Proof of Lemma 1: The first step, analogous to Myerson (1981), establishes that given the information acquisition of the buyers, it is sufficient for them to report their posteriors. Let $Y$ be the action space in $\mathcal{M}$, and $\tau$ be the distribution of posteriors that the buyer acquires in equilibrium. For each $\mu \in \text{supp}(\tau)$, the buyer will choose some strategy $\xi : \mu \rightarrow \Delta(Y)$. Let $\hat{x}(\xi(\mu_1), \ldots, \xi(\mu_N))$ be the vector of probabilities that buyers receive the item by playing according to strategy $\xi$; similarly, define $\hat{t}(\xi(\mu_1), \ldots, \xi(\mu_N))$ to be the vector of expected transfers. One can then define the direct revelation mechanism $\mathcal{M}'$ where each buyer reports her posterior $\mu_i$, and the probabilities of receiving the item and transfers are given by

$$ x(\mu_1, \ldots, \mu_N) = \hat{x}(\xi(\mu_1), \ldots, \xi(\mu_N)) $$

$$ t(\mu_1, \ldots, \mu_N) = \hat{t}(\xi(\mu_1), \ldots, \xi(\mu_N)) $$

Hence each buyer receives the same expected utility as in $\mathcal{M}$ for each possible report of posterior; since $\xi$ was an equilibrium strategy in $\mathcal{M}$, it is optimal in $\mathcal{M}'$ to report one’s true posterior.

Similarly, any distribution of posteriors $\tau'$ will yield a weakly lower payoff than $\tau$, as the same set of payoffs is feasible in $\mathcal{M}'$ as from acquiring $\tau'$ in mechanism $\mathcal{M}$ and then choosing $\xi(\mu)$ for each $\mu \in \text{supp}(\tau')$. Hence it will be optimal to acquire $\tau$ in $\mathcal{M}'$.

The above shows that it is without loss to consider mechanisms in which the seller recommends that the buyer acquire $\tau$, and report their posterior $\mu$;
there will then be a unique \( x \) for each reported \( \mu \). It is also clear that for each \( x \), there must be a unique \( t \), since otherwise the buyer could misreport her type \( \mu \) in order to get a lower \( t \). To complete the proof, one must show conversely that for each \( x \), there is a unique \( \mu \in \text{supp}(\tau) \) that receives the item with probability \( x \). Suppose otherwise; let \( 1_x(s) \) be the indicator function on the signal space that takes the value 1 if, upon receiving signal \( s \), the buyer receives the item with probability \( x \), and 0 otherwise. This is a measurable function with respect to \( \pi \), and so the buyer’s ex-ante payoff is given by

\[
\sum_{\theta \in \Theta} \int_S \int_0^1 (x\theta - t(x)) 1_x(s)\mu_0(\theta) dx d\pi(s|\theta) - H(\mu_0) + \int_{\Delta(\Theta)} H(\mu) d\tau(\mu)
\]

where \( t(x) \) is the transfer associated with \( x \). If the set of signal realizations for which the same \( x \) is chosen is of measure greater than 0 with respect to \( \pi \), then there exists \( \hat{\pi} \) in which all signal realizations \( s \) for which \( x \) is chosen are merged into one signal \( \hat{s} \), upon whose reception the buyer again chooses \( x \). If \( \mu(\cdot|s) \) is not the same almost everywhere for all such \( s \), then the cost of information acquisition is strictly lower, and hence an improvement for the buyer. Hence it is without loss that there is a unique \( \mu \) for which \( x \) is chosen almost everywhere. □

**Proof of Lemma 2:** To see that (IR-A) is implied by the other constraints, let \( \bar{x} \equiv \min \{ x \in X \} \). By standard single-crossing arguments from (IC-I), \( E_{\mu(x)}[\theta] \) is increasing in \( x \). Thus, for all \( x \in X \),

\[
\bar{x} E_{\mu(x)}[\theta] - t(\bar{x}) \geq \bar{x} E_{\mu(\bar{x})}[\theta] - t(\bar{x}) \geq 0
\]

Furthermore, the buyer can acquire no information, which is costless. Therefore, by (IC-I),

\[
\int \int [x(\mu)\theta - t(x(\mu))] d\mu(\theta) d\tau(\mu) - [H(\mu_0) - \int H(\mu) d\tau(\mu)] \geq
\int \int [\bar{x} \theta - t(\bar{x})] d\mu(\theta) d\tau(\mu) - [H(\mu_0) - H(\mu_0)]
\]
\[ = \int [x^* \theta - t(x^*)] \, d\mu_0(\theta) \]

\[ \geq 0 \]

where the last inequality is by (IR-I).

For part (ii), we show that if there is a deviation ex interim that is an improvement for the buyer, then there exists some \( \hat{\pi} \) that is an improvement ex ante for the buyer. By Bayes’ rule and Fubini’s theorem, the buyer’s objective in (IC-A) can be written as the linear operator of \( \pi(\cdot|\theta) \),

\[
F(\pi) \equiv \sum_{\theta \in \Theta} \int_X [x \theta - t(x)] + H\left( \frac{d\pi(x|\theta) \mu_0(\theta)}{\sum_{\theta' \in \Theta} d\pi(x|\theta') \mu_0(\theta')} \right) \, d\pi(x|\theta) \mu_0(\theta) \tag{15}
\]

where \( \frac{d\pi(x|\theta)}{\sum_{\theta' \in \Theta} d\pi(x|\theta')} \) is the Radon-Nikodym derivative of the measure \( d\pi(x|\theta) \mu_0(\theta) \) with respect to \( \sum_{\theta' \in \Theta} d\pi(x|\theta') \mu_0(\theta') \). By assumption, since \( \pi \) is a valid signal (i.e. it generates posteriors via Bayes’ rule), the measures \( \{\pi(\cdot|\theta)\}_{\theta \in \Theta} \) are absolutely continuous with their sum and so this Radon-Nikodym derivative is well defined.

