our model, with the sets $\mathcal{P}_a$ given by $\mathcal{P}_a = \{\mu \in \Delta(\Omega) : v(a) = \min_\theta v(a_\theta(\mu))\}$ for each $a \in A$. Thus, our results yield a robust signaling scheme through a convex program.

7. Application: Signaling in Unobservable Queues

We conclude the paper with an illustrative application of our methodology to study information sharing in a service system where arriving customers must choose whether to join an unobservable queue to obtain service. The model is based on that of Lingenbrink and Iyer (2019), with the difference being that here customers are not expected utility maximizers. Instead, inspired by the literature on the psychology of waiting in queues (Maister 1984), we consider customers who exhibit uncertainty aversion. Formally, the customers have a mean-standard deviation utility (Nikolova and Stier-Moses 2014, Cominetti and Torrico 2016, Lianeas et al. 2019), where the disutility for joining the queue is the sum of the mean waiting time and a multiple of its standard deviation. Moreover, the setting has a key difference from the model in Section 2: the customers' (receiver's) prior belief is endogenously determined from the equilibrium queue dynamics. We show that our theoretical results carry over to this setting and establish that the optimal signaling scheme has an intricate "sandwich" structure.

We consider a service system modeled as an unobservable $M/M/1/C$ first-in, first-out queue-that is, a single-server queue with Poisson arrivals with rate λ, independent exponential service times with unit mean, and queue capacity $C > 0$. Upon arrival, each customer chooses whether to join the queue to receive service ($a = 1$) or leave without obtaining service ($a = 0$); a customer cannot join the queue if there are C customers already in the queue. We assume that customers are averse to waiting but cannot observe the queue length before making the decision to join. Instead, the service provider can observe the queue length and communicate this information to arriving customers. The service provider aims to maximize the queue throughput; if service is offered at a fixed price, this translates to maximizing the revenue rate.

As the customer cannot join the queue if the queue length is C, the relevant state space is $\Omega = \{0, 1, \ldots, C-1\}$, where the state $\bar{\omega} \in \Omega$ describes the queue length upon a customer's arrival. To focus on throughput maximization, we set $v(\omega, a) = a$ for $\omega \in \Omega$ and $a \in \{0, 1\}$.

For an arriving customer, the payoff-relevant variable X is the customer's waiting time until service completion. When $\bar{\omega} = n \in \Omega$, the waiting time X is distributed as the sum of $n + 1$ independent unit exponentials (the waiting times for n customers in the queue plus the customer's own service time). To capture uncertainty aversion on the part of the customers (Maister 1984), we focus on the following differential utility function for joining the queue:

$$\hat{\rho}(\mu) \triangleq \tau - \left(\mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu[X]}\right), \tag{11}$$

where $\mu \in \Delta(\Omega)$ is the customer's belief about the queue length, $\tau > 0$ captures the customer's value for service, and $\beta \geq 0$ captures the customer's degree of risk consciousness. It is straightforward to verify that $\hat{\rho}(\cdot)$ satisfies convexity (Assumption 2) and monotonicity (Assumption 3); see Online Appendix E for details.

Using the results from Section 5 and the same approach as in Lingenbrink and Iyer (2019) to handle endogenous priors, we obtain that the service provider's signaling problem can be optimized by solving the following linear program:

$$\max_{t_0, t_1} \sum_{\omega \in \Omega} t_1(\omega)$$

$$\text{subject to} \quad t_1 \in \mathrm{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\}), \tag{12a}$$

$$t_0 \in \mathrm{Conv}(L_0 \cup \{\mathbf{0}\}), \tag{12b}$$

$$t_0(\omega+1) + t_1(\omega+1) = \lambda t_1(\omega), \quad \text{for each } \omega < C-1, \tag{12c}$$

$$\sum_{\omega \in \Omega} t_0(\omega) + \sum_{\omega \in \Omega} t_1(\omega) + \lambda t_1(C-1) = 1. \tag{12d}$$

Here, the objective captures the probability that an arriving customer joins the queue, whereas Constraint (12c) captures the detailed balance conditions on the steady-state distribution of the queue. The constraint (12d) is the normalization condition for the steady-state distribution (with $\lambda_1 t_1(C-1)$ being the probability that the queue is at capacity).

Using the monotonicity of the differential utility function, our first result shows that the optimal signaling scheme induces a threshold structure on the customers' actions. The proof uses a perturbation argument similar to Lingenbrink and Iyer (2019) and is presented in Online Appendix E.1.

Lemma 4. *An optimal signaling scheme induces a threshold structure in the customers' actions: there exists an $m \in \Omega$ such that an arriving customer joins the queue if the queue length is strictly less than m and leaves if it is strictly greater than m.*

Although the preceding lemma provides insights into the structure of the customers' actions, it does not reveal the structure of an optimal signaling scheme. The following main result of this section characterizes the intricate structure of an optimal signaling scheme using the canonical set of signals (as described in Section 5). The proof is given in Online Appendix E.2.

Proposition 4. *Suppose the optimal solution $t = (t_0, t_1)$ to (12) satisfies $t_0 \neq 0$. Then, there exists an optimal signaling scheme with signals $S = \{\mathsf{Join}_1, \mathsf{Join}_2, \ldots, \mathsf{Join}_J, \mathsf{Leave}\}$ for some $J \le |\Omega|$ such that joining the queue is optimal under each signal Join_j, and leaving is optimal under the signal Leave. Furthermore, we have the following:*

(1) *For each $j \le J$, a customer's utility upon receiving the signal Join_j is 0; that is, $\hat{\rho}(\mu_j) = 0$, where μ_j is the induced belief upon receiving signal Join_j.*

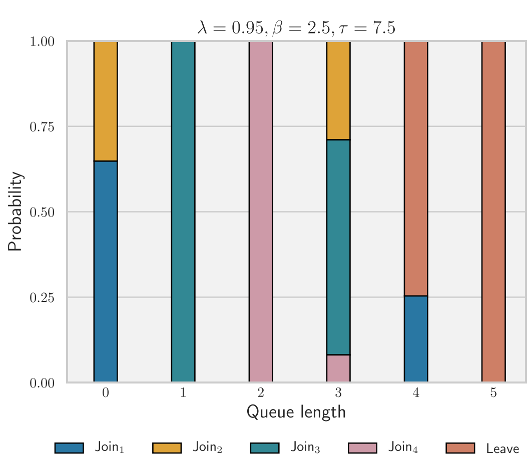
(2) *For each $j \le J$, either the induced belief μ_j puts all its weight on a state $\omega^j \in K_1 \cap K_0$ or there exists two states $\omega_0^j \in L_0$ and $\omega_1^j \in K_1$ such that μ_j puts positive weight only on the states ω_0^j and ω_1^j. (In the former case, we define $\omega_0^j = \omega_1^j = \omega^j$.) These states form a sandwich structure: $\omega_1^1 \le \omega_1^2 \le \cdots \le \omega_1^J \le \omega_0^J \le \omega_0^{J-1} \le \cdots \le \omega_0^1$.*

(3) *For each $j \le J$, we have $\mathbf{E}[X|\mathsf{Join}_{j-1}] \le \mathbf{E}[X|\mathsf{Join}_j]$ and $\mathrm{Var}(X|\mathsf{Join}_{j-1}) \ge \mathrm{Var}(X|\mathsf{Join}_j)$.*

The states in the preceding proposition implies that, assuming not all customers join the queue, the customer's utility upon receiving any Join_j signal is 0, and hence $\mathbf{E}_\mu[X|\mathsf{Join}_j] + \beta\sqrt{\mathrm{Var}_\mu[X|\mathsf{Join}_j]} = \tau$ for all $j \le J$. Thus, the optimal signaling mechanism compensates for the higher expected waiting times under some signals with lower variance, leading to the sandwich structure in the induced posterior beliefs.

We numerically illustrate the preceding result in Figure 5 for $(\lambda, \beta, \tau) = (0.95, 2.5, 7.5)$ and $C = 100$. We depict an optimal signaling scheme that uses five signals (Join_1, Join_2, Join_3, Join_4, and Leave), the first four of which persuade the customer to join. The plot on the left shows the probabilities of sending each signal conditional on the queue length. For instance, if the queue length is $\omega = 3$ upon a customer's arrival, the signals Join_2, Join_3, Join_4 are sent with probabilities 0.29, 0.63, and 0.08, respectively. As established in Lemma 4, the plot reveals the threshold structure induced on the customers' actions: a customer joins for queue lengths smaller than 4 and leaves for queue lengths larger than 4. The table on the right shows the customer's induced posterior belief upon receiving each signal. We see that each signal Join_j puts weight only on two states, and these states form a sandwich structure. The sandwich structure also orders the signals in the increasing order of mean waiting times; because the customer's differential utility upon receiving any Join_j signal is 0, the variance of the waiting time is decreasing. This ordering provides a convenient vocabulary for the service provider to communicate with the customers: rather than sending an abstract signal such as Join_j, the service provider can directly convey the induced expected waiting times $\mathbf{E}[X|\mathsf{Join}_j]$ (or the induced uncertainty $\sqrt{\mathrm{Var}(X|\mathsf{Join}_j)}$).

Figure 5. (Color online) Optimal Signaling Schemes



Signal s	Posterior belief	$\mathbf{E}[X\mid s]$
Join_1	$0.76 \cdot e_0 + 0.24 \cdot e_4$	1.97
Join_2	$0.59 \cdot e_0 + 0.41 \cdot e_3$	2.24
Join_3	$0.64 \cdot e_1 + 0.36 \cdot e_3$	2.72
Join_4	$0.93 \cdot e_2 + 0.07 \cdot e_3$	3.07
Leave	$0.76 \cdot e_4 + 0.24 \cdot e_5$	5.24

Acknowledgments

The authors thank the area editor, the anonymous associate editor, and the referees for their constructive feedback. Portions of this work were done when the second author was with Cornell University. A preliminary version of this work appeared as a one-page abstract at the 15th Conference on Web and Internet Economics (WINE 2019) and as a poster at the 2019 Workshop on Behavioral EC; the authors thank the anonymous reviewers at these conferences for their feedback.

References

Acerbi C, Tasche D (2002) Expected shortfall: A natural coherent alternative to value at risk. *Econom. Notes* 31(2):379–388.

Ahmadi-Javid A (2012) Entropic value-at-risk: A new coherent risk measure. *J. Optim. Theory Appl.* 155(3):1105–1123.

Alizamir S, de Véricourt F, Wang S (2020) Warning against recurring risks: An information design approach. *Management Sci.* 66(10):4612–4629.

Allais M (1979) The so-called Allais paradox and rational decisions under uncertainty. Allais M, Hagen O, eds. *Expected Utility Hypotheses and the Allais Paradox: Contemporary Discussions of the Decisions Under Uncertainty with Allais' Rejoinder* (Springer, Dordrecht, Netherlands), 437–681.

Anunrojwong J, Iyer K, Manshadi V (2020) Information design for congested social services: Optimal need-based persuasion. Biró P, Hartline J, Ostrovsky M, Procaccia A, eds. *Proc. 21st ACM Conf. Econom. Comput.* (ACM, New York), 349–350.

Arieli I, Babichenko Y (2019) Private Bayesian persuasion. *J. Econom. Theory* 182(July):185–217.

Artzner P, Delbaen F, Eber J-M, Heath D (2001) Coherent measures of risk. *Math. Finance* 9(3):203–228.

Aumann RJ, Maschler M (1995) *Repeated Games with Incomplete Information* (MIT Press, Cambridge, MA).

Bárány I, Onn S (1995) Carathéodory's theorem, colourful and applicable. Bárány I, Böröczky K, eds. *Intuitive Geometry*, Bolyai Society Mathematical Studies, Vol. 6 (János Bolyai Mathematical Society, Budapest), 11–21.

Beauchêne D, Li J, Li M (2019) Ambiguous persuasion. *J. Econom. Theory* 179(January):312–365.

Bergemann D, Bonatti A (2015) Selling cookies. *Amer. Econom. J. Microeconom.* 7(3):259–294.

Bergemann D, Bonatti A (2019) Markets for information: An introduction. *Annual Rev. Econom.* 11(1):85–107.

Bergemann D, Morris S (2016) Bayes correlated equilibrium and the comparison of information structures in games. *Theoret. Econom.* 11(2):487–522.

Bergemann D, Morris S (2019) Information design: A unified perspective. *J. Econom. Literature* 57(1):44–95.

Bergemann D, Bonatti A, Smolin A (2018) The design and price of information. *Amer. Econom. Rev.* 108(1):1–48.

Bimpikis K, Papanastasiou Y (2019) Inducing exploration in service platforms. Hu M, ed. *Sharing Economy: Making Supply Meet Demand* (Springer, Cham, Switzerland), 193–216.

Bimpikis K, Ehsani S, Mostagir M (2019) Designing dynamic contests. *Oper. Res.* 67(2):339–356.

Candogan O (2019) Persuasion in networks: Public signals and k-cores. Karlin A, Immorlica N, Johari R, eds. *Proc. 2019 ACM Conf. Econom. Comput.* (ACM, New York), 133–134.

Candogan O (2020) Information design in operations. Druehl C, Elmaghraby W, eds. *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*, INFORMS TutORials in Operations Research (INFORMS, Catonsville, MD), 176–201.

Candogan O, Drakopoulos K (2020) Optimal signaling of content accuracy: Engagement vs. misinformation. *Oper. Res.* 68(2):497–515.

Cominetti R, Torrico A (2016) Additive consistency of risk measures and its application to risk-averse routing in networks. *Math. Oper. Res.* 41(4):1510–1521.

Das S, Kamenica E, Mirka R (2017) Reducing congestion through information design. *Proc. 55th Annual Allerton Conf. Comm. Control Comput.* (*Allerton*) (IEEE Press, New York), 1279–1284.

DellaVigna S (2009) Psychology and economics: Evidence from the field. *J. Econom. Literature* 47(2):315–372.

Drakopoulos K, Jain S, Randhawa R (2021) Persuading customers to buy early: The value of personalized information provisioning. *Management Sci.* 67(2):828–853.

Dughmi S, Xu H (2016) Algorithmic Bayesian persuasion. Wichs D, Mansour Y, eds. *Proc. 48th Annual ACM Sympos. Theory Comput.* (ACM, New York), 412–425.

Ellsberg D (1961) Risk, ambiguity, and the Savage axioms. *Quart. J. Econom.* 75(4):643–669.

Föllmer H, Schied A (2016) *Stochastic Finance: An Introduction in Discrete Time* (De Gruyter, Berlin).

Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *J. Math. Econom.* 18(2):141–153.

Hu J, Weng X (2021) Robust persuasion of a privately informed receiver. *Econom. Theory* 72(3):909–953.

Jorion P (2006) *Value at Risk: The New Benchmark for Managing Financial Risk* (McGraw-Hill, New York).