Suppose that for some subset of allocations \( Y = \{x\} \) that are recommended with positive probability according to \( \pi \), there is some action \( \hat{x}(x) \) that the buyer strictly prefers, i.e.

\[
\sum_{\theta} \int_Y [\hat{x}(x) \theta - t(\hat{x}(x))] \mu_0(\theta) \, d\pi(x|\theta) > \sum_{\theta} \int_Y [x \theta - t(x)] \mu_0(\theta) \, d\pi(x|\theta)
\]

This same ex-interim payoff could be achieved by using the recommendation strategy \( \hat{\pi}(x|\theta) \) where, instead of recommending \( x \), \( \hat{x}(x) \) is recommended, i.e.

\[
d\hat{\pi}(x|\theta) = \begin{cases} 
0, & x \in Y \\
d\pi(x|\theta) + \int_{y \in Y \setminus \hat{x}(y) = x} d\pi(y|\theta), & x \notin Y 
\end{cases}
\]

Moreover, since \( H \) is concave, the information cost is reduced because the buyer no longer distinguishes between the cases where \( x \) was recommended and \( \{y \in Y : \hat{x}(y) = x\} \) was recommended, and instead generates a single posterior
that is the weighted average (according to \( \tau \)) of \( \mu(\cdot|x) \) and \( \{ \mu(\cdot|y) : \hat{x}(y) = x \} \).
Thus the buyer could improve her expected payoff at least as much by an ex-ante deviation for any \( \pi \). □

**Proof of Lemma 3:** By Lemma A, \( \exists \epsilon > 0 \) such that \( \mu(\theta|x) > \epsilon, \forall \theta, x \). Hence \( H(\mu(\cdot|x)) \) and \( \frac{\partial H}{\partial \mu(\theta)}(\mu(\cdot|x)) \) are bounded. By (15), one can view the buyer’s objective as a linear operator of \( \pi(\cdot|\theta) \).

Consider the set of finite signed measures \( \{ \hat{\pi}(\cdot|\theta) \}_{\theta \in \Theta} \) that are absolutely continuous with respect to \( \pi \), and endow it with the norm

\[
\| \hat{\pi} \| = \left( \sum_{\theta \in \Theta} \int \left( \frac{d\hat{\pi}(x|\theta)}{d\pi(x|\theta)} \right)^2 d\pi(x|\theta) \mu_0(\theta) \right)^{1/2}
\]

Thus \( \{ \hat{\pi}(\cdot|\theta) \}_{\theta \in \Theta} \) constitutes a normed vector space. Of particular interest are those \( \hat{\pi} \) such that \( \hat{\pi}(\cdot|\theta) \) is a conditional probability measure. For such \( \hat{\pi} \), consider the vector \( \epsilon(\hat{\pi} - \pi) \). As the linear operator

\[
A(x, \theta) = x\theta - t(x) + h(x, \theta)
\]

is bounded, in the limit,

\[
\lim_{\epsilon \to 0} \frac{1}{\epsilon\| \hat{\pi} - \pi \|} [F(\pi + \epsilon(\hat{\pi} - \pi)) - F(\pi) - \epsilon \sum_{\theta \in \Theta} \int_X A(x, \theta) d(\hat{\pi} - \pi)(x|\theta) \mu_0(\theta)] = 0
\]

and so \( F \) is Fréchet differentiable. Hence in order to be optimal, one must have that for all conditional probability measures \( \hat{\pi} \),

\[
\sum_{\theta \in \Theta} \int_X A(x, \theta) d(\hat{\pi} - \pi)(x|\theta) \mu_0(\theta) = 0
\]

and so \( A(x, \theta) = A(x', \theta) \) almost everywhere with respect to \( \pi \). Thus (3) is necessary.

For the sufficiency of (3), suppose that \( \pi \) is suboptimal, and that instead some \( \hat{\pi} \) is better for the buyer. First, the conditional distribution \( \hat{\mu}(\cdot|x) \) must be weak* continuous with respect to \( x \) almost everywhere: suppose not, and
that there exists some point $x^*$ around which there exists $\epsilon > 0$ such that, for every $\delta > 0$, the open ball $B_\delta(x^*)$ contains two subsets of positive measure $X_1^\epsilon, X_2^\epsilon$ such that $|\mu(\cdot|x_1) - \mu(\cdot|x_2)| > \epsilon$, for all $x_i \in X_i^\epsilon$, respectively. Then for sufficiently small $\delta$, the alternative signal that recommends $x^*$ instead of any other $x \in B_\delta(x^*)$ will be an improvement, as the information cost will be strictly lower by the strong concavity of $H$, while by the compactness of $M$, the loss from recommending $x^*$ instead vanishes as $\delta \to 0$ (recalling that, by (IC-I), $t(\cdot)$ must be continuous in $x$). That is, indicating this alternative recommendation by $\hat{\pi}_\delta$, for small enough $\delta$,