Kahneman D, Tversky A (1972) Subjective probability: A judgment of representativeness. *Cognitive Psych.* 3(3):430–454.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.

Kamenica E (2019) Bayesian persuasion and information design. *Annual Rev. Econom.* 11(1):249–272.

Kamenica E, Gentzkow M (2011) Bayesian persuasion. *Amer. Econom. Rev.* 101(6):2590–2615.

Kolotilin A, Mylovanov T, Zapechelnyuk A, Li M (2017) Persuasion of a privately informed receiver. *Econometrica* 85(6):1949–1964.

Lianeas T, Nikolova E, Stier-Moses NE (2019) Risk-averse selfish routing. *Math. Oper. Res.* 44(1):38–57.

Lingenbrink D, Iyer K (2018) Signaling in online retail: Efficacy of public signals. Preprint, submitted May 15, http://dx.doi.org/10.2139/ssrn.3179262.

Lingenbrink D, Iyer K (2019) Optimal signaling mechanisms in unobservable queues. *Oper. Res.* 67(5):1397–1416.

Lipnowski E, Mathevet L (2018) Disclosure to a psychological audience. *Amer. Econom. J. Microeconom.* 10(4):67–93.

Machina MJ (1982) "Expected utility" analysis without the independence axiom. *Econometrica* 50(2):277–323.

Machina MJ (1995) Non-expected utility and the robustness of the classical insurance paradigm. *Geneva Papers Risk Insurance Theory* 20(1):9–50.

Maister DH (1984) *The Psychology of Waiting Lines.* HBS Background Note 684-064, Harvard Business School, Boston.

Markowitz HM (1952) Portfolio selection. *J. Finance* 7(1):77–91.

Nikolova E, Stier-Moses NE (2014) A mean-risk model for the traffic assignment problem with stochastic travel times. *Oper. Res.* 62(2):366–382.

Papanastasiou Y, Bimpikis K, Savva N (2018) Crowdsourcing exploration. *Management Sci.* 64(4):1727–1746.

Pavan A, Inostroza N (2021) Persuasion in global games with application to stress testing. CEPR Discussion Paper DP16484, Centre for Economic Policy Research, London.

Rabin M (1998) Psychology and economics. *J. Econom. Literature* 36(1):11–46.

Rayo L, Segal I (2010) Optimal information disclosure. *J. Political Econom.* 118(5):949–987.

Romanyuk G, Smolin A (2019) Cream skimming and information design in matching markets. *Amer. Econom. J. Microeconom.* 11(2):250–276.

Savage LJ (1954) *The Foundations of Statistics* (John Wiley & Sons, New York).

Taneva I (2019) Information design. *Amer. Econom. J. Microeconom.* 11(4):151–185.

Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* 5(4):297–323.

Yang P, Iyer K, Frazier PI (2019) Information design in spatial resource competition. (Abstract.) Caragiannis I, Mirrokni V, Nikolova E, eds. Proc. 15th Internat. Conf. Web Internet Econom. (Springer, Cham, Switzerland), 346.

Zheng S, Chen Y (2021) Optimal advertising for information products. Biró P, Chawla S, Echenique F, eds. *Proc. 22nd ACM Conf. Econom. Comput.* (ACM, New York), 888–906.

Ziegler G (2020) Adversarial bilateral information design. Working paper, Northwestern University, Evanston, IL.

Jerry Anunrojwong is a PhD candidate in the Decision, Risk, and Operations Division at Columbia Business School. His research interests include robustness in mechanism design and revenue management, data-driven decision making, and information design, with applications to online markets and service systems.

Krishnamurthy Iyer is an associate professor in the Department of Industrial and Systems Engineering at the University of Minnesota. His research interests include game theory, mechanism design, and stochastic modeling, with applications in markets and service systems.

David Lingenbrink received his PhD in the School of Operations Research and Information Engineering at Cornell University and is currently a software engineer at Google in New York City. His research interests include optimal information sharing and persuasion.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Persuading Risk-Conscious Agents: A Geometric Approach

Online Appendix

Jerry Anunrojwong, Krishnamurthy Iyer and David Lingenbrink

Appendix A: Relation to the Concavification Approach

In their seminal work, Kamenica and Gentzkow (2011) (hereafter [KG11]) present a convex analytic concavification approach to find the optimal signaling scheme in persuasion problems. Furthermore, in Proposition 4 of the online appendix of [KG11], the authors establish that the number of signals in an optimal signaling scheme can be upper-bounded by the cardinality $|\Omega|$ of the state space. While the focus in [KG11] is on the case of an expected utility maximizing receiver, the approach presented therein also applies to the case of risk-conscious receiver. In this appendix, we discuss how the concavification approach relates to ours, and show how their results provide an alternative path to arrive at the convex programming formulation (7).

A.1. Convex Analytic Argument of Kamenica and Gentzkow (2011)

To describe their approach in detail, define $\hat{v}(s) \triangleq \sum_{\omega \in \Omega} s(\omega) v(\omega, a(s))$ for $s \in \Delta(\Omega)$, where $a(s) \in \arg\max_{a \in A} \rho(s, a)$ denotes the receiver's optimal action under belief $s \in \Delta(\Omega)$. Using this definition and Lemma 1, the persuasion problem (4) can be written as the following problem:

$$\max_{\eta \in \Delta(\Delta(\Omega))} \mathbf{E}_\eta \left[\hat{v}(\bar{s})\right] \quad \text{subject to} \quad \mathbf{E}_\eta[\bar{s}] = \mu^*.$$

Thus, letting $V(\mu) \triangleq \sup_{\eta \in \Delta(\Delta(\Omega))} \{\mathbf{E}_\eta\left[\hat{v}(\bar{s})\right] : \mathbf{E}_\eta[\bar{s}] = \mu\}$, the sender's largest payoff from persuasion is given by $V(\mu^*)$.

The main result of [KG11] is that $V(\cdot)$ is the smallest concave function that dominates $\hat{v}$. In particular, the authors establish that, for all $\mu \in \Delta(\Omega)$,

$$V(\mu) = \sup\{x : (x, \mu) \in \mathsf{Conv}(\mathsf{hyp}(\hat{v}))\}, \tag{EC.1}$$

where $\mathsf{hyp}(\hat{v}) \triangleq \{(x, s) : x \in \mathbb{R}, s \in \Delta(\Omega), x \le \hat{v}(s)\}$ denotes the hypograph of $\hat{v}$. Furthermore, the authors show that any representation of $(\mu^*, V(\mu^*))$ as a convex combination of elements of $\mathsf{hyp}(\hat{v})$ yields an optimal signaling scheme. Note that (EC.1) yields a convex program to compute $V(\mu^*)$.

To obtain the bound on the number of signals in the optimal signaling scheme, the authors apply the Fenchel-Bunt theorem (see the online appendix of [KG11] for the detailed argument) to $\mathsf{hyp}(\hat{v})$ to show that any $(x, \mu) \in \mathsf{Conv}(\mathsf{hyp}(\hat{v}))$ can be written as a convex combination of at most

$|\Omega|$ elements of $\mathsf{hyp}(\hat{v})$. From this, the authors deduce the existence of an optimal signaling scheme with at most $|\Omega|$ signals.

Our approach has many parallels with the preceding approach. First, our approach requires us to compute the set $\mathcal{P}_a$ for each $a \in A$, and compute the convex hull $\mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$. In contrast, the preceding approach requires computing the set $\mathsf{Conv}(\mathsf{hyp}(\hat{v}))$. Second, we formulate the convex optimization problem (7) with variables $t_a \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$ for each $a \in A$. In contrast, the characterization (EC.1) of $V(\mu^*)$ suggests a convex optimization formulation with variables $(x, \mu) \in \mathsf{Conv}(\mathsf{hyp}(\hat{v}))$. Finally, we use the Caratheodory theorem to split each optimal t_a into at most $\mathsf{Cara}(\mathcal{P}_a) \leq |\Omega|$ signals (see Proposition 1). In contrast, the preceding approach directly splits the point $(\mu^*, V(\mu^*))$ into at most $|\Omega|$ signals using the Fenchel-Bunt theorem.

We note that since our splitting argument applies to each t_a separately, our argument furnishes an upper-bound of at most $\mathsf{Cara}(\mathcal{P}_a) \leq |\Omega|$ signals per action, and $\sum_a \mathsf{Cara}(\mathcal{P}_a) \leq |\Omega| \cdot |A|$ signals in total, in the optimal signaling scheme. In comparison, the argument of [KG11] provides an upper-bound of at most $|\Omega|$ signals in total in the optimal signaling scheme. Thus, when the state space is large and the action space is small, our bound might be better, whereas the bound of $|\Omega|$ might be better if the action space is large. We further note that the advantage of our approach is that in some cases, it provides a canonical set of signals, as we show in the case of binary persuasion in Section 5.

A.2. Alternative Argument Yielding Convex Formulation (7)

In this section, we present an alternative argument to arrive at the convex programming formulation (7) using the upper-bound result in [KG11]. In fact, we show that this formulation can be arrived at using any finite bound larger than $|\Omega|$ on the number of signals in the optimal signaling scheme.

To begin, let $B \geq |\Omega|$ be a bound on the number of signals required in the optimal signaling scheme. Hence, for any receiver action $a \in A$, there are at most B signals that induce it under the optimal signaling scheme. Thus, it suffices to consider a signaling scheme, with signals $S = \{(a, i) : a \in A, 1 \leq i \leq B\}$, where each signal $(a, i) \in S$ induces the receiver to take action $a \in A$. Thus, the sender's optimization problem (3) can be written as

$$\begin{gathered}\max_{\pi} \sum_{\omega \in \Omega} \sum_{(a,i) \in S} \pi(\omega, (a, i)) v(\omega, a) \\ \text{subject to, } a \in \arg\max_{a' \in A} \rho(\mu_{(a,i)}, a'), \quad \text{for all } (a, i) \in S, \\ \pi(\omega, S) = \mu^*(\omega), \quad \text{for all } \omega \in \Omega.\end{gathered}$$

where $\mu_{(a,i)}(\omega) = \frac{\pi(\omega,(a,i))}{\sum_{\omega'} \pi(\omega',(a,i))}$ denotes the receiver's belief after receiving the signal $(a, i) \in S$. Note that the constraint $a \in \arg\max_{a' \in A} \rho(\mu_{(a,i)}, a')$ is essentially an *obedience constraint*, which requires that the receiver, upon receiving the signal (a,i), finds it optimal to choose action $a \in A$. Using the definition of $\mathcal{P}_a$, this constraint can be equivalently written as $\mu_{(a,i)} \in \mathcal{P}_a$. Letting $t_a(\omega) = \sum_{i=1}^B \pi(\omega,(a,i))$ for each $a \in A$, the problem can be reformulated as

$$\begin{aligned}
\max_{\pi} & \sum_{\omega\in\Omega}\sum_{a\in A} t_a(\omega)v(\omega,a) \\
\text{subject to, } & \mu_{(a,i)} \in \mathcal{P}_a, \quad \text{for all } (a,i)\in S, \\
& \sum_{a\in A} t_a(\omega) = \mu^*(\omega), \quad \text{for all } \omega\in\Omega. \\
& t_a(\omega) = \sum_{i=1}^B \pi\left(\omega,(a,i)\right), \quad \text{for all } \omega\in\Omega \text{ and } a\in A.
\end{aligned}$$

Let $k(a,i) = \sum_{\omega'\in\Omega}\pi(\omega',(a,i))$. Note that $k(a,i)\geq 0$, and $\sum_{i=1}^B k(a,i)\leq 1$. Since $\mu_{(a,i)}(\omega) = \frac{\pi(\omega,(a,i))}{\sum_{\omega'}\pi(\omega',(a,i))} = \frac{\pi(\omega,(a,i))}{k(a,i)}$, we obtain for all $a\in A$ and $\omega\in\Omega$,

$$t_a(\omega) = \sum_{i=1}^B \pi\left(\omega,(a,i)\right) = \sum_{i=1}^B k(a,i)\mu_{(a,i)}(\omega) = \sum_{i=1}^B k(a,i)\mu_{(a,i)}(\omega) + \left(1-\sum_{i=1}^B k(a,i)\right)\mathbf{0}.$$

Since $\mu_{(a,i)} \in \mathcal{P}_a$ for all $1\leq i\leq B$, we deduce that $t_a \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$. Hence, the problem can be written as

$$\begin{aligned}
\max_{\pi} & \sum_{\omega\in\Omega}\sum_{a\in A} t_a(\omega)v(\omega,a) \\
\text{subject to, } & t_a \in \mathsf{Conv}(\mathcal{P}_a\cup\{\mathbf{0}\}), \quad \text{for all } a\in A, \\
& \sum_{a\in A} t_a(\omega) = \mu^*(\omega), \quad \text{for all } \omega\in\Omega. \\
& \mu_{(a,i)} \in \mathcal{P}_a, \quad \text{for all } (a,i)\in S, \\
& t_a(\omega) = \sum_{i=1}^B \pi(\omega,(a,i)), \quad \text{for all } \omega\in\Omega \text{ and } a\in A.
\end{aligned}$$

The formulation (7) then follows from an application of the Caratheodory's theorem, which implies that in the preceding optimization problem, the last two constraints are redundant (implied by the first constraint) as long as $B\geq|\Omega|$, and the optimization can be done directly over $\{t_a\}$ instead of π. The optimal signaling scheme π (along with the beliefs $\mu_{(a,i)}$) can then be obtained from the optimal t_a using (8), as described in the discussion following Theorem 1.

Appendix B: Proofs from Section 3

In this section, we provide the missing proofs from Section 3. Before we proceed, we recall the definition of a Bayes-plausible measure (Kamenica and Gentzkow 2011):

Definition EC.1. A measure $\eta\in\Delta(\Delta(\Omega))$ is Bayes-plausible for prior $\mu^*\in\Delta(\Omega)$ if $\mathbf{E}_\eta[\bar{s}] = \mu^*$, where $\bar{s}\in\Delta(\Omega)$ is distributed as η.

With this definition, we begin with the proof of Lemma 1.

Proof of Lemma 1. Consider a signaling scheme $\pi \in \Delta(\Omega \times \Delta(\Omega))$ satisfying $\mu_s = s$ for almost all $s \in \Delta(\Omega)$. Let $\eta(\cdot) \triangleq \pi(\Omega, \cdot) \in \Delta(\Delta(\Omega))$. By definition, we have for any $\omega \in \Omega$,

$$\mu^*(\omega) = \pi(\omega, \Delta(\Omega)) = \mathbf{E}_\pi[\mathbf{I}\{\bar{\omega} = \omega\}] = \mathbf{E}_\pi\left[\mathbf{E}_\pi[\mathbf{I}\{\bar{\omega} = \omega\}|\bar{s}]\right] = \mathbf{E}_\pi[\mu_{\bar{s}}(\omega)] = \mathbf{E}_\pi[\bar{s}(\omega)] = \mathbf{E}_\eta[\bar{s}(\omega)].$$

Thus, η is Bayes-plausible.

Next, let $\eta \in \Delta(\Delta(\Omega))$ be Bayes-plausible, and let $\pi(\omega, ds) \triangleq s(\omega)\eta(ds)$ for all $\omega \in \Omega$ and $s \in \Delta(\Omega)$. Observe that $\pi(\omega, \Delta(\Omega)) = \int_{\Delta(\Omega)} s(\omega)\eta(ds) = \mathbf{E}_\eta[\bar{s}(\omega)] = \mu^*(\omega)$, where the final equality follows from the Bayes-plausibility of η. Hence, π is a valid signaling scheme. Using Bayes' rule, for any $s \in \Delta(\Omega)$ (η-almost surely), we have $\mu_s(\omega) = \mathbf{P}_\pi(\bar{\omega} = \omega | \bar{s} = s) = \pi(\omega, ds)/\pi(\Omega, ds) = s(\omega)$ for all $\omega \in \Omega$. Thus, under π, the receiver's belief μ_s upon receiving signal $s \in \Delta(\Omega)$ satisfies $\mu_s = s$ almost surely. $\square$

LEMMA EC.1. *The sender's persuasion problem* (4) *is equivalent to the following optimization problem over Bayes-plausible measures η:*

$$\begin{aligned} &\max_{\eta \in \Delta(\Delta(\Omega))} \mathbf{E}_\eta\left[\sum_{\omega \in \Omega} \bar{s}(\omega) v(\omega, a(\bar{s}))\right] \\ &\textit{subject to,} \quad a(s) \in \arg\max_{a \in A} \rho(s, a), \quad \textit{for all } s \in \Delta(\Omega), \\ &\mathbf{E}_\eta[\bar{s}(\omega)] = \mu^*(\omega), \quad \textit{for each } \omega \in \Omega. \end{aligned} \tag{EC.2}$$

Proof. From Lemma 1, we obtain that for each Bayes-plausible η, there exists a corresponding signaling scheme $\pi \in \Delta(\Omega \times \Delta(\Omega))$ satisfying $\mu_s = s$ for almost all $s \in \Delta(\Omega)$, and conversely for each such scheme π, the corresponding measure $\eta(\cdot) = \pi(\Omega, \cdot)$ is Bayes-plausible. Furthermore, observe that for any Bayes-plausible measure η, the sender's expected utility under the corresponding signaling scheme $\pi(\omega, ds) = s(\omega)\eta(ds)$ can be written as

$$\begin{aligned} \mathbf{E}_\pi[v(\bar{\omega}, a(\bar{s}))] = \mathbf{E}_\pi\left[\mathbf{E}_\pi\left[v(\bar{\omega}, a(\bar{s}))\middle|\bar{s}\right]\right] &= \mathbf{E}_\pi\left[\sum_{\omega \in \Omega} \mu_{\bar{s}}(\omega) v(\omega, a(\bar{s}))\right] \\ &= \mathbf{E}_\pi\left[\sum_{\omega \in \Omega} \bar{s}(\omega) v(\omega, a(\bar{s}))\right] = \mathbf{E}_\eta\left[\sum_{\omega \in \Omega} \bar{s}(\omega) v(\omega, a(\bar{s}))\right], \end{aligned}$$

where the third equality follows from the fact that $\mu_s = s$ for almost all $s \in \Delta(\Omega)$, and the final equality follows from the definition of η. Since the sender's expected utility can be written as a function of the receiver's strategy and the probability measure η, we obtain the reformulation in the lemma statement. $\square$

Proof of Lemma 2. Fix any Bayes-plausible measure η and an optimal receiver strategy $a(\cdot)$. For each $a \in A$, define $b_a \triangleq \mathbf{P}_\eta(a(\bar{s}) = a) \in [0, 1]$ to be the probability that the receiver chooses action a under the corresponding signaling scheme. For each $a \in A$, if $b_a = 0$, then let η_a be any probability measure with support contained in $\mathcal{P}_a$. Otherwise, define η_a to the measure obtained by conditioning η on the event $a(\bar{s}) = a$. More precisely, we have $\eta_a(ds) \triangleq \frac{1}{b_a}\mathbf{I}\{a(s) = a\}\eta(ds)$ if $b_a > 0$. Note that, by the definition of the sets $\mathcal{P}_a$, the support of η_a is contained in $\mathcal{P}_a$ for each $a \in A$. The following equations are immediate from the definitions:

$$\sum_{a\in A} b_a \eta_a = \eta, \quad \sum_{a\in A} b_a = 1.$$

We let $m_a(\omega) \triangleq \mathbf{E}_{\eta_a}[\bar{s}(\omega)]$ for each $\omega \in \Omega$. Note that if $b_a > 0$, then $m_a(\omega) = \mathbf{E}_\eta[\bar{s}(\omega)|a(\bar{s}) = a]$. Thus, $m_a(\omega)$ is the *mean-posterior belief* of the receiver that the state is ω, given that she chooses the action a. From the Bayes-plausibility of η, we obtain for each $\omega \in \Omega$:

$$\sum_{a\in A} b_a m_a(\omega) = \sum_{a\in A} b_a \, \mathbf{E}_{\eta_a}\left[\bar{s}(\omega)\right] = \mathbf{E}_\eta\left[\bar{s}(\omega)\right] = \mu^*(\omega),$$

where the second equality follows from the fact that $\sum_{a\in A} b_a\eta_a = \eta$. Moreover, it is straightforward to verify that $m_a \in \mathsf{Conv}(\mathcal{P}_a)$, since m_a is the mean of the posterior distribution η_a with support contained in the closed set $\mathcal{P}_a$. Finally, note that for each $\omega \in \Omega$, we have

$$\begin{aligned}\sum_{a\in A} b_a m_a(\omega) v(\omega, a) &= \sum_{a\in A} b_a \, \mathbf{E}_{\eta_a}\left[\bar{s}(\omega)\right] v(\omega, a) \\ &= \sum_{a\in A} b_a \, \mathbf{E}_{\eta_a}\left[\bar{s}(\omega) v(\omega, a(\bar{s}))\right] \\ &= \mathbf{E}_\eta\left[\bar{s}(\omega) v(\omega, a(\bar{s}))\right],\end{aligned}$$

where the first equality uses the definition of m_a, the second equality follows from the fact that $a(\bar{s}) = a$ when $\bar{s} \sim \eta_a$, and the third equality follows from the fact that $\sum_{a\in A} b_a\eta_a = \eta$.

Conversely, suppose we have $\{(b_a, m_a) : b_a \in [0,1], m_a \in \mathsf{Conv}(\mathcal{P}_a)\}_{a\in A}$ with $\sum_{a\in A} b_a m_a = \mu^*$. By the definition of the convex hull, $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ implies the existence of $\{(\mu_i^a, \lambda_i^a) : i = 1, \dots, j_a\}$ such that $\mu_i^a \in \mathcal{P}_a$ and $\lambda_i^a \geq 0$ for each $i \leq j_a$ with $\sum_{i=1}^{j_a} \lambda_i^a = 1$ and $m_a = \sum_{i=1}^{j_a} \lambda_i^a \mu_i^a$. Define $\eta \in \Delta(\Delta(\Omega))$ to be the discrete distribution that selects the posterior μ_i^a with probability $b_a\lambda_i^a$. Then, we have for all $\omega \in \Omega$,

$$\mathbf{E}_\eta[\bar{s}(\omega)] = \sum_{a\in A}\sum_{i=1}^{j_a} b_a \lambda_i^a \mu_i^a(\omega) = \sum_{a\in A} b_a \left(\sum_{i=1}^{j_a} \lambda_i^a \mu_i^a(\omega)\right) = \sum_{a\in A} b_a m_a(\omega) = \mu^*(\omega).$$

This proves the Bayes-plausibility of η. Finally, define the strategy $a(\cdot) : \Delta(\Omega) \to A$ so that $a(\mu_i^a) = a$ for each $i \leq j_a$ and $a \in A$, and for other values of μ, let $a(\mu)$ be an arbitrary element in $\arg\max_{a\in A} \rho(\mu, a)$. Since $\mu_i^a \in \mathcal{P}_a$, it is straightforward to verify that the strategy $a(\cdot)$ is optimal. Finally, we have for each $\omega \in \Omega$,

$$\sum_{a\in A} b_a m_a(\omega) v(\omega, a) = \sum_{a\in A} b_a \sum_{i=1}^{j_a} \lambda_i^a \mu_i^a(\omega) v(\omega, a)$$

$$
\begin{aligned}
&= \sum_{a\in A}\sum_{i=1}^{j_a} (b_a\lambda_i^a)\cdot \mu_i^a(\omega)\cdot v(\omega, a(\mu_i^a)) \\
&= \sum_{a\in A}\sum_{i=1}^{j_a} \eta(\mu_i^a)\cdot \mu_i^a(\omega)\cdot v(\omega, a(\mu_i^a)) \\
&= \mathbf{E}_\eta\left[\bar{s}(\omega)v(\omega,\bar{s})\right].
\end{aligned}
$$

Here, the first equation follows from the fact that $m_a = \sum_{i=1}^{j_a}\mu_i^a\lambda_i^a$, the second equation follows from the fact that $a(\mu_i^a) = a$, and the third equation follows from the definition of η. This completes the proof of the lemma. □

Proof of Theorem 1. Note that for any Bayes-plausible $\eta \in \Delta(\Delta(\Omega))$, Lemma 2 guarantees a corresponding $\{(b_a, m_a)\}_{a\in A}$ with $b_a \in [0,1]$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ satisfying (5) and (6). Conversely, for each such $\{(b_a, m_a)\}_{a\in A}$, there exists a Bayes-plausible η. Thus, using Lemma EC.1, we can reframe the sender's persuasion problem (EC.2) as

$$
\begin{aligned}
\max_{\{b_a, m_a: a\in A\}} \quad & \sum_{\omega\in\Omega}\sum_{a\in A} b_a m_a(\omega) v(\omega, a) \\
\text{subject to,} \quad & \sum_{a\in A} b_a m_a = \mu^*, \\
& m_a \in \mathsf{Conv}(\mathcal{P}_a), \quad b_a \in [0,1] \quad \text{ for each } a \in A.
\end{aligned}
$$

Substituting $t_a = b_a m_a$ for each $a \in A$, we obtain that for each feasible $\{(b_a, m_a)\}_{a\in A}$, we have $t = \{t_a\}_{a\in A}$ feasible for (7), with equal objective values. Conversely, for any feasible $t = \{t_a\}_{a\in A}$, $t_a \in \mathsf{Conv}(\mathcal{P}_a \cup \{\mathbf{0}\})$ implies $t_a = b_a m_a + (1-b_a)\mathbf{0}$ for some $b_a \in [0,1]$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$. It follows that such $\{(b_a, m_a)\}_{a\in A}$ is feasible for the preceding program, with the same objective value. Thus, we obtain that the preceding program and (7) are equivalent. □

Proof of Proposition 1. Since the optimal m_a lies in the set $\mathsf{Conv}(\mathcal{P}_a)$, it follows that m_a can be written as a convex combination of at most $\mathsf{Cara}(\mathcal{P}_a)$ points in $\mathcal{P}_a$. As detailed in the discussion preceding the proposition statement, one can then construct an optimal signaling scheme using such a convex combination that sends, for each $a \in A$, at most $\mathsf{Cara}(\mathcal{P}_a)$ signals for which the receiver's optimal action is a. Hence, the total number of signals is at most $\sum_{a\in A}\mathsf{Cara}(\mathcal{P}_a)$.

Using Caratheodory's theorem, we have $\mathsf{Cara}(\mathcal{P}_a) \le \dim(\mathcal{P}_a) + 1$, where $\dim(H)$ is the dimension of the smallest affine space containing H. Since the set $\mathcal{P}_a \subseteq \Delta(\Omega)$ lies in an affine space of dimension $\mathbb{R}^{|\Omega|-1}$, we obtain $\mathsf{Cara}(\mathcal{P}_a) \le |\Omega|$. □

Appendix C: Proofs from Section 4

Proof of Proposition 2. We begin by showing that the two parts of the proposition statement are equivalent, and hence it suffices to prove the second part.

Suppose there is information the sender would share, i.e, there exists a belief μ with $\hat{\nu}(\mu) > \mathbf{E}_\mu[v(\bar{\omega}, a_{\mu^*}]$. Then, for $a = a_\mu$ and $m_a = \mu$, we have $m_a \in \mathcal{P}_a \subseteq \mathsf{Conv}(\mathcal{P}_a)$, and

$\sum_{\omega\in\Omega} m_a(\omega)\left(v(\omega,a)-v(\omega,a_{\mu^*})\right) = \mathbf{E}_\mu[v(\bar\omega,a_\mu)-v(\bar\omega,a_{\mu^*})] > 0$. Conversely, suppose there is no information the sender would share. Then, for any $\mu\in\mathcal{P}_a$, we have $\mathbf{E}_\mu[v(\bar\omega,a)] = \hat\nu(\mu) \le \mathbf{E}_\mu[v(\bar\omega,a_{\mu^*})]$. Using linearity of expectation, we obtain $\mathbf{E}_\mu[v(\bar\omega,a)] \le \mathbf{E}_\mu[v(\bar\omega,a_{\mu^*})]$ for all $\mu\in\mathsf{Conv}(\mathcal{P}_a)$, implying that $\sum_{\omega\in\Omega} m_a(\omega)\left(v(\omega,a)-v(\omega,a_{\mu^*})\right)\le 0$ for all $m_a\in\mathsf{Conv}(\mathcal{P}_a)$ and $a\in A$. Thus, the two parts of the proposition statement are equivalent, and we next prove the second part.

Suppose for all $a\in A$ and $m_a\in\mathsf{Conv}(\mathcal{P}_a)$, we have $\sum_{\omega\in\Omega} m_a(\omega)\left(v(\omega,a)-v(\omega,a_{\mu^*})\right)\le 0$. Then, for any feasible $t=\{t_a\}_{a\in A}$, let $b_a\in[0,1]$ and $m_a\in\mathsf{Conv}(\mathcal{P}_a)$ be such that $t_a = b_a m_a$. We have

$$\begin{aligned}
\sum_{a\in A}\sum_{\omega\in\Omega} t_a(\omega)v(\omega,a) &= \sum_{a\in A} b_a\left(\sum_{\omega\in\Omega} m_a(\omega)v(\omega,a)\right)\\
&\le \sum_{a\in A} b_a\left(\sum_{\omega\in\Omega} m_a(\omega)v(\omega,a_{\mu^*})\right)\\
&= \sum_{\omega\in\Omega}\mu^*(\omega)v(\omega,a_{\mu^*})\\
&= \hat\nu(\mu^*),
\end{aligned}$$

where in the penultimate equality we use $\mu^* = \sum_{a\in A} b_a m_a$. Since this holds for all feasible t, we obtain $V(\mu^*)=\hat\nu(\mu^*)$, and hence the sender does not benefit from persuasion.