$$
\sum_{\theta \in \Theta} \int_X [x\theta - t(x) + H(\frac{d\hat{\pi}_\delta(x|\theta)\mu_0(\theta)}{\sum_{\theta' \in \Theta} d\hat{\pi}_\delta(x|\theta')\mu_0(\theta')})]d\hat{\pi}_\delta(x|\theta)\mu_0(\theta) \\
- \sum_{\theta \in \Theta} \int_X [x\theta - t(x) + H(\frac{d\pi(x|\theta)\mu_0(\theta)}{\sum_{\theta' \in \Theta} d\pi(x|\theta')\mu_0(\theta')})]d\pi(x|\theta)\mu_0(\theta) \\
= \sum_{\theta \in \Theta} \hat{\pi}(B_\delta(x^*)|\theta)[x^*\theta - t(x^*)] + H(\int_{B_\delta(x^*)} d\hat{\pi}_\delta(x|\theta)\mu_0(\theta)) \\
- \sum_{\theta \in \Theta} \int_{B_\delta(x^*)} [x\theta - t(x) + H(\frac{d\pi(x|\theta)\mu_0(\theta)}{\sum_{\theta' \in \Theta} d\pi(x|\theta')\mu_0(\theta')})]d\pi(x|\theta)\mu_0(\theta) > 0
$$

Next, consider the case where $\hat{\pi}$ is absolutely continuous with respect to $\pi$. For any $\alpha \in (0, 1)$, consider the conditional probability measures $(1-\alpha)\pi + \alpha\hat{\pi}$. This will also be an improvement for the buyer over $\pi$, since

$$
\sum_{\theta \in \Theta} \int_X [x\theta - t(x) + H(\frac{d\pi(x|\theta)\mu_0(\theta)}{\sum_{\theta' \in \Theta} d\pi(x|\theta')\mu_0(\theta')})]d\pi(x|\theta)\mu_0(\theta) \\
< (1-\alpha) \sum_{\theta \in \Theta} \int_X [x\theta - t(x) + H(\frac{d\pi(x|\theta)\mu_0(\theta)}{\sum_{\theta' \in \Theta} d\pi(x|\theta')\mu_0(\theta')})]d\pi(x|\theta)\mu_0(\theta) \\
+ \alpha \sum_{\theta \in \Theta} \int_X [x\theta - t(x) + H(\frac{d\hat{\pi}(x|\theta)\mu_0(\theta)}{\sum_{\theta' \in \Theta} d\hat{\pi}(x|\theta')\mu_0(\theta')})]d\hat{\pi}(x|\theta)\mu_0(\theta)
$$
\[
\leq \sum_{\theta \in \Theta} \int_X \left[ x \theta - t(x) + H \left( \frac{((1 - \alpha)d\pi + \alpha d\hat{\pi})(x|\theta)\mu_0(\theta)}{\sum_{\theta' \in \Theta}((1 - \alpha)d\pi + \alpha d\hat{\pi})(x|\theta')\mu_0(\theta')} \right) ) \right] ((1 - \alpha)d\pi + \alpha d\hat{\pi})(x|\theta)\mu_0(\theta)
\]

(17)

where the second inequality is from merging recommendations of the same \( x \), and the fact that \( \pi \neq \hat{\pi} \) and \( H \) is concave. Subtracting (16) from (17), dividing by \( \alpha \), and taking the limit as \( \alpha \to 0 \), this becomes the Fréchet derivative as above in the direction of \( \hat{\pi} - \pi \):

\[
0 < \sum_{\theta \in \Theta} \int_X \left[ x \theta - t(x) + h(x, \theta) \right] (d\hat{\pi} - d\pi)(x|\theta)\mu_0(\theta)
\]

yielding that for some positive measure of \( x \) and \( \hat{x} \) with respect to \( \pi \) and some positive measure of \( \hat{x} \) with respect to both \( \pi, \hat{\pi} \),

\[
\sum_{\theta \in \Theta} [x \theta - t(x) + h(x, \theta)] < \sum_{\theta \in \Theta} [\hat{x} \theta - t(\hat{x}) + h(\hat{x}, \theta)]
\]

and so, for some \( \theta \),

\[
x \theta - t(x) + h(x, \theta) < \hat{x} \theta - t(\hat{x}) + h(\hat{x}, \theta)
\]

contradicting (3).

Now suppose that \( \hat{\pi} \) is singular with respect to \( \pi \). Since \( \pi \) is a recommendation strategy, for any \( x \in X \), the open ball of radius \( \epsilon \) has measure \( \pi(B_\epsilon(x)|\theta) > 0 \). Then construct the alternative measure \( \hat{\pi}_\epsilon \) defined by partitioning \([0,1]\) into intervals \( I \) of length between \( \epsilon/2 \) and \( \epsilon \) whose endpoints are not mass points of \( \hat{\pi} \), and set, for all \( x \in I \),

\[
d\hat{\pi}_\epsilon(x|\theta) = \frac{\int_{I \cap X} d\hat{\pi}(\hat{x}|\theta)}{\int_{I \cap X} d\pi(\hat{x}|\theta)} d\pi(x|\theta)
\]