Next, suppose there exists an $a\in A$ and $m_a\in\mathsf{Conv}(\mathcal{P}_a)$ with $\sum_{\omega\in\Omega} m_a(\omega)\left(v(\omega,a)-v(\omega,a_{\mu^*})\right)>0$. If μ^* lies in the interior of $\mathcal{P}_{\mu^*}$, then there exists a δ-ball B_δ around μ^* contained in $\mathcal{P}_{\mu^*}$. Let $\mu'\in B_\delta$ be such that $\mu^*=\gamma\mu'+(1-\gamma)m_a$ for some $\gamma\in(0,1)$. Then, let $t_{a'}=\gamma\mu'$ for $a'=a_{\mu^*}$, let $t_a=(1-\gamma)m_a$, and let $t_{a'}=\mathbf{0}$ for $a'\neq a_{\mu^*}, a$. It follows that $t=\{t_{a'}\}_{a'\in A}$ is feasible for (7). We have

$$\begin{aligned}
V(\mu^*) &\ge \sum_{\omega\in\Omega}\sum_{a'\in A} t_{a'}(\omega)v(\omega,a')\\
&= \gamma\sum_{\omega\in\Omega}\mu'(\omega)v(\omega,a_{\mu^*}) + (1-\gamma)\sum_{\omega\in\Omega} m_a(\omega)v(\omega,a)\\
&> \gamma\sum_{\omega\in\Omega}\mu'(\omega)v(\omega,a_{\mu^*}) + (1-\gamma)\sum_{\omega\in\Omega} m_a(\omega)v(\omega,a_{\mu*})\\
&= \sum_{\omega\in\Omega}\mu^*(\omega)v(\omega,a)\\
&= \hat\nu(\mu^*).
\end{aligned}$$

Thus, the sender (strictly) benefits from persuasion. □

Appendix D: Proofs from Section 5

In this section, we provide the proofs of the results in Section 5. We begin with the following simple argument showing that the quasiconvexity of $\hat\rho(\cdot)$ implies Assumption 2.

Proof of Lemma 3. The proof follows from the fact that if $\hat\rho$ is quasiconvex, then $\hat\rho(\gamma\mu+(1-\gamma)\mu')\le\max\{\hat\rho(\mu),\hat\rho(\mu')\}$ for $\gamma\in[0,1]$ and $\mu,\mu'\in\Delta(\Omega)$. □

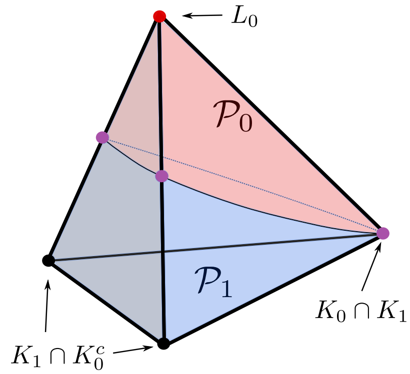


Figure EC.1 The sets L_0, K_0 and K_1 when $K_0 \neq L_0$. Here, K_{01} is the set of the purple vertices.

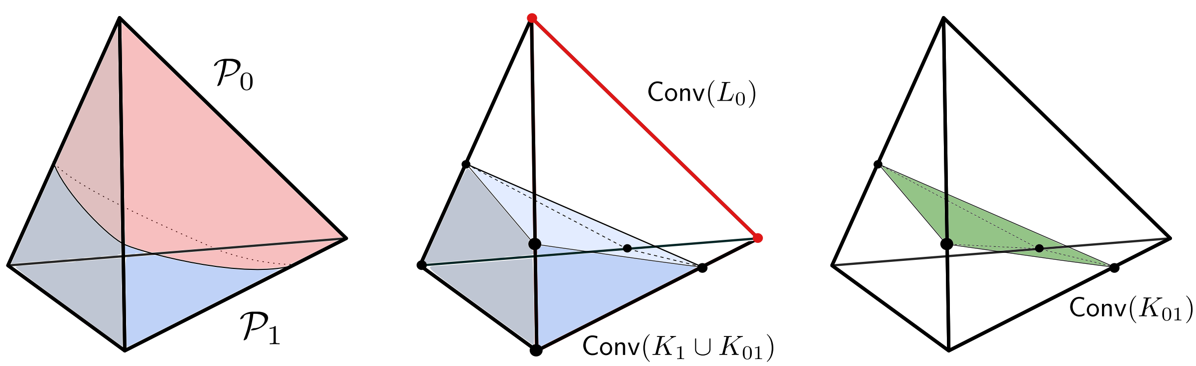


Figure EC.2 Geometry of $\Delta(\Omega)$ when $\mathsf{Conv}(K_{01})$ is not a hyperplane. Here, $\mathsf{Conv}(L_0)$ is an edge of the simplex, shown in red.

We next focus on the proof of Theorem 3. The proof of the theorem rests on two helper lemmas that characterize the geometry of the sets $\Delta(\Omega)$ and $\mathsf{Conv}(\mathcal{P}_1)$. The first lemma shows that the set $\Delta(\Omega)$ can be viewed as the union of two regions, each of which is the convex hull of a finite set of points. Fig. 4 and Fig. EC.2 illustrate the geometric intuition behind this lemma.

Recall that $L_0 \triangleq K_0 \cap K_1^c$ denotes the set of states for which action 0 is uniquely optimal for the receiver.

Lemma EC.2. *$\Delta(\Omega) = \mathsf{Conv}(L_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$.*

Proof. Let $\mu \in \Delta(\Omega)$. Since $\Omega = L_0 \cup K_1$, we have $\mu \in \mathsf{Conv}(L_0 \cup K_1) = \mathsf{Conv}(L_0 \cup K_1 \cup K_{01})$. Consider any convex decomposition of $\mu = \sum_{\phi \in L_0 \cup K_1 \cup K_{01}} \alpha_\phi \phi$. If the convex decomposition does not place positive weights either on the elements of L_0 or on those of $K_1 \setminus K_{01}$, then we are done. Otherwise, let $\omega_0 \in L_0$ and $\omega_1 \in K_1 \setminus K_{01}$ be such that $\alpha_\phi > 0$ for $\phi \in \{\omega_0, \omega_1\}$. Let $\mu' \triangleq \frac{1}{\alpha_{\omega_0} + \alpha_{\omega_1}}(\alpha_{\omega_0}\omega_0 + \alpha_{\omega_1}\omega_1)$, and note that $\mu = (\alpha_{\omega_0} + \alpha_{\omega_1})\mu' + \sum_{\phi \in L_0 \setminus \{\omega_0\}} \alpha_\phi \phi + \sum_{\phi \in K_{01} \cup K_1 \setminus \{\omega_1\}} \alpha_\phi \phi$.

Since $\mu' \in \mathsf{Conv}(\{\omega_0, \omega_1\})$, one can write μ' as either a convex combination of ω_0 and $\chi(\omega_0, \omega_1)$ or a convex combination of ω_1 and $\chi(\omega_0, \omega_1)$. In either scenario, using this decomposition for μ', we obtain a convex decomposition of μ that places positive weight on at least one fewer element of $L_0 \cup (K_1 \setminus K_{01})$. Continuing with this process, we obtain a convex decomposition of μ that places no positive weight either on elements on L_0 or on elements of $K_1 \setminus K_{01}$, yielding the lemma statement. □

The second lemma uses this result to establish that $\mathsf{Conv}(\mathcal{P}_1)$ is a convex polytope with extreme points in the set $K_1 \cup K_{01}$.

LEMMA EC.3. *Under Assumption 2,* $\mathsf{Conv}(\mathcal{P}_1) = \mathsf{Conv}(K_1 \cup K_{01})$.

Proof. We have $K_1 \subseteq \mathcal{P}_1$, and furthermore, as $\mathcal{P}_1$ is closed, we have $K_{01} \subseteq \mathcal{P}_1$. Hence, we obtain that $\mathsf{Conv}(K_1 \cup K_{01}) \subseteq \mathsf{Conv}(\mathcal{P}_1)$. Thus, to prove the lemma statement, we must show $\mathsf{Conv}(\mathcal{P}_1) \subseteq \mathsf{Conv}(K_1 \cup K_{01})$. Morever, since $\mathsf{Conv}(\mathcal{P}_1)$ is the smallest convex set containing $\mathcal{P}_1$, it suffices to show $\mathcal{P}_1 \subseteq \mathsf{Conv}(K_1 \cup K_{01})$.

Let $\mu \in \mathcal{P}_1$. By Lemma EC.2, $\mu \in \mathsf{Conv}(L_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$. Suppose for the sake of contradition, $\mu \notin \mathsf{Conv}(K_1 \cup K_{01})$. Then, it follows that $\mu \in \mathsf{Conv}(L_0 \cup K_{01})$ and $\mu \notin \mathsf{Conv}(K_{01})$. Consider any convex decomposition of μ as follows:

$$\mu = \sum_{\omega \in L_0} \alpha^0_\omega \omega + \sum_{\omega \in K_{01} \cap K_1} \alpha^0_\omega \omega + \sum_{\phi \in K_{01} \cap K_1^c} \alpha^0_\phi \phi.$$

Define $H^0_0 = \{\omega \in L_0 : \alpha^0_\omega > 0\}$, $H^1_0 = \{\omega \in K_{01} \cap K_1 : \alpha^0_\omega > 0\}$, and $H^2_0 = \{\phi \in K_{01} \cap K_1^c : \alpha^0_\phi > 0\}$. Since $\mu \notin \mathsf{Conv}(K_{01})$, it follows that $\sum_{\omega \in L_0} \alpha^0_\omega > 0$. Thus, we obtain that H^0_0 is non-empty.

In the following, we inductively define sets $\{H^i_n : i = 0, 1, 2\}$ for $n \geq 1$ as long as both H^0_{n-1} and H^1_{n-1} are non-empty, such that the following four properties hold: (1) $H^0_n \subseteq L_0$; (2) $H^1_n \subseteq K_{01} \cap K_1$; (3) for any $\chi(\omega_0, \omega_1) \in H^2_n \subseteq K_{01}$, either $\chi(\omega_0, \omega_1) \in K_{01} \cap K_1^c$ or $\omega_0 \in H^0_n$; and (4) μ is a strict convex combination of elements in $H^0_n \cup H^1_n \cup H^2_n$. Towards that end, first note that the sets $\{H^i_0 : i = 0, 1, 2\}$ satisfy the aforementioned four properties.

Next, suppose for some $n \geq 1$ the sets $\{H^i_{n-1} : i = 0, 1, 2\}$ satisfy the four properties, with both H^0_{n-1} and H^1_{n-1} being non-empty. Consider a strict convex decomposition of μ in terms of elements in $\cup_{i=0}^2 H^i_{n-1}$ with coefficients $\alpha^{n-1}_\phi > 0$ for $\phi \in \cup_{i=0}^2 H^i_{n-1}$. Choose some $\omega_0 \in H^0_{n-1}$ and $\omega_1 \in H^1_{n-1}$, and let $\beta_{n-1} = \alpha^{n-1}_{\omega_0} / (\alpha^{n-1}_{\omega_0} + \alpha^{n-1}_{\omega_1})$. Observe that $\beta_{n-1}\omega_0 + (1 - \beta_{n-1})\omega_1$ either (1) equals $\chi(\omega_0, \omega_1)$; or (2) is a strict convex combination of ω_0 and $\chi(\omega_0, \omega_1)$; or (3) is a strict convex combination of $\chi(\omega_0, \omega_1)$ and ω_1. We split the analysis into the three respective cases:

1. If $\beta_{n-1}\omega_0 + (1 - \beta_{n-1})\omega_1$ equals $\chi(\omega_0, \omega_1)$, let $H^i_n = H^i_{n-1} \setminus \{\omega_i\}$ for $i \in \{0, 1\}$ and $H^2_n = H^2_{n-1} \cup \{\chi(\omega_0, \omega_1)\}$. Since $\alpha^{n-1}_{\omega_i} > 0$ for $i \in \{0, 1\}$, we have $\beta_{n-1} \in (0, 1)$ and hence $\chi(\omega_0, \omega_1) \in K_{01} \cap K_1^c$. Finally, letting $\alpha^n_\phi = \alpha^{n-1}_\phi$ for all $\phi \in \cup_{i=0}^2 H^i_n \setminus \{\chi(\omega_0, \omega_1)\}$ and $\alpha^n_\phi = (\alpha^{n-1}_\phi + \alpha^{n-1}_{\omega_0} + \alpha^{n-1}_{\omega_1}) > 0$

for $\phi=\chi(\omega_0,\omega_1)$, we obtain a strict convex combination of μ in terms of elements of $\cup_{i=0}^2 H_n^i$. Thus, the four properties continue to hold for $\{H_n^i : i=0,1,2\}$.

2. If $\beta_{n-1}\omega_0+(1-\beta_{n-1})\omega_1$ is a strict convex combination of ω_0 and $\chi(\omega_0,\omega_1)$, let $H_n^0=H_{n-1}^0$, $H_n^1=H_{n-1}^1\setminus\{\omega_1\}$ and $H_n^2=H_{n-1}^2\cup\{\chi(\omega_0,\omega_1)\}$. Note that properties (1) and (2) hold trivially, and since $\omega_0\in H_n^0$, property (3) continues to hold. Using the strict convex combination of $\beta_{n-1}\omega_0+(1-\beta_{n-1})\omega_1$ in terms of ω_0 and $\chi(\omega_0,\omega_1)$, we obtain a strict convex combination of μ in terms of elements of $\cup_{i=0}^2 H_n^i$, and hence property (4) also holds.
3. If $\beta_{n-1}\omega_0+(1-\beta_{n-1})\omega_1$ is a strict convex combination of $\chi(\omega_0,\omega_1)$ and ω_1, let $H_n^0=H_{n-1}^0\setminus\{\omega_0\}$, $H_n^1=H_{n-1}^1$ and $H_n^2=H_{n-1}^2\cup\{\chi(\omega_0,\omega_1)\}$. Again, properties (1) and (2) hold trivially. Since $\beta_{n-1}\omega_0+(1-\beta_{n-1})\omega_1$ with $\beta_{n-1}\in(0,1)$ is a strict convex combination of $\chi(\omega_0,\omega_1)$ and ω_1, it follows that $\chi(\omega_0,\omega_1)\neq\omega_1$ and hence $\chi(\omega_0,\omega_1)\in K_{01}\cap K_1^c$. Thus property (3) holds. Finally, using the strict convex combination of $\beta_{n-1}\omega_0+(1-\beta_{n-1})\omega_1$ in terms of $\chi(\omega_0,\omega_1)$ and ω_1, we obtain a strict convex combination of μ in terms of elements of $\cup_{i=0}^2 H_n^i$, and hence property (4) also holds.