Clearly, \( \hat{\pi}_\epsilon \) is absolutely continuous with respect to \( \pi \). By the compactness of \( M \) and the Portmanteau theorem,

\[
\lim_{\epsilon \to 0} \sum_{\theta \in \Theta} \int_X \left[ x \theta - t(x) + H \left( \frac{d\hat{\pi}_\epsilon(x|\theta)\mu_0(\theta)}{\sum_{\theta' \in \Theta} d\hat{\pi}_\epsilon(x|\theta')\mu_0(\theta')} \right) \right] d\hat{\pi}_\epsilon(x|\theta)\mu_0(\theta)
\]
But for low enough $\epsilon$, that would mean that $\hat{\pi}_\epsilon$ is also better than $\pi$, which we saw was impossible for any measure that is absolutely continuous with respect to $\pi$. □

**Proof of Lemma 4:** I define a system of partial differential equations defining the motion of $(x, t(x), \mu(\cdot|x))$, and show that they have a unique solution. I then verify that the necessary and sufficient conditions of Lemma 3 are satisfied.

I start by deriving a differentiable law of motion that satisfies (3), which will be used to show sufficiency. Thus I show that there exists a differentiable locus of points on which the buyer’s choice has its support; one can then convert it to a mechanism in recommendation strategies by dropping the values of $x$ that are not in the support, and invoking Lemma 3 on the remaining values of $x$ to verify that it is optimal for the buyer. First, to define $t'(x)$, any solution that is optimal for the buyer must satisfy (IC-I). It is well known from Myerson (1981) that in order to do so,

$$
\lim_{\epsilon \to 0} \frac{h(x + \epsilon, \theta) - h(x, \theta)}{\epsilon} = E_{\mu(\cdot|x)}[\theta]
$$

(18)

So, one can define

$$
\frac{\partial h}{\partial x}(x, \theta) \equiv \lim_{\epsilon \to 0} \frac{h(x + \epsilon, \theta) - h(x, \theta)}{\epsilon} = E_{\mu(\cdot|x)}[\theta] - \theta
$$

(19)

This implicitly defines the law of motion of beliefs from $\mu(\cdot|x)$. By (2), for $\mu(\cdot|x)$ to be differentiable,

$$
\frac{\partial h}{\partial x}(x, \theta) = \sum_{\theta'' \in \Theta} \frac{\partial^2 H}{\partial \mu(\theta'') \partial \mu(\theta)}(\mu(\cdot|x)) \frac{\partial \mu}{\partial x}(\theta''|x)(1 - \mu(\theta|x))
$$
\[
- \sum_{\theta'' \in \Theta} \sum_{\theta' \neq \theta} \frac{\partial^2 H}{\partial \mu(\theta'') \partial \mu(\theta')} (\mu(\cdot|x)) \frac{\partial \mu}{\partial x} (\theta''|x) \mu(\theta'|x)
\]  

(20)

Thus, for any constant \(C_{\mu(\cdot|x)}\),

\[
\sum_{\theta' \in \Theta} \frac{\partial^2 H}{\partial \mu(\theta') \partial \mu'(\theta')} (\mu(\cdot|x)) \frac{\partial \mu}{\partial x} (\theta'|x) = -(\theta + C_{\mu(\cdot|x)}), \forall \theta
\]

(21)

is a solution to (20), as by plugging these values into (18), (19) is satisfied. Since \(H\) is strongly concave, the Hessian \(H(\mu(\cdot|x))\) is negative definite, and so

\[
\begin{pmatrix}
\frac{\partial \mu}{\partial x}(\theta_1|x) \\
\vdots \\
\frac{\partial \mu}{\partial x}(\theta_K|x)
\end{pmatrix} = -H^{-1}(\mu(\cdot|x))
\begin{pmatrix}
\theta_1 + C_{\mu(\cdot|x)} \\
\vdots \\
\theta_K + C_{\mu(\cdot|x)}
\end{pmatrix}
\]

(22)

Lastly, in order to be a probability distribution, \(\sum_{\theta \in \Theta} \frac{\partial \mu}{\partial x}(\theta|x) = 0\), which means that, indicating the \((i, j)\)th entry of \(H^{-1}\) by \(H^{-1}_{(i,j)}\),

\[
C_{\mu(\cdot|x)} = -\frac{\sum_{i=1}^{K} \sum_{j=1}^{K} \theta_j H^{-1}_{(i,j)}(\mu(\cdot|x))}{\sum_{i=1}^{K} \sum_{j=1}^{K} H^{-1}_{(i,j)}(\mu(\cdot|x))}
\]

(23)

It now remains to be shown that the system of differential equations defined by (18) and (22) has a solution, in order to demonstrate that the assumption of differentiability yields a valid solution. Since \(H\) is twice Lipschitz continuously differentiable and strongly concave, \(H(\mu)\) is Lipschitz continuous in \(\mu\) and bounded away from 0, and so \(H^{-1}\) is Lipschitz continuous as well. Lastly, by (23), \(C_{\mu(\cdot|x)}\) is defined by the ratio of Lipschitz continuous functions, and so \(C_{\mu}\) is itself Lipschitz continuous in \(\mu\). By the Picard-Lindelöf theorem (Coddington and Levinson, Theorem 5.1), there exists an interval \([x-a, x+b]\) on which the system \((x, t(x), \mu(\cdot|x))\) has a unique solution.

By the fundamental theorem of calculus, it then follows that (3) is satisfied for all pairs \(x, x' \in [x-a, x+b]\). Hence any distribution \(\tau\) over \(\{\mu(\cdot|x) : x \in [x-a, x+b]\}\) is optimal for the buyer given prior \(\mu_0 = \int d\tau(\mu(\cdot|x))\) by Lemma.
3, and so (18) and (22) are sufficient for (IC-A) to be satisfied, with

$$\frac{\partial}{\partial x} \{E_{\mu(|x)}[\theta]\} = - \sum_{\theta, \theta' \in \Theta} \left[ \frac{\partial^2 H}{\partial \mu(\theta) \partial \mu(\theta')} (\mu(\cdot|x)) \right] \left( \frac{\partial \mu(\theta')}{\partial x} \right) \left( \frac{\partial \mu(\theta)}{\partial x} \right) (\theta|x) > 0 \quad (24)$$

as is easily derived from multiplying (21) by $\frac{\partial \mu(\theta|x)}{\partial x}$ and summing over $\theta$; the inequality is due to the negative-definiteness of the Hessian matrix.\(^1\)

To see that one can set $[x - a, x + b] = [0, 1]$, suppose that the maximal such value of $a$ were less than $x$. Beliefs $\mu(\cdot|x - a)$ must still be in the interior of the simplex by Lemma A since $x + b - t(x + b) - (x - a) + t(x - a) \leq b - a + \max\{\theta \in \Theta\}$. Thus, the conditions of the Picard-Lindel"of theorem are still satisfied, and so this cannot be the supremum. The same reasoning applies to $b$.

For necessity, one must show that any incentive-compatible solution to the buyer’s problem must be identical to that given above. To do so, fix $x^*$, and suppose that there exists $\tilde{\tau}$ that places positive measure, for some subset of allocations $\{x\}$, on beliefs $(\tilde{t}(x), \hat{\mu}(\cdot|x)) \neq (t(x), \mu(\cdot|x))$, where the beliefs on the right-hand side are those derived from (18) and (22). Consider the distribution $\tilde{\tau}$ over $\{\mu(\cdot|x)\}$ whose pushforward measure over $x \in [0, 1]$ is uniform. Then, by Lemma 3, $\alpha \hat{\tau} + (1 - \alpha)\tilde{\tau}$ is optimal for the buyer for any $\alpha \in (0, 1)$ given prior $\tilde{\mu}_0 = \alpha \mu_0 + \int_{(\mu(\cdot|x))} \tilde{d} \tau (\mu(\cdot|x))$. It is immediate that in order to satisfy (IC-I), the transfers conditional on $x$ must be the same under the mechanisms that generate $\hat{\tau}$ and $\tilde{\tau}$, respectively. Thus, by (2) and (3),

$$H(\hat{\mu}(\cdot|x)) + \frac{\partial H}{\partial \mu(\theta)} (\hat{\mu}(\cdot|x)) (1 - \hat{\mu}(\theta|x)) - \sum_{\theta' \neq \theta} \frac{\partial H}{\partial \mu(\theta')} (\hat{\mu}(\cdot|x)) \hat{\mu}(\theta'|x)$$

$$= H(\mu(\cdot|x)) + \frac{\partial H}{\partial \mu(\theta)} (\mu(\cdot|x)) (1 - \mu(\theta|x)) - \sum_{\theta' \neq \theta} \frac{\partial H}{\partial \mu(\theta')} (\mu(\cdot|x)) \mu(\theta'|x) \quad (25)$$

Multiplying the above by $\hat{\mu}(\theta|x)$ and $\mu(\theta|x)$, then summing over $\theta \in \Theta$ and

---

\(^1\)As remarked in the discussion following Lemma 3, any set of triplets $(x, t(x), \mu(\cdot|x))$ that satisfies (3) and on which $\tau$ has its support is incentive compatible, and so the monotonicity of $E_{\mu(|x)}[\theta]$ is implied anyway.
taking the difference between the former and the latter, one gets

\[
\sum_{\theta \in \Theta} \left( \frac{\partial H}{\partial \mu(\theta)}(\mu(\cdot|x)) - \frac{\partial H}{\partial \mu(\theta)}(\hat{\mu}(\cdot|x)) \right) (\mu(\theta|x) - \hat{\mu}(\theta|x)) = 0 \tag{26}
\]

By the intermediate value theorem, there exists some \( \alpha \in [0, 1] \) such that for \( \tilde{\mu} \equiv \alpha \mu(\cdot|x) + (1 - \alpha) \hat{\mu}(\cdot|x) \),

\[
\frac{\partial H}{\partial \mu(\theta)}(\mu(\cdot|x)) - \frac{\partial H}{\partial \mu(\theta)}(\hat{\mu}(\cdot|x)) = \sum_{\theta' \in \Theta} \frac{\partial^2 H}{\partial \mu(\theta) \partial \mu(\theta')} (\tilde{\mu})(\mu(\theta'|x) - \hat{\mu}(\theta'|x)) \tag{27}
\]

Combining (26) and (27), one gets

\[
\sum_{\theta \in \Theta} \sum_{\theta' \in \Theta} \frac{\partial^2 H}{\partial \mu(\theta) \partial \mu(\theta')} (\tilde{\mu})(\mu(\theta'|x) - \hat{\mu}(\theta'|x)) (\mu(\theta|x) - \hat{\mu}(\theta|x)) = 0
\]

But by the negative-definiteness of \( H \), the left-hand side must be negative, contradiction. □