Note that in all three cases, we have $|H_n^0|+|H_n^1|<|H_{n-1}^0|+|H_{n-1}^1|$. Thus, this inductive process stops with sets $H_n^i, i=0,1,2$ for some $n\geq 0$ with either H_n^0 or H_n^1 empty. If H_n^0 is empty, then $\mu\in\mathsf{Conv}(H_n^1\cup H_n^2)\subseteq\mathsf{Conv}(K_{01})$, contradicting the assumption that $\mu\notin\mathsf{Conv}(K_{01}\cup K_1)$. Thus, it must be that H_n^1 is empty. Furthermore, consider any $\phi=\chi(\omega_0,\omega_1)\in H_n^2$ for which $\omega_0\notin H_n^0$. By property (3), it must be that $\chi(\omega_0,\omega_1)\in K_{01}\cap K_1^c$. Choose any $\omega'\in H_n^0$, and define $\beta=\alpha_{\omega'}^n/(\alpha_{\omega'}^n+\alpha_\phi^n)$. Since $\chi(\omega_0,\omega_1)\in K_{01}\cap K_1^c$ and $\omega'\in H_n^0$, it is straightforward to verify that $\beta\omega'+(1-\beta)\chi(\omega_0,\omega_1)$ can be written as a strict convex combination of ω_0, ω', $\chi(\omega_0,\omega_1)$ and $\chi(\omega',\omega_1)$. Using such a strict convex combination and adding $\omega_0\in L_0$ to the set H_n^0 (and if needed, adding $\chi(\omega',\omega_1)$ to H_n^2), we can without loss of generality assume that for any $\phi=\chi(\omega_0,\omega_1)\in H_n^2$ we have $\omega_0\in H_n^0$. We thus obtain the following strict convex decomposition:

$$\mu=\sum_{\omega\in H_n^0}\gamma_\omega\omega+\sum_{\phi\in H_n^2}\gamma_\phi\phi,$$

where $\sum_{\omega\in H_n^0}\gamma_\omega>0$, and for which if $\chi(\omega_0,\omega_1)\in H_n^2$ then $\omega_0\in H_n^0$.

This can be further rewritten as the following convex combination

$$\mu=\sum_{\omega\in H_n^0}\hat{\gamma}_\omega\omega+\sum_{\chi(\omega_0,\omega_1)\in H_n^2}(\tilde{\gamma}_{\omega_0}\omega_0+\tilde{\gamma}_{\chi(\omega_0,\omega_1)}\chi(\omega_0,\omega_1)),$$

where if $\tilde{\gamma}_{\chi(\omega_0,\omega_1)}>0$, then $\tilde{\gamma}_{\omega_0}>0$. Note that for any such (ω_0,ω_1), by definition of $\chi(\omega_0,\omega_1)$, the belief $\xi(\omega_0,\omega_1)\triangleq\frac{1}{(\tilde{\gamma}_{\omega_0}+\tilde{\gamma}_{\chi(\omega_0,\omega_1)})}\left(\tilde{\gamma}_{\omega_0}\omega_0+\tilde{\gamma}_{\chi(\omega_0,\omega_1)}\chi(\omega_0,\omega_1)\right)$ lies in the set $\mathcal{P}_1^c$. Thus, μ is a convex combination of elements in $H_n^0\subseteq L_0$ and the elements $\{\xi(\omega_0,\omega_1)\}$, all of which belong to $\mathcal{P}_1^c$. From Assumption 2, we then obtain that μ itself is an element of $\mathcal{P}_1^c$, contradicting the fact that

$\mu \in \mathcal{P}_1$. This proves that our initial assumption that $\mu \notin \mathsf{Conv}(K_1 \cup K_{01})$ must be false, and hence $\mu \in \mathsf{Conv}(K_1 \cup K_{01})$.

Thus, $\mu \in \mathcal{P}_1$ implies $\mu \in \mathsf{Conv}(K_1 \cup K_{01})$, and hence $\mathcal{P}_1 \subseteq \mathsf{Conv}(K_1 \cup K_{01})$. Thus, we conclude that $\mathsf{Conv}(\mathcal{P}_1) = \mathsf{Conv}(K_1 \cup K_{01})$. $\square$

With these two helper lemmas in place, we are now ready to prove Theorem 3.

Proof of Theorem 3. Since the objectives of the programs (7) and (10) are identical, to prove the result it suffices to show that optimal solution of each program is a feasible for the other.

First, consider any optimal solution to (10). Since $\mathsf{Conv}(L_0 \cup \{\mathbf{0}\}) \subseteq \mathsf{Conv}(\mathcal{P}_0 \cup \{\mathbf{0}\})$ and $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\}) \subseteq \mathsf{Conv}(\mathcal{P}_1 \cup \{\mathbf{0}\})$, it is a feasible solution to (7).

Next, consider any optimal solution $t = (t_0, t_1)$ to (7) with $t_1 \in \mathsf{Conv}(\mathcal{P}_1 \cup \{\mathbf{0}\})$ and $t_0 \in \mathsf{Conv}(\mathcal{P}_0 \cup \{\mathbf{0}\})$. By Lemma EC.3, it follows that $t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$. If $t_0 = \mathbf{0}$, then $t_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$ and we are done. Instead, suppose $t_0 \neq \mathbf{0}$. Let $t_0 = b_0 m_0$ where $b_0 \in (0,1]$ and $m_0 \in \mathsf{Conv}(\mathcal{P}_0)$. By Lemma EC.2, we obtain $m_0 \in \mathsf{Conv}(L_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$. If $m_0 \in \mathsf{Conv}(K_1 \cup K_{01})$, then the solution $\hat{t} = (\hat{t}_0, \hat{t}_1)$ with $\hat{t}_0 = 0$ and $\hat{t}_1 = t_1 + t_0$ is feasible for (7) and achieves larger utility for the sender than (t_0, t_1), contradicting the latter's optimality. Hence, $m_0 \in \mathsf{Conv}(L_0 \cup K_{01})$. If $m_0 \notin \mathsf{Conv}(L_0)$, then $m_0 = (1-\alpha)\nu_0 + \alpha\nu_1$ with $\nu_0 \in \mathsf{Conv}(L_0)$ and $\nu_1 \in \mathsf{Conv}(K_{01})$ and $\alpha > 0$. However, we then have that $\tilde{t} = (\tilde{t}_0, \tilde{t}_1)$ with $\tilde{t}_0 = b_0(1-\alpha)\nu_0$ and $\tilde{t}_1 = t_1 + b_0\alpha\nu_1$ is feasible for (7) and has larger utility for the sender than (t_0, t_1), once again contradicting the latter's optimality. Thus, we must have $m_0 \in \mathsf{Conv}(L_0)$ and hence $t_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$. Taken together, we obtain that (t_0, t_1) is feasible for (10). $\square$

Finally, we conclude this section with the proof of Proposition 3.

Proof of Proposition 3. First, note that if $v(\omega,1) - v(\omega,0) = v > 0$ for all $\omega \in \Omega$, then for any feasible $t = (t_0, t_1)$ for (10), we have $\sum_{\omega\in\Omega} v(\omega,1)t_1(\omega) + v(\omega,0)t_0(\omega) = \sum_{\omega\in\Omega} v(\omega,0)\mu^*(\omega) + v\sum_{\omega\in\Omega} t_1(\omega)$. Thus, it suffices to show that for any $t = (t_0, t_1)$ that does not have the threshold structure, one can find another feasible $\tilde{t} = (\tilde{t}_1, \tilde{t}_0)$ that satisfies $\sum_{\omega\in\Omega} t_1(\omega) < \sum_{\omega\in\Omega} \tilde{t}_1(\omega)$.

Toward that end, let $t = (t_0, t_1)$ be a feasible solution to (10) such that there exists $\omega^\dagger, \omega' \in \Omega$ satisfying $\omega^\dagger \prec \omega'$ with $t_1(\omega^\dagger) < \mu^*(\omega^\dagger)$ and $t_1(\omega') > 0$. Since $t_1 + t_0 = \mu^*$, we obtain $t_0(\omega^\dagger) > 0$. As $t_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$, this implies that $\omega^\dagger \in L_0$. By Assumption 3, this implies that $\omega' \in L_0$ as well.

Now, $t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$ implies that

$$t_1 = \sum_{j\in K_1} \alpha_j e_j + \sum_{\phi \in K_{01}\setminus K_1} \alpha_\phi \phi,$$

with $\alpha_j, \alpha_\phi \geq 0$ and $\sum_{j\in K_1} \alpha_j + \sum_{\phi\in K_{01}\setminus K_1} \alpha_\phi \leq 1$. Since $t_1(\omega') > 0$, we deduce that there exists an $\hat{\omega} \in K_1$ such that $\alpha_\phi > 0$ for $\phi = \chi(\omega', \hat{\omega})$. For fixed small enough $\epsilon > 0$, define $\tilde{t} = (\tilde{t}_0, \tilde{t}_1)$ as

$$\tilde{t}_1 = t_1 + \epsilon\chi(\omega^\dagger, \hat{\omega}) - \epsilon'\chi(\omega', \hat{\omega})$$
$$\tilde{t}_0 = t_0 - \epsilon\chi(\omega^\dagger, \hat{\omega}) + \epsilon'\chi(\omega', \hat{\omega}),$$

where $\epsilon' > 0$ satisfies $\epsilon'(1-\gamma(\omega',\hat{\omega})) = \epsilon(1-\gamma(\omega^\dagger,\hat{\omega}))$. Clearly, we have $\tilde{t}_1 + \tilde{t}_0 = t_1 + t_0 = \mu^*$. Furthermore, for small enough $\epsilon > 0$, we have $\tilde{t}_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$. Finally, using the definition of $\chi(\omega^\dagger,\hat{\omega})$ and $\chi(\omega',\hat{\omega})$ and the choice of ϵ', we have

$$\tilde{t}_0 = t_0 - \epsilon\gamma(\omega^\dagger,\hat{\omega})e_{\omega^\dagger} + \epsilon'\gamma(\omega',\hat{\omega})e_{\omega'}.$$

Since $t_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$ with $t_0(\omega^\dagger) > 0$, and $\omega' \in L_0$, we obtain that $\tilde{t}_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$ for small enough $\epsilon > 0$. Thus, putting it all together, for small enough $\epsilon > 0$, $\tilde{t}$ is feasible for (10).

Finally, we have

$$\begin{aligned}
\sum_{\omega\in\Omega} \tilde{t}_1(\omega) &= \sum_{\omega\in\Omega} t_1(\omega) + \epsilon\gamma(\omega^\dagger,\hat{\omega}) - \epsilon'\gamma(\omega',\hat{\omega}) \\
&= \sum_{\omega\in\Omega} t_1(\omega) + \epsilon(1-\gamma(\omega^\dagger,\hat{\omega}))\left(\frac{\gamma(\omega^\dagger,\hat{\omega})}{1-\gamma(\omega^\dagger,\hat{\omega})} - \frac{\gamma(\omega',\hat{\omega})}{1-\gamma(\omega',\hat{\omega})}\right) \\
&> \sum_{\omega\in\Omega} t_1(\omega),
\end{aligned}$$

where we have used the fact that since $\omega^\dagger \prec \omega'$, we have $\gamma(\omega^\dagger,\hat{\omega}) > \gamma(\omega',\hat{\omega})$. $\square$

Appendix E: Proofs from Section 7

In this section, we provide the proofs of the results from Section 7. To begin, let $\Omega = \{0, 1, \ldots, C-1\}$ denote the set of queue-lengths for which an arriving customer could possibly join the queue. (We assume that immediately upon arrival a customer is informed whether the queue-length equals C or not.) Let X denote the waiting time until service completion faced by an arriving customer if she decides to join the queue. Under the belief e_n, an arriving customer knows that there are exactly n customers already in queue, and thus X has the distribution of the sum of $n+1$ independent unit exponential random variables.

Recall the definition (11) of the differential utility function: $\hat{\rho}(\mu) = \tau - \left(\mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu[X]}\right)$ for some $\tau > 0$ and $\beta \geq 0$. It is straightforward to verify that $\hat{\rho}(e_n) = \tau - \left(n+1+\beta\sqrt{n+1}\right)$, and hence $\hat{\rho}(e_n)$ is strictly decreasing in n. As $K_1 = \{n \in \Omega : \hat{\rho}(e_n) \geq 0\}$, and $L_0 = \{n \in \Omega : \hat{\rho}(e_n) < 0\}$, we obtain that $K_1 = \{0, 1, \ldots, M\}$ and $L_0 = \{n \in \Omega : n > M\}$ for some $M < C$.

For $n, m \in \Omega$ and $\gamma \in [0,1]$, the belief $\mu = \gamma e_n + (1-\gamma)e_m$ assigns probability γ to the state with n customers already in queue, and probability $1-\gamma$ to the state with m customers already in queue; consequently, under the belief μ, the distribution of X is a convex mixture (with weight γ) of its distribution under beliefs e_n and e_m. Thus, we obtain $\mathbf{E}_\mu[X] = 1 + m + (n-m)\gamma$ and $\mathrm{Var}_\mu[X] = 1 + m + (n-m)\gamma + (n-m)^2\gamma(1-\gamma)$. Finally, for $n \in L_0$ and $m \in K_1$, $\gamma(n,m) = \sup\{\gamma \in [0,1] : \hat{\rho}(\gamma e_n + (1-\gamma)e_m) \geq 0\}$ is the largest value of $\gamma \in [0,1]$ for which $\gamma e_n + (1-\gamma)e_m \in \mathcal{P}_1$, and $\chi(n,m) = \gamma(n,m)e_n + (1-\gamma(n,m))e_m$ is the corresponding belief. Note that under the belief

$\chi(n,m)$, the customer is indifferent between joining and leaving the queue. Using the expression for $\hat{\rho}$ and after some straightforward algebra, we obtain for $n \in L_0$ and $m \in K_1$,

$$\gamma(n,m) = \frac{2(\tau - 1 - m) + \beta^2(n-m+1) - \beta\sqrt{h(n,m)}}{2(n-m)(1+\beta^2)}, \tag{EC.3}$$

where $h(n,m) \triangleq \beta^2(n-m+1)^2 + 4(\tau-1-m)(n-m+1) + 4(1+\beta^2)(1+m) - 4(\tau-1-m)^2$.

The following result implies that differential utility function $\hat{\rho}(\cdot)$ defined in (11) is convex and monotone (Assumptions 2 and 3).