**Proof of Theorem 1:** By Lemma 1, any contour mechanism can be implemented by recommendation strategies. Conversely, by Lemmas 3 and 4, the contour mechanism satisfies (IC-A) and (IC-I). Since \( \mathbf{t}(0) \leq 0 \) and (IC-I) is satisfied, (IR-I) is satisfied by standard arguments (e.g. Myerson, 1981). By Lemma 2, (IR-I) implies (IR-A). Hence all four constraints are satisfied □

**Proof of Proposition 1:** Immediate from (18) and (22) defining an autonomous system of differential equations. □

**Proof of Theorem 2:** I first establish that an optimal mechanism exists. It is clear that any contour mechanism’s revenue can be increased if \( \mathbf{t}(0) < 0 \), and so it is without loss of optimality to restrict attention to ones with \( \mathbf{t}(0) = 0 \). Within this set, let \( \{C_m\}_{m=1}^\infty \) be a sequence of such contour mechanisms, and let \( \tau_m \) be the corresponding distributions over posteriors. By Lemma A, there exists \( \epsilon > 0 \) such that for all \( m \), \( \mu(\theta|x) \geq \epsilon \). As shown in the proof of Lemma 4 in equations (18) and (22), the functions \( \mathbf{t}'(x) \) and \( \frac{\partial H}{\partial \mu} (\cdot|x) \) are Lipschitz continuous on any compact set in the interior of the simplex, no matter what \( \mu(\cdot|x) \) is, and so \( \{\mathbf{t}_m\} \) and \( \{\mu_m(\cdot|x)\} \) are equi-Lipschitz continuous. Therefore,
by the Arzelà-Ascoli theorem, there exists a subsequence of \( \{(C_m, \tau_m)\}_{m=1}^{\infty} \) such that \( C_m \to C \) uniformly and \( \tau_m \to \tau \) in the weak* topology, with support within the same compact set. By Coddington and Levinson, Theorem 7.1, the solutions of differential equations for a sequence of starting points converge uniformly to a solution of the differential equations for the limit point as well, so the limit values of \( (t(x), \mu(\cdot|x)) \) in \( C \) satisfy (3). Therefore \( \tau \) is an incentive-compatible distribution by Lemma 3. This implies that the set of feasible payoffs to the seller is compact, and so a maximum exists.

Given the existence of an optimal mechanism, it follows that by Theorem 1, any implementable mechanism can be expressed by some \( C \). As \( v_C(\mu) = -\infty \) for all \( \mu \) not contained in \( C \), the support of \( co(v_C) \) must be contained in \( C \) with probability 1. Hence optimization over mechanisms satisfying (8) yields the overall optimal mechanism. That \( t(0) = 0 \) follows from being able to increase \( t(x) \) by some \( \epsilon > 0 \) without violating either (IC-A) or (IR-I) for \( \mu \) otherwise.

**Proof of Corollary 1:** This follows immediately from Kamenica and Gentzkow (2011, Proposition 4 in their Online Appendix).

**Proof of Proposition 2:** Suppose that, given \( C \), some \( \tau \) is optimal such that \( x^* \equiv \sup\{x : \exists \mu \in \text{supp}(\tau) : x(\mu) = x\} < 1 \). Then the mechanism \( \hat{C} \) in which, starting from \( (x(\cdot)m\hat{t}(\cdot)) \), \( 1 - x^* \) is added to all values of \( x \leq x^* \), and all triplets corresponding to \( x > x^* \) are excluded, also satisfies (3). Thus \( \tau \) remains optimal, where the choice of \( x \) under \( \hat{C} \), \( \hat{x}(\mu) \) equals \( x(\mu) + 1 - x^* \), and \( t(x) = \hat{t}(x) \), by Proposition 1. By Lemma 4, one can then complete \( \hat{C} \) to apply to values of \( x < 1 - x^* \). Since, by (18), \( \hat{t}'(x) > 0 \), one can then increase \( \hat{t} \) by \( \int_0^{1-x^*} \hat{t}'(x)dx \) for \( \hat{x}(\mu) \geq 1 - x^* \) while maintaining (3) and (IR-I).

**Proof of Theorem 3:** For each choice of \( C \), there will either be as much information revelation as possible in the case of convex \( \hat{t} \), or none in the case of concave \( \hat{t} \), by Kamenica and Gentzkow (2011, Proposition 1). Thus it must also be true for the optimal \( C \).

**Proof of Lemma 5:** Fix \( \tau \), and suppose that it is not of the form described in the statement of the lemma. The first step is to show that there is a mean-
preserving spread of this form. With binary states, one can rewrite (12) as

$$\int_\mu^1 x(\mu)d\tau(\mu) \leq \frac{1 - [\tau(\mu < \hat{\mu})]^N}{N}$$

Differentiating this when it holds with equality, one gets

$$-x(\hat{\mu})d\tau(\hat{\mu}) = -[\tau(\mu < \hat{\mu})]^{N-1}d\tau(\hat{\mu})$$

$$\implies \tau(\mu < \hat{\mu}) = [x(\hat{\mu})]^\frac{1}{N-1}$$

$$\implies d\tau(\mu) = \frac{1}{N-1}[x(\mu)]^{\frac{1}{N-1}-1}x'(\mu)d\mu$$

with boundary condition $\tau(\mu \leq \bar{\mu}) = 1$, where $x(\bar{\mu}) = 1$. Let

$$\mu^* \equiv \inf\{\hat{\mu} : \tau(\mu < \hat{\mu}) = [x(\hat{\mu})]^\frac{1}{N-1}, \forall \hat{\mu} > \hat{\mu}\}$$

Note that (28) does not depend on the exact distribution below $\mu$. Thus, to find a mean-preserving spread, one need only consider the distribution between $\underline{\mu}$ and $\mu^*$.