LEMMA EC.4. *For $\tau > 0, \beta \geq 0$, let $\hat{\rho}(\mu) = \tau - \left(\mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu[X]}\right)$. Then,*

1. *The differential utility function $\hat{\rho}(\mu)$ is convex in μ, and hence satisfies Assumption 2.*
2. *Under the usual total order on Ω, $\hat{\rho}(\mu)$ satisfies the monotonicity condition in Assumption 3.*

Proof. From the fact that expectation is linear in the belief, it is straightforward to show that $\mathrm{Var}_\mu[X] = \mathbf{E}_\mu(X^2) - (\mathbf{E}_\mu(X))^2$ is concave in the belief μ. Since $\sqrt{x}$ is concave and strictly increasing in x, using Jensen's inequality, we obtain that $\sqrt{\mathrm{Var}_\mu[X]}$ is also concave in μ. From these facts, we conclude that $\hat{\rho}(\mu) = \tau - \left(\mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu[X]}\right)$ is convex in μ. Thus, we obtain that the set $\mathcal{P}_a^c = \{\mu : \hat{\rho}(\mu) < 0\}$ is convex, and hence $\hat{\rho}(\cdot)$ satisfies Assumption 2.

Since $K_1 = \{0, 1, \ldots, M\}$ and $L_0 = \{n \in \Omega : n > M\}$, if $m \in K_1$ then $m < n$ for all $n \in L_0$. Next, let $m \in K_1$ and $n, \ell \in L_0$ with $n < \ell$. For $\gamma \in (0,1]$, let $\mu = \gamma e_n + (1-\gamma)e_m$ and $\mu' = \gamma e_\ell + (1-\gamma)e_m$. Since $\mathbf{E}_\mu[X] < \mathbf{E}_{\mu'}[X]$ and $\mathrm{Var}_\mu[X] < \mathrm{Var}_{\mu'}[X]$, we obtain that $\hat{\rho}(\mu) > \hat{\rho}(\mu')$ for all $\gamma \in (0,1]$. Thus, we obtain that $\gamma(n,m) > \gamma(\ell,m)$. To summarize, the usual order on Ω satisfies the property that if $n < \ell$, then either $n \in K_1$ or for all $m \in K_1$, we have $\gamma(n,m) > \gamma(\ell,m)$. Thus, we obtain that $\hat{\rho}(\cdot)$ satisfies the monotonicity condition (Assumption 3). □

E.1. Proof of Lemma 4

Our proof follows a similar approach as in Lingenbrink and Iyer (2019): we show that any feasible solution where the customers' actions do not have a threshold structure can be perturbed to obtain another feasible solution corresponding to a signaling scheme with higher throughput.

Proof of Lemma 4. First, note that if $K_1 = \Omega$. i.e., $M = C-1$, then joining the queue is always optimal. In this case, the full-information mechanism (a mechanism inducing a threshold structure on the actions) has the highest throughput. Hence, hereafter we focus on the case where K_1 is a strict subset of Ω, i.e, $M < C-1$.

Consider a feasible solution $t = (t_1, t_0)$ to the linear program (12). Since $t_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$ and $L_0 = \{M+1, \ldots, C-1\}$, we have $t_0(k) = 0$ for all $k \leq M$. Then the constraint (12c) implies that $t_1(k+1) = \lambda t_1(k)$ for all $k < M$, and hence $t_1(k) = \lambda^k t_1(0)$ for all $k \leq M$.

Let N be the largest value of n such that $t_1(k) = \lambda^k t_1(0)$ for all $k \leq n$. (Note the preceding argument implies $N \geq M$.) If either $N \geq C-2$ or $t_1(N+2) = 0$, then we obtain that t induces a threshold structure in the customers' actions. Hence, for the rest of the proof assume that $N < C-2$ and $t_1(N+2) > 0$. Using (12c), we then obtain $0 < t_1(N+1) < \lambda t_1(N)$, $t_0(N+1) > 0$ and $t_0(n) = 0$ for $n \leq N$. We now show that $t = (t_1, t_0)$ cannot be an optimal solution to (12). We do this by constructing another feasible solution $\tilde{t} = (\tilde{t}_1, \tilde{t}_0)$ that has a larger objective value than t.

For some small $\epsilon > 0$ and $\beta > 0$ to be chosen later, consider $\tilde{t} = (\tilde{t}_1, \tilde{t}_0)$ with

$$\tilde{t}_1 = \frac{1}{Z}\left(t_1 - \beta \sum_{m=N+2}^{C-1} t_1(m) e_m + (\beta+\epsilon) \sum_{m=N+2}^{C-1} t_1(m) e_{N+1}\right)$$

and $\tilde{t}_0(0) = 0$, $\tilde{t}_0(n) = \lambda \tilde{t}_1(n-1) - \tilde{t}_1(n)$ for all $n > 0$, where Z is chosen to satisfy (12d); it is straightforward to verify that

$$Z = 1 + \lambda\epsilon \sum_{m=N+2}^{C-1} t_1(m).$$

In particular, for $\epsilon > 0$, we obtain that $Z > 1$.

We begin by showing that $\tilde{t}$ is feasible for (12) for all small enough $\epsilon, \beta > 0$. Note that $\tilde{t}$ satisfies (12c) and (12d) by definition. Thus, we verify the conditions (12a) and (12b).

First, note that for $\ell \leq N$, we have $\tilde{t}_1(\ell) = \frac{1}{Z} t_1(\ell)$. On the other hand, we have $\tilde{t}_1(N+1) = \frac{1}{Z}(t_1(N+1) + (\beta+\epsilon)\sum_{m=N+2}^{C-1} t_1(m))$ and $\tilde{t}_1(\ell) = \frac{(1-\beta)}{Z} t_1(\ell)$ for $\ell > N+1$. Using this and the definition of $\tilde{t}_0$, it is straightforward to verify that for small enough $\epsilon, \beta > 0$, we have $\tilde{t}_0 \geq 0$ and $\tilde{t}_0(0) = 0$ for all $\ell \leq N$. Further using the fact that $\sum_\ell \tilde{t}_0(\ell) \leq 1$, we obtain $\tilde{t}_0 \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$ for all small enough $\epsilon, \beta > 0$, and hence $\tilde{t}$ satisfies (12a).

Next, since $t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, we can write the decomposition

$$t_1 = \sum_{j=0}^{M} \alpha(j) e_j + \sum_{j=0}^{M} \sum_{i=M+1}^{C-1} \delta(i,j) \chi(i,j),$$

where $\alpha(j) \geq 0$ for $j \leq M$, $\delta(i,j) \geq 0$ for $j \leq M < i$, and $\sum_{j=0}^{M} \alpha(j) + \sum_{j=0}^{M} \sum_{i=M+1}^{C-1} \delta(i,j) \leq 1$. Thus, we have

$$\begin{aligned}
Z\tilde{t}_1 &= t_1 - \beta \sum_{m=N+2}^{C-1} t_1(m)(e_m - e_{N+1}) + \epsilon \sum_{m=N+2}^{C-1} t_1(m) e_{N+1} \\
&= t_1 - \beta \sum_{m=N+2}^{C-1} \sum_{j=0}^{M} \delta(m,j)\gamma(m,j)(e_m - e_{N+1}) \\
&\quad + \epsilon \sum_{m=N+2}^{C-1} \sum_{j=0}^{M} \delta(m,j)\gamma(m,j) e_{N+1}.
\end{aligned}$$

Now, it can be readily verified that for $0 \leq j \leq M$ and $m > N+1$,

$$\gamma(m,j)(e_m - e_{N+1}) = \chi(m,j) - \left(\frac{\gamma(m,j)}{\gamma(N+1,j)} \chi(N+1,j) + \left(1 - \frac{\gamma(m,j)}{\gamma(N+1,j)}\right) e_j\right).$$

Thus, after some algebra, we obtain

$$\begin{aligned} Z\tilde{t}_1 = t_1 &- \beta \sum_{m=N+2}^{C-1} \sum_{j=0}^{M} \delta(m,j)\chi(m,j) \\ &+ \beta \sum_{m=N+2}^{C-1} \sum_{j=0}^{M} \delta(m,j) \left(\frac{\gamma(m,j)}{\gamma(N+1,j)} \chi(N+1,j) + \left(1 - \frac{\gamma(m,j)}{\gamma(N+1,j)}\right) e_j \right) \\ &+ \epsilon \sum_{m=N+2}^{C-1} \sum_{j=0}^{M} \delta(m,j)\gamma(m,j) e_{N+1}. \end{aligned}$$

Now, since $t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, we have the first term $t_1 - \beta \sum_{m=N+2}^{C-1} \sum_{j=0}^{M} \delta(m,j)\chi(m,j) \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$ for small enough $\beta > 0$. Moreover, since $\hat{\rho}(\cdot)$ satisfies Assumption 3 (Lemma EC.4), we have $\gamma(N+1,j) > \gamma(m,j)$ for all $j \leq M$ and $m > N+1$. Hence, for all small enough $\beta > 0$, we obtain that second term lies in the $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, but not in $\mathsf{Conv}(K_{01} \cup \{\mathbf{0}\})$. This in turn implies that the second and the third term together lie in $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$ for small enough $\beta, \epsilon > 0$. Taken together, this implies that $Z\tilde{t}_1$ (and hence $\tilde{t}_1$) lies in the cone generated by $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$. Since $\sum_\omega \tilde{t}_1(\omega) \leq 1$, we conclude that $\tilde{t}_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, and hence $\tilde{t}$ satisfies (12b).

Having shown the feasibility of $\tilde{t}$, we now verify that the objective value under $\tilde{t}$ is greater than that under t. Using (12d) and (12c), we obtain

$$\sum_{\ell=0}^{C-1} \tilde{t}_1(\ell) = \frac{1}{\lambda}\left(1 - \tilde{t}_1(0)\right) > \frac{1}{\lambda}\left(1 - t_1(0)\right) = \sum_{\ell=0}^{C-1} t_1(\ell).$$

Here, the inequality follows from the fact that $\tilde{t}_1(0) = \frac{1}{Z} t_1(0) < t_1(0)$ since $Z > 1$. Thus, the objective of (12) for $\tilde{t}$ is strictly greater than that for t. Thus, it follows that t cannot be optimal for (12).

To conclude the proof, we obtain from the preceding argument that for any optimal $t = (t_1, t_0)$, if N denotes the largest value of n such that $t_1(k) = \lambda^k t_1(0)$ for all $k \leq n$, then either $N \geq C-2$ or $t_1(N+2) = 0$. In either case, we obtain that t induces a threshold structure in the customers' actions. $\square$

E.2. Proof of Proposition 4

Proof. The optimal signaling scheme can be found from the optimal solution to the linear program (12), using an argument similar to the discussion surrounding (8). Moreover, from similar argument as in Proposition 1, it follows that there exists an optimal signaling scheme that uses at most $|\Omega| = C$ signals per action. As $\hat{\rho}$ satisfies Assumption 2 (Lemma EC.4), all the signals that induce the customer to leave ($a = 0$) can be coalesced into single signal, which we call Leave. Taken together, we obtain that there exists a signaling scheme with the signal set $\{\mathsf{Join}_1, \mathsf{Join}_2, \ldots, \mathsf{Join}_J, \mathsf{Leave}\}$ for some $J \leq |C|$, where each Join_i signal induces the customer to join the queue, and the Leave signal induces the customer to leave.

We will now show that there exists such an optimal signaling scheme that satisfies the three properties listed in the proposition statement. For ease of exposition, we will split the proof into three parts, one part for each property.

Part 1. From the constraints (12a) and (12b), we observe that the optimal signaling scheme can be implemented with the canonical set of join signals consisting of pure signals in K_1 and binary mixed signals in K_{01}. The customer's utility upon receiving a binary mixed signal in K_{01} or a signal in $K_1 \cap K_0$ is zero by definition. Thus, to prove that a customer's utility upon receiving each signal Join_j is zero, we need to prove that there exists a signaling scheme that never sends the pure signals in $K_1 \cap K_0^c$. To show this, we use a perturbation argument, and show that any canonical signaling scheme that sends a pure signal in $K_1 \cap K_0^c$ with positive probability must be sub-optimal.

Let $t = (t_0, t_1)$ be a feasible solution to (12) with $t_0 \neq \mathbf{0}$. If t does not have a threshold structure, then Lemma 4 implies that t cannot be optimal for (12). Thus, henceforth we assume that t has a threshold structure, i.e., there exists an $m \in L_0 \cup \{C\}$ such that $t_1(\omega) > 0$ for $\omega < m$, $t_1(\omega) = 0$ for $\omega \geq m+1$, and $t_0(\omega) = 0$ for $\omega \notin \{m, m+1\}$. Furthermore, since $\sum_{\omega\in\Omega} t_0(\omega) > 0$, we obtain that $m < C$, and thus one can let $m \in L_0$ with $t_0(m) > 0$.

Now, suppose $t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$ but $t_1 \notin \mathsf{Conv}((K_1 \cap K_0) \cup K_{01} \cup \{\mathbf{0}\})$. The former condition implies that we can write t_1 as

$$t_1 = \sum_{\omega_1 \in K_1} \beta_{\omega_1} e_{\omega_1} + \sum_{\phi \in K_{01}\setminus K_1} \beta_\phi \phi.$$

where $\sum_{\omega_1\in K_1} \beta_{\omega_1} + \sum_{\phi\in K_{01}\setminus K_1} \beta_\phi \leq 1$. Furthermore, since $t_1 \notin \mathsf{Conv}((K_1 \cap K_0) \cup K_{01} \cup \{\mathbf{0}\})$, it follows that there exists an $\omega_1^* \in K_1 \cap K_0^c$ such that $\beta_{\omega_1^*} > 0$. Using these facts, we will now construct a feasible $\hat{t}$ with strictly higher objective, thus implying that t cannot be optimal.

For small enough $\epsilon > 0$, let $\hat{t} = (\hat{t}_0, \hat{t}_1)$ be defined as

$$\begin{aligned}
\hat{t}_1 &= \frac{1}{Z}\left(t_1 + \epsilon e_m\right) \\
\hat{t}_0 &= \frac{1}{Z}\left(t_0 - \epsilon e_m + \lambda\epsilon e_{m+1}\mathbf{I}\{m \neq C-1\}\right),
\end{aligned}$$

where Z is chosen so that $\hat{t}$ satisfies (12d): $Z = 1 + \lambda\epsilon > 1$. It is straightforward to verify that $\hat{t}$ satisfies (12c). Furthermore, since $t_0(m) > 0$, we have for small enough $\epsilon > 0$, $t_0 - \epsilon e_m \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$. Moreover, if $m \neq C-1$, then $e_{m+1} \in \mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$. This implies that $\hat{t}_0$ itself lies in $\mathsf{Conv}(L_0 \cup \{\mathbf{0}\})$, and thus satisfies (12a).

Thus, to show the feasibility of $\hat{t}$, it remains to show that (12b) holds. We have

$$\begin{aligned}
\hat{t}_1 &= \frac{1}{Z}\left(t_1 + \epsilon e_m\right) \\
&= \frac{1}{Z}\left(\sum_{\omega_1\in K_1, \omega_1 \neq \omega_1^*} \beta_{\omega_1} e_{\omega_1} + \sum_{\phi \in K_{01}\setminus K_1} \beta_\phi \phi\right) + \frac{1}{Z}\left(\beta_{\omega_1^*} e_{\omega_1^*} + \epsilon e_m\right).
\end{aligned}$$

Now, since $Z > 1$, the first term lies in the set $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$. Furthermore, since $\omega_1^* \in K_1 \in K_0^c$ and $m \in L_0$, for small enough $\epsilon > 0$, the second term can be written as a convex combination of ω_1^*, $\chi(m, \omega_1^*) \in K_{01}$ and $\mathbf{0}$. Thus, we obtain that the second term lies in the set $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$ as well. Finally, since $\sum_{\omega\in\Omega} \hat{t}_1(\omega) = \frac{1}{Z}\left(\sum_{\omega\in\Omega} t_1(\omega) + \epsilon\right) = \frac{1}{Z}\left(1 - \lambda t_1(C-1) - \sum_{\omega\in\Omega} t_0(\omega) + \epsilon\right) < 1$ for small enough $\epsilon > 0$, we obtain that $\hat{t}_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$. Thus, $\hat{t}$ is feasible for (12).

Finally, note that,

$$\sum_{\omega\in\Omega} \hat{t}_1(\omega) - \sum_{\omega\in\Omega} t_1(\omega) = \frac{\epsilon}{Z}\left(1 - \lambda \sum_{\omega\in\Omega} t_1(\omega)\right) > 0,$$

where the final inequality follows from the fact that the throughput $\lambda \sum_{\omega\in\Omega} t_1(\omega)$ must be strictly less than the mean service rate, assumed to equal one. Thus, we obtain that $\hat{t}$ achieves a strictly higher objective than t, and hence t cannot be optimal for (12). Summarizing, we obtain that any optimal $t = (t_0, t_1)$ with $t_0 \neq \mathbf{0}$ must satisfy $t_1 \in \mathsf{Conv}((K_1 \cap K_0) \cup K_{01} \cup \{\mathbf{0}\})$, and thus under the optimal signaling mechanism, for any signal that induces the customer to join, the induced beliefs are in the set $(K_1 \cap K_0) \cup K_{01}$, implying that her utility for joining must be zero.

Part 2. From the proof of part (1), it follows that for each $j \leq J$, the signal Join_j either induces a posterior belief $\omega^j \in K_1 \cap K_0$ or a belief $\chi(\omega_0^j, \omega_1^j)$ between a pair of states $\omega_0^j \in L_0$ and $\omega_1^j \in K_1$. In the former case, we let $\omega_0^j = \omega_1^j = \omega^j$. We next show that the set of such state-pairs forms a sandwich structure, in the sense stated in the proposition statement.

Consider two signals Join_i and Join_j inducing beliefs on state-pairs $(\omega_0^i, \omega_1^i) = (n, m)$ and $(\omega_0^j, \omega_1^j) = (\ell, k)$ respectively, with $m, k \in K_1$ and $n, \ell \in K_0$. Without loss of generality, we assume $m \leq k$.

Now, if $n \in K_0 \cap K_1$, then $m = n$, and using the fact that $\hat{\rho}(e_i)$ is strictly decreasing, we obtain that $m = k$, and hence $k = m = n \leq \ell$, which conforms with the sandwich structure. Similarly, if $\ell \in K_0 \cap K_1$, then $k = \ell$, and hence $m \leq k = \ell \leq n$, which again conforms with the sandwich structure. Thus, hereafter we assume that $n, \ell \in K_0 \cap K_1^c = L_0$. Hence, we obtain $m \leq k < n$ and $m \leq k < \ell$. Thus, we have three possible cases: (1) $m \leq k < \ell \leq n$; (2) $m = k < n < \ell$; and (3) $m < k < n < \ell$. The first two cases conform with the sandwich structure stated in the proposition. Thus, we focus on the third case and show that such a signaling scheme cannot be optimal.

Thus, suppose $m < k < n < \ell$. We next show that any strict convex combination of $\chi(n, m)$ and $\chi(\ell, k)$ can be written as either a convex combination of $e_m, \chi(\ell, m)$, $\chi(\ell, k)$ and $\chi(n, k)$, or a convex combination of $e_m, \chi(\ell, m), \chi(n, m)$ and $\chi(n, k)$ with positive weight on e_m in either case. For notational simplicity, we denote $\gamma(i, j)$ by γ_{ij}. For any $\alpha \in (0, 1)$, we obtain from straightforward (but tedious) algebra,

$$\begin{aligned}
&\alpha\chi(n, m) + (1-\alpha)\chi(\ell, k) \\
&= \alpha\left(\gamma_{nm} e_n + (1-\gamma_{nm}) e_m\right) + (1-\alpha)\left(\gamma_{\ell k} e_\ell + (1-\gamma_{\ell k}) e_k\right)
\end{aligned}$$

$$= \alpha(1-\gamma_{nm}) \left(1 - \frac{\gamma_{nm}\gamma_{\ell k}(1-\gamma_{nk})(1-\gamma_{\ell m})}{\gamma_{\ell m}\gamma_{nk}(1-\gamma_{\ell k})(1-\gamma_{nm})}\right) e_m + \alpha \frac{\gamma_{nm}\gamma_{\ell k}(1-\gamma_{nk})}{\gamma_{\ell m}\gamma_{nk}(1-\gamma_{\ell k})}\chi(\ell,m)$$
$$+ \frac{\gamma_{nm}(1-\gamma_{nk})}{\gamma_{nk}(1-\gamma_{\ell k})}\left((1-\alpha)\frac{\gamma_{nk}(1-\gamma_{\ell k})}{\gamma_{nm}(1-\gamma_{nk})} - \alpha\right)\chi(\ell,k) + \frac{\alpha\gamma_{nm}}{\gamma_{nk}}\chi(n,k) \tag{EC.4}$$
$$= (1-\alpha)(1-\gamma_{nm})\frac{\gamma_{nk}(1-\gamma_{\ell k})}{\gamma_{nm}(1-\gamma_{nk})}\left(1 - \frac{\gamma_{nm}\gamma_{\ell k}(1-\gamma_{nk})(1-\gamma_{\ell m})}{\gamma_{\ell m}\gamma_{nk}(1-\gamma_{\ell k})(1-\gamma_{nm})}\right) e_m + (1-\alpha)\frac{(1-\gamma_{\ell k})}{(1-\gamma_{nk})}\chi(n,k)$$
$$+ \left(\alpha - (1-\alpha)\frac{\gamma_{nk}(1-\gamma_{\ell k})}{\gamma_{nm}(1-\gamma_{nk})}\right)\chi(n,m) + (1-\alpha)\frac{\gamma_{\ell k}}{\gamma_{\ell m}}\chi(\ell,m). \tag{EC.5}$$

In Lemma EC.5, we show that for $m<k<n<\ell$, $0 < \frac{\gamma_{nm}\gamma_{\ell k}(1-\gamma_{nk})(1-\gamma_{\ell m})}{\gamma_{\ell m}\gamma_{nk}(1-\gamma_{\ell k})(1-\gamma_{nm})} < 1$. Thus, the coefficient of e_m in both (EC.4) and (EC.5) is positive. If $\frac{\alpha}{1-\alpha} \le \frac{\gamma_{nk}(1-\gamma_{\ell k})}{\gamma_{nm}(1-\gamma_{nk})}$, then the coefficient of $\chi(\ell,k)$ in (EC.4) is non-negative. On the other hand, if $\frac{\alpha}{1-\alpha} \ge \frac{\gamma_{nk}(1-\gamma_{\ell k})}{\gamma_{nm}(1-\gamma_{nk})}$, then the coefficient of $\chi(n,m)$ in (EC.5) is non-negative. Thus, taken together, we obtain that for any $\alpha \in (0,1)$, the convex combination $\alpha\chi(n,m) + (1-\alpha)\chi(\ell,k)$ can be written as either a convex combination of $e_m, \chi(\ell,m), \chi(\ell,k)$ and $\chi(n,k)$ or a convex combination of $e_m, \chi(\ell,m), \chi(n,m)$ and $\chi(n,k)$, with positive weight on e_m in either case.

Now, since $m<k$ with $k \in K_1$, it follows that $m \in K_1 \cup K_0^c$. This in turn implies that, using arguments as in the proof of Part (1), one can appropriately combine the weight on e_m with an $\omega \in L_0$ to construct an additional signal which induces a posterior belief $\chi(m,\omega)$, leading to an increased probability of customer joining the queue. Thus, if $m<k<n<\ell$, the signaling scheme cannot be optimal.

To conclude, we obtain that for any pair of signals Join_i and Join_j with $\omega_1^i \le \omega_1^j$, it must be that $\omega_1^i \le \omega_1^j \le \omega_0^j \le \omega_1^i$, implying the sandwich structure stated in the proposition statement.

Part 3. Finally, we show that in the optimal signaling scheme, for $i<j$, we have $\mathbf{E}[X|\mathsf{Join}_i] \le \mathbb{E}[X|\mathsf{Join}_j]$ and $\mathrm{Var}[X|\mathsf{Join}_i] \ge \mathrm{Var}[X|\mathsf{Join}_j]$. Since $\mathbf{E}[X|\mathsf{Join}_i] + \beta\sqrt{\mathrm{Var}[X|\mathsf{Join}_i]} = \mathbf{E}[X|\mathsf{Join}_j] + \beta\sqrt{\mathrm{Var}[X|\mathsf{Join}_j]} = \tau$ from Part (1), it suffices to prove $\mathbf{E}[X|\mathsf{Join}_i] \le \mathbf{E}[X|\mathsf{Join}_j]$.

To prove this, let $m,k \in K_1$ and $n,\ell \in K_0$ be such that Join_i induces posterior beliefs over the state-pair $\{m,n\}$ and Join_j induces beliefs over the pair $\{k,\ell\}$. The sandwich structure from Part (2) implies that $m \le k \le \ell \le n$. Suppose $k<\ell$, implying that $\ell, n \in L_0$. We have $\mathbf{E}[X|\mathsf{Join}_i] = \gamma_{nm}(n+1) + (1-\gamma_{nm})(m+1) = 1 + \gamma_{nm}n + (1-\gamma_{nm})m$ and $\mathbf{E}[X|\mathsf{Join}_j] = 1 + \gamma_{\ell k}\ell + (1-\gamma_{\ell k})k$. In Lemma EC.6, we show that for $m \in K_1$ and $n \in L_0$, the function $\gamma_{nm}n + (1-\gamma_{nm})m$ is non-decreasing in m (for fixed n) and non-increasing in n (for fixed m). Thus, we have for $m \le k < \ell \le n$,

$$\mathbf{E}[X|\mathsf{Join}_i] = \gamma_{nm}n + (1-\gamma_{nm})m \le \gamma_{nk}n + (1-\gamma_{nk})k \le \gamma_{\ell k}\ell + (1-\gamma_{\ell k})k = \mathbf{E}[X|\mathsf{Join}_j].$$

On the other hand, if $k=\ell$, then $k=\ell \in K_1 \cap K_0$, implying that $n \in L_0$, and hence $m<k$ and $\gamma_{nk}=0$. Once again, we have $\mathbf{E}[X|\mathsf{Join}_i] = \gamma_{nm}n + (1-\gamma_{nm})m \le \gamma_{nk}n + (1-\gamma_{nk})k = k = \mathbf{E}[X|\mathsf{Join}_j]$. □

We use the following two lemmas in the proof of Proposition 4.

Lemma EC.5. *Let $\hat{\rho}(\mu) = \mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu(x)}$, and $m, k \in K_1$ and $n, \ell \in L_0$ be such that $m < k < n < \ell$. Then, we have*

$$0 < \frac{\gamma_{nm}\gamma_{\ell k}(1-\gamma_{nk})(1-\gamma_{\ell m})}{\gamma_{\ell m}\gamma_{nk}(1-\gamma_{\ell k})(1-\gamma_{nm})} < 1.$$

Proof. Since $n, \ell \in L_0$, it follows from definition that the ratio in the lemma statement is positive. Thus, it remains to show that the ratio is strictly less than one. To simplify the notation, let $f(n,m) \triangleq \log\left(\frac{\gamma_{nm}}{1-\gamma_{nm}}\right)$. Then, the upper-bound condition in the lemma statement is equivalent to requiring

$$f(n,k) - f(n,m) > f(\ell,k) - f(\ell,m),$$

i.e., that f has decreasing differences. Now, using the expression for γ_{nm} from (EC.3), we obtain

$$\begin{aligned} f(n,m) &= \log\left(\frac{\gamma_{nm}}{1-\gamma_{nm}}\right) \\ &= \log\left(\frac{2(\tau-1-m) + \beta^2(n-m+1) - \beta\sqrt{h(n,m)}}{2(n-m)(1+\beta^2) - 2(\tau-1-m) - \beta^2(n-m+1) + \beta\sqrt{h(n,m)}}\right). \end{aligned}$$

We use the preceding expression to extend the definition of f to non-integer values of m and n. Then, by a straightforward calculation, it follows that

$$\frac{\partial^2 f(n,m)}{\partial n \partial m} = -\frac{2\beta(\beta^2+1)(n-m)}{(h(n,m))^{3/2}} < 0.$$

Thus, by the mean value theorem, there exists $\xi_1 \in [n,\ell]$ and $\xi_2 \in [m,k]$ with

$$(f(\ell,k) - f(\ell,m)) - (f(n,k) - f(n,m)) = \frac{\partial^2}{\partial x \partial y} f(x,y)\bigg|_{x=\xi_1, y=\xi_2} (\ell-n)(k-m) < 0. \quad \square$$

Lemma EC.6. *For $m \in K_1$ and $n \in L_0$, let $g(n,m) \triangleq \gamma_{nm} n + (1-\gamma_{nm})m$. Then, $g(n,m)$ is non-decreasing in m (for fixed n), and non-increasing in n (for fixed m).*

Proof. Using (EC.3), we obtain

$$g(n,m) = m + \gamma_{nm}(n-m) = m + \frac{2(\tau-1-m) + \beta^2(n-m+1) - \beta\sqrt{h(n,m)}}{2(1+\beta^2)},$$

which implies

$$\begin{aligned} \frac{\partial g(n,m)}{\partial n} &= \frac{\beta\left(\beta\sqrt{h(n,m)} + \beta^2(m-n-1) + 2m - 2\tau + 2\right)}{2(1+\beta^2)\sqrt{h(n,m)}}, \\ \frac{\partial g(n,m)}{\partial m} &= \frac{\beta\left(\beta\sqrt{h(n,m)} + \beta^2(n-m-1) + 2n - 2\tau + 2\right)}{2(1+\beta^2)\sqrt{h(n,m)}}. \end{aligned}$$

Using the definition of $h(n,m)$, we obtain

$$\beta^2 h(n,m) - \left(\beta^2(m-n-1)+2m-2\tau+2\right)^2 = -4(1+\beta)^2\left((\tau-m-1)^2-\beta^2(m+1)\right) \le 0,$$
$$\beta^2 h(n,m) - \left(\beta^2(n-m-1)+2n-2\tau+2\right)^2 = -4(1+\beta)^2\left((\tau-n-1)^2-\beta^2(n+1)\right) \ge 0,$$

where we make use of the fact that $m+1+\beta\sqrt{m+1} \le \tau < n+1+\beta\sqrt{n+1}$ since $m \in K_1$ and $n \in L_0$. Thus, we get

$$\beta\sqrt{h(n,m)} \le |\beta^2(m-n-1)+2m-2\tau+2| = -\left(\beta^2(m-n-1)+2m-2\tau+2\right)$$
$$\beta\sqrt{h(n,m)} \ge |\beta^2(n-m-1)+2n-2\tau+2| \ge -\left(\beta^2(n-m-1)+2n-2\tau+2\right),$$

where the first line follows because $\beta^2(m-n-1)+2m-2\tau+2 \le 0$ as $\tau \ge m+1$ and $m \le n$. Taken together, we obtain $\frac{\partial g(n,m)}{\partial n} \le 0$ and $\frac{\partial g(n,m)}{\partial m} \ge 0$. $\square$

Appendix F: Examples of Risk-conscious Utility

In this section, we discuss the evidence for systematic deviations from expected utility maximization (EUM), using the examples of risk-conscious utility provided in Table 1.