I show that for any other $\tau$ satisfying (12) not of the form of the lemma, there exists a mean-preserving spread that satisfies (12); by Zorn’s lemma, there will then be a maximal element, that must be of the form of the lemma. First, suppose that there is an atom at some $\mu^* \in (\underline{\mu}, \mu^*)$. Then there for sufficiently small $\epsilon > 0$, (12) does not hold with equality at $\hat{\mu}$, $\forall \hat{\mu} \in (\mu^*, \mu^* + \epsilon)$ or else (12) would be violated at $\mu^*$.

Moreover,

$$\lim_{\epsilon \to 0} \tau(\mu \in (\mu^* - \epsilon, \mu^* + \epsilon)) = \tau(\mu^*)$$

Consider the following mean-preserving spread: replace $\tau$ by $\hat{\tau}^\epsilon$ which, for all $\mu \in [\mu^* - \epsilon^2, \mu^* + \epsilon]$, assigns all mass to $\{\mu^* - \epsilon^2, \mu^* + \epsilon\}$, while preserving $E_{\hat{\tau}^\epsilon}[\mu] = \mu_0$. By Bayes’ rule,

$$\lim_{\epsilon \to 0} \tau(\mu \in [\underline{\mu}, \mu^* - \epsilon^2]) + \frac{1}{1 + \epsilon}\tau(\mu^*) \leq \lim_{\epsilon \to 0} \hat{\tau}^\epsilon(\mu < \mu^* + \epsilon) \leq \lim_{\epsilon \to 0} \tau(\mu \in [\underline{\mu}, \mu^* + \epsilon] \setminus \{\mu^*\}) + \frac{1}{1 + \epsilon}\tau(\mu^*)$$
Since clearly
\[
\lim_{\epsilon \to 0} \tau(\mu \in [\mu, \mu_* - \epsilon^2]) = \lim_{\epsilon \to 0} \frac{1}{1 + \epsilon} \tau(\mu_* - \epsilon) = \tau(\mu_* - \epsilon)
\]
then by the squeeze theorem,
\[
\lim_{\epsilon \to 0} \hat{\tau}\epsilon(\mu < \mu_* + \epsilon) = \lim_{\epsilon \to 0} \tau(\mu < \mu_* + \epsilon)
\]
Thus \(\hat{\tau}\epsilon\) does not violate (12) at \(\mu_* + \epsilon\). For all \(\mu \leq \mu_* - \epsilon^2\), the right-hand side of (12) is the same as under \(\tau\), while by Jensen’s inequality,
\[
\int_{\mu}^{1} x(s) d\hat{\tau}\epsilon(s) \leq \int_{\mu}^{1} x(s) d\tau(s)
\]
Hence (12) is satisfied everywhere by \(\hat{\tau}\epsilon\) for \(\epsilon\) sufficiently small.

Alternatively, suppose that there are no such atoms. Then \(\tau\) is continuous for \(\mu \in (\mu, \mu^*)\). Consider \(\mu_* \in \text{supp}(\tau)\) such that \(\mu_* \in (0, \mu^*)\) and (12) does not hold with equality. By assumption, such a point exists. Then for sufficiently small \(\epsilon\), (12) does not hold with equality for all \(\mu \in (\mu_* - \epsilon^2, \mu_* + \epsilon)\). Thus the construction of the previous paragraph can be used to create a mean-preserving spread that does not violate (12) here either.

Finally, note that for a fixed \(\mu\), \(E[\mu]\) is decreasing in \(\mu^*\). There is therefore a unique \(\mu^*\) for which \(E_\tau[\mu] = \mu_0\). If one increases \(\mu\), then if \(\tau(\mu)\) does not increase as well, the new resultant distribution \(\hat{\tau}_\mu\) will strictly first-order stochastically dominate \(\tau\). As this implies \(E_{\hat{\tau}_\mu}[\mu] > \mu_0\), this is impossible. □

**Proof of Proposition 3:** By Jensen’s inequality, any mean-preserving spread of any \(\tau\) is a weak improvement for the seller. By Lemma 5, any \(\tau\) has a feasible mean-preserving spread unless it satisfies (12) with equality above some \(\mu^*\), and no other posterior aside from \(\mu\) is in the support. Hence some such \(\tau\) will be optimal. That this can be implemented by a second-price auction with a reserve price \(r\) can be seen by setting \(r = \int_{\mu}^{\mu^*} \hat{v}(\mu) d\mu\) and using the revenue equivalence theorem (Myerson, 1981). □
Before presenting the proofs of Proposition 4 and Theorem 4, I introduce some additional notation and a useful lemma, analogous to Lemma 5. Consider the pushforward measure $\sigma$ as generated by $x(\mu)$ where $\mu$ is distributed according to $\tau$. One can then write (12) as

$$\int_{x^*}^1 x d\sigma(x) \leq \frac{1 - \sigma(x < x^*)}{N}, \forall x^* \in [0, 1]$$  \hspace{1cm} (29)

**Lemma B:** For any $\sigma$ satisfying (29), there exists a mean-preserving spread $\hat{\sigma}$ over $x \in [0, 1]$ that

(i) satisfies (29) with equality between some $x^*$ and 1;

(ii) sets $\sigma((0, x^*)) = 0$; and

(iii) has an atom at $x = 0$.