F.1. Cumulative Prospect Theory

As discussed in the introduction, behavioral economists have documented a number of "paradoxes" that cannot be explained by expected utility maximization. In particular, the Allais paradox (Allais 1979) suggests that individuals choose a certain small gain over a small chance of a large gain (risk-averse), but choose a small chance of a large loss over a certain small loss (risk-seeking). They also perceive events relative to a reference point, and overweight extreme events but underweight typical events. Cumulative prospect theory, introduced by Tversky and Kahneman (1992), incorporates this idea of *rank-dependent expected utility* by weighting the subjective probability over outcomes differently from the true distribution, that is, they replace the true weight $\mu(\omega)$ in the expected utility $\sum_\omega \mu(\omega)u(\omega,a)$ with the subjective weight $f_\omega(\mu)$, yielding a risk-conscious utility of the form $\rho(\mu,a) = \sum_{\omega\in\Omega} f_\omega(\mu)u(\omega,a)$.

F.2. Mean-standard-deviation utility

The expression for mean-standard-deviation (mean-stdev) utility is $\mathbf{E}_\mu[u(\omega,a)] - \beta\sqrt{\mathrm{Var}_\mu(g(\omega,a))}$. There are two components to the utility: the receiver wants the first metric u to have high mean, and the second metric g to have low variability.

This functional specification of utility is often used when the agent is uncertainty-averse but is more likely to base her decision on a few summary statistics of the distribution (in this case, the first two moments) rather than the entire distribution. The canonical Markovitz portfolio theory in finance (Markowitz 1952, 1987) considers the utility as a function of mean return and volatility captured by standard deviation of return. There is a rich literature on traffic assignment and routing with stochastic travel times (Nikolova and Stier-Moses 2014, Cominetti and Torrico 2016, Lianeas et al. 2019) using the mean-stdev utility that also deals with strategic behavior of agents like our work. The distribution of waiting time in queues in practical settings is also often summarized with the mean and the standard deviation of meeting time: see, e.g., Ang et al. (2016) for emergency department wait time prediction.

F.3. Maximin Utility

The expression for maximin utility is $\rho(\mu, a) = \min_\theta \mathbf{E}_\mu[u(\omega, a; \theta)]$. This utility formulation arises when the agent faces uncertainty about the utility parameter θ, and seeks to hedge against the worst-case realization. The notion of using the maximin criterion for decision making has a long history; see, e.g., French (1986) for a discussion. More specifically, Armbruster and Delage (2015) and Hu and Mehrotra (2015) propose and analyze decision making problems with maximin utility in the form stated here, and many other subsequent works in decision theory and finance adopt it (Post and Kopa 2017, Hu et al. 2018, Guo and Xu 2021).

F.4. Risk Measures

A risk measure is a numerical metric assigned to an uncertain event to capture a potential loss in decision making problems under uncertainty. Risk measures are often designed to capture the "tail" of the distribution of losses, so they are necessarily nonlinear in the distribution (belief). Commonly used risk-measures are the Value-at-Risk $\mathsf{VaR} = \min\{t \in \mathbb{R} : \mathbf{P}_\mu(\ell(\omega, a) > t) \le 1 - \alpha\}$, and the Conditional Value-at-Risk $\mathsf{CVaR} = \mathbf{E}_\mu[\ell(\omega, a) | \ell(\omega, a) > \tau]$, where $\ell(\omega, a)$ is the loss under state ω and action a.

References

Acerbi C, Tasche D (2002) Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes* 31(2):379–388.

Ahmadi-Javid A (2012) Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications* 155(3):1105–1123.

Alizamir S, de Véricourt F, Wang S (2020) Warning against recurring risks: An information design approach. *Management Science* 66(10):4612–4629, URL `http://dx.doi.org/10.1287/mnsc.2019.3420`.

Allais M (1979) The so-called allais paradox and rational decisions under uncertainty. Allais M, Hagen O, eds., *Expected Utility Hypotheses and the Allais Paradox: Contemporary Discussions of the Decisions under Uncertainty with Allais' Rejoinder*, 437–681 (Springer Netherlands, Dordrecht), URL `http://dx.doi.org/10.1007/978-94-015-7629-1_17`.

Ang E, Kwasnick S, Bayati M, Plambeck EL, Aratow M (2016) Accurate emergency department wait time prediction. *Manufacturing & Service Operations Management* 18(1):141–156, URL http://dx.doi.org/10.1287/msom.2015.0560.

Anunrojwong J, Iyer K, Manshadi V (2020) Information design for congested social services: Optimal need-based persuasion. *Proceedings of the 21st ACM Conference on Economics and Computation*, 349–350, EC '20 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450379755, URL http://dx.doi.org/10.1145/3391403.3399504.

Arieli I, Babichenko Y (2019) Private bayesian persuasion. *Journal of Economic Theory* 182:185 – 217, URL http://dx.doi.org/10.1016/j.jet.2019.04.008.

Armbruster B, Delage E (2015) Decision making under uncertainty when preference information is incomplete. *Management Science* 61(1):111–128, URL http://dx.doi.org/10.1287/mnsc.2014.2059.

Artzner P, Delbaen F, Eber J, Heath D (2001) Coherent measures of risk. *Mathematical Finance* 9:203–228, URL http://dx.doi.org/10.1111/1467-9965.00068.

Aumann RJ, Maschler M (1995) *Repeated Games with Incomplete Information* (MIT press).

Bárány I, Onn S (1995) Carathéodory's theorem, colourful and applicable. *Intuitive Geometry* 6:11–21.

Beauchêne D, Li J, Li M (2019) Ambiguous persuasion. *Journal of Economic Theory* 179:312–365, URL http://dx.doi.org/10.1016/j.jet.2018.10.008.

Bergemann D, Bonatti A (2015) Selling cookies. *American Economic Journal: Microeconomics* 7(3):259–94, URL http://dx.doi.org/10.1257/mic.20140155.

Bergemann D, Bonatti A (2019) Markets for information: An introduction. *Annual Review of Economics* 11(1):85–107, URL http://dx.doi.org/10.1146/annurev-economics-080315-015439.

Bergemann D, Bonatti A, Smolin A (2018) The design and price of information. *American Economic Review* 108(1):1–48, URL http://dx.doi.org/10.1257/aer.20161079.

Bergemann D, Morris S (2016) Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics* 11(2):487–522.

Bergemann D, Morris S (2019) Information design: A unified perspective. *Journal of Economic Literature* 57(1):44–95, URL http://dx.doi.org/10.1257/jel.20181489.

Bimpikis K, Ehsani S, Mostagir M (2019) Designing dynamic contests. *Operations Research* 67(2):339–356, URL `http://dx.doi.org/10.1287/opre.2018.1823`.

Bimpikis K, Papanastasiou Y (2019) Inducing exploration in service platforms. Hu M, ed., *Sharing Economy: Making Supply Meet Demand*, chapter 10, 193–216 (Springer).

Candogan O (2019) Persuasion in networks: Public signals and k-cores. *Proceedings of the 2019 ACM Conference on Economics and Computation*, 133–134, EC '19 (Association for Computing Machinery, New York, NY, USA), URL `http://dx.doi.org/10.1145/3328526.3329618`.

Candogan O (2020) *Information Design in Operations*, 176–201 (INFORMS TutORials in Operations Research), URL `http://dx.doi.org/10.1287/educ.2020.0217`.

Candogan O, Drakopoulos K (2020) Optimal signaling of content accuracy: Engagement vs. misinformation. *Operations Research* 68(2):497–515, URL `http://dx.doi.org/10.1287/opre.2019.1897`.

Cominetti R, Torrico A (2016) Additive consistency of risk measures and its application to risk-averse routing in networks. *Mathematics of Operations Research* 41(4):1510–1521.

Das S, Kamenica E, Mirka R (2017) Reducing congestion through information design. *Proceedings of the 22th Conference on Economics and Computation (EC)55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* .

DellaVigna S (2009) Psychology and economics: Evidence from the field. *Journal of Economic Literature* 47(2):315–72, URL `http://dx.doi.org/10.1257/jel.47.2.315`.

Drakopoulos K, Jain S, Randhawa R (2021) Persuading customers to buy early: The value of personalized information provisioning. *Management Science* 67(2):828–853, URL `http://dx.doi.org/10.1287/mnsc.2020.3580`.

Dughmi S, Xu H (2016) Algorithmic Bayesian persuasion. *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, 412–425, STOC '16 (Association for Computing Machinery), URL `http://dx.doi.org/10.1145/2897518.2897583`.

Ellsberg D (1961) Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics* 75(4):643–669.

Föllmer H, Schied A (2016) *Stochastic Finance: An Introduction in Discrete Time* (De Gruyter), ISBN 9783110463453, URL `http://dx.doi.org/doi:10.1515/9783110463453`.

French S (1986) *Decision Theory: An Introduction to the Mathematics of Rationality* (Ellis Horwood).

Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18:141–153, URL http://dx.doi.org/10.1016/0304-4068(89)90018-9.

Guo S, Xu H (2021) Statistical robustness in utility preference robust optimization models. *Mathematical Programming* 190:679–720.

Hu J, Bansal M, Mehrotra S (2018) Robust decision making using a general utility set. *European Journal of Operations Research* 269(2):699–714, URL http://dx.doi.org/https://doi.org/10.1016/j.ejor.2018.02.018.

Hu J, Mehrotra S (2015) Robust decision making over a set of random targets or risk-averse utilities with an application to portfolio optimization. *IIE Transactions* 47(4):358–372, URL http://dx.doi.org/10.1080/0740817X.2014.919045.

Hu J, Weng X (2021) Robust persuasion of a privately informed receiver. *Economic Theory* 72:909–953.

Inostroza N, Pavan A (2021) Persuasion in global games with application to stress testing. Technical report, CEPR Discussion Paper No. DP16484.

Jorion P (2006) *Value at Risk: The New Benchmark for Managing Financial Risk* (McGraw-Hill, New York).

Kahneman D, Tversky A (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3(3):430–454.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.

Kamenica E (2019) Bayesian persuasion and information design. *Annual Review of Economics* 11(1):249–272, URL http://dx.doi.org/10.1146/annurev-economics-080218-025739.

Kamenica E, Gentzkow M (2011) Bayesian persuasion. *American Economic Review* 101(6):2590–2615.

Kolotilin A, Mylovanov T, Zapechelnyuk A, Li M (2017) Persuasion of a privately informed receiver. *Econometrica* 85(6):1949–1964.

Lianeas T, Nikolova E, Stier-Moses NE (2019) Risk-averse selfish routing. *Mathematics of Operations Research* 44(1):38–57, URL http://dx.doi.org/10.1287/moor.2017.0913.

Lingenbrink D, Iyer K (2018) Signaling in online retail: Efficacy of public signals. *SSRN Electronic Journal* URL `http://dx.doi.org/10.2139/ssrn.3179262`.

Lingenbrink D, Iyer K (2019) Optimal signaling mechanisms in unobservable queues. *Operations Research* 67(5):1397–1416, URL `http://dx.doi.org/10.1287/opre.2018.1819`.

Lipnowski E, Mathevet L (2018) Disclosure to a psychological audience. *American Economic Journal: Microeconomics* 10(4):67–93, URL `http://dx.doi.org/10.1257/mic.20160247`.

Machina MJ (1982) "Expected utility" analysis without the independence axiom. *Econometrica* 50(2):277–323.

Machina MJ (1995) Non-expected utility and the robustness of the classical insurance paradigm. *The Geneva Papers on Risk and Insurance Theory* 20(1):9–50.

Maister DH (1984) *The psychology of waiting lines* (Harvard Business School, Boston).

Markowitz H (1987) *Mean-Variance Analysis in Portfolio Choice and Capital Markets* (Blackwell, New York).

Markowitz HM (1952) Portfolio selection. *Journal of Finance* 7(1):77–91.

Nikolova E, Stier-Moses NE (2014) A mean-risk model for the traffic assignment problem with stochastic travel times. *Operations Research* 62(2):366–382.

Papanastasiou Y, Bimpikis K, Savva N (2018) Crowdsourcing exploration. *Management Science* 64(4):1727–1746, URL `http://dx.doi.org/10.1287/mnsc.2016.2697`.

Post T, Kopa M (2017) Portfolio choice based on third-degree stochastic dominance. *Management Science* 63(10):3381–3392, URL `http://dx.doi.org/10.1287/mnsc.2016.2506`.

Rabin M (1998) Psychology and economics. *Journal of Economic Literature* 36(1):11–46.

Rayo L, Segal I (2010) Optimal information disclosure. *Journal of Political Economy* 118(5):949–987.

Romanyuk G, Smolin A (2019) Cream skimming and information design in matching markets. *American Economic Journal: Microeconomics* 11(2):250–76, URL `http://dx.doi.org/10.1257/mic.20170154`.

Savage LJ (1954) *The Foundations of Statistics* (John Wiley & Sons, New York).

Taneva I (2019) Information design. *American Economic Journal: Microeconomics* 11(4):151–85, URL `http://dx.doi.org/10.1257/mic.20170351`.

Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5(4):297–323.

Yang P, Iyer K, Frazier P (2019) Information design in spatial resource competition. *The 15th Conference on Web and Internet Economics*, WINE 2019, URL arXiv:1909.12723.

Zheng S, Chen Y (2021) Optimal advertising for information products. *Proceedings of the 22nd ACM Conference on Economics and Computation*, 888–906, EC '21 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450385541, URL http://dx.doi.org/10.1145/3465456.3467649.

Ziegler G (2020) Adversarial bilateral information design. Technical report.