**Proof:** Suppose that (29) is satisfied for all $x \geq x^*$. As in the proof of Lemma 5, it is easy to show that in order to find a mean-preserving spread, one need only consider the distribution between 0 and $x^*$, since (29) for $x > x^*$ does not depend on the exact distribution of lower values, but only on their cumulative distribution up to $x$.

If there is an atom at some $x^*_* \in (0, x^*)$, then for sufficiently small $\epsilon > 0$, (29) does not hold with equality at $\hat{x}$, $\forall \hat{x} \in (x^*_*, x^*_* + \epsilon)$, or else (29) would be violated at $x^*_*$ itself. Moreover,

$$\lim_{\epsilon \to 0} \sigma(x^*_* - \epsilon, x^*_* + \epsilon) = \sigma(x^*_*)$$

Consider the following mean-preserving spread: replace $\sigma$ with $\hat{\sigma}^\epsilon$, which, for all $x \in [x_\epsilon - \epsilon^2, x_\epsilon + \epsilon]$, assigns all mass to $\{x_\epsilon - \epsilon^2, x_\epsilon + \epsilon\}$, while preserving $E_{\hat{\sigma}^\epsilon}[x] = E_{\sigma}[x]$. By Bayes’ rule,

$$\lim_{\epsilon \to 0} \sigma([0, x_\epsilon - \epsilon^2]) + \frac{1}{1 + \epsilon} \sigma(x_\epsilon) \leq \lim_{\epsilon \to 0} \hat{\sigma}^\epsilon([0, x_\epsilon + \epsilon]) \leq \lim_{\epsilon \to 0} \sigma([0, x_\epsilon + \epsilon]\{x_\epsilon\}) + \frac{1}{1 + \epsilon} \sigma(x_\epsilon)$$

$$\implies \lim_{\epsilon \to 0} \hat{\sigma}^\epsilon([0, x_\epsilon + \epsilon]) = \lim_{\epsilon \to 0} \sigma([0, x_\epsilon + \epsilon])$$

and so $\hat{\sigma}^\epsilon$ does not violate (29) at $x_\epsilon + \epsilon$. For all $x \leq x_\epsilon - \epsilon^2$, the right-hand
side of (29) is the same as under $\sigma$, while $\int_x^1 s \hat{\sigma}'(s) = \int_x^1 s \sigma(s)$. Thus, (29) is satisfied everywhere for $\hat{\sigma}'$ for $\epsilon$ sufficiently small.

Now suppose instead that there are no such atoms. Then $\sigma$ is continuous for $x \in (0, x^*)$. Consider $x_* \in \text{supp}(\sigma)$ such that $x_* \in (0, x^*)$ and (29) does not hold with equality. By assumption, such a point exists. Then, for sufficiently small $\epsilon$, (29) does not hold with equality for all $x \in (x_* - \epsilon^2, x_* + \epsilon)$. Thus the construction of the previous paragraph can be used to create a mean-preserving spread that does not violate (29) here either.

By Zorn’s lemma, there then exists a maximal mean-preserving spread, which must satisfy (i)-(iii). □

**Proof of Proposition 4:** Since $H$ is quadratic, $H$ is independent of $\mu$. By (22) and (23), this means that $\frac{\partial \mu}{\partial x}(\theta|x)$ is constant, i.e. not dependent on $x$ or $\mu$. Thus, for any contour mechanism $C$, all values of $\mu(\cdot|x)$ are linear in $x$. By (24), so is $E_{\mu(\cdot|x)}[\theta]$, and as a result by (18) $t$ is quadratic in $x$, with initial conditions $t(0) = 0$ and $t'(0) = E_{\mu}[\theta]$. Letting $\sigma$ be the pushforward measure over $X$ defined by $\tau$ and $x(\mu)$, any mean-preserving spread $\hat{\sigma}$ over $X$ also defines a mean-preserving spread $\hat{\tau}$ over $\mu$ given $C$, and vice versa. Any such mean-preserving spread increases the seller’s expected payoff due to $t(x)$ being quadratic in $x$ (and hence convex). By Lemma B, a maximal mean-preserving spread places an atom at $x = 0$ while satisfying (12) with equality for all $x > x^*$ for some $x^*$, while placing measure 0 on $x \in (0, x^*)$. By the revenue equivalence theorem of Myerson (1981), this can be implemented by a second-price auction with a reserve price. □

**Proof of Theorem 4:** (i) The information acquisition cost is given by

$$c(\tau_N) = \int [H(\mu_0) - H(\mu)]d\tau_N(\mu)$$

By (12), the buyer’s probability of winning $E_{\tau_N}[x_N(\mu)] \to 0$, so her expected utility converges to 0 as well. Thus (with some abuse of notation), $\tau_N \to \delta_{\mu_0}$ in the weak* topology, where $\delta_{\mu_0}$ is the Dirac measure that places probability 1 on $\mu_0$. Therefore, $E_{\mu}[\theta] \to E_{\mu_0}[\theta]$.  

15
(ii) Again, by (12), \( E_{\tau_N}[x_N(\mu)] \to 0 \). By Proposition 1, \( x'(\mu) \) is determined for any \( \mu \) regardless of \( \mu \). By (2) and (3), \( \frac{\partial \mu}{\partial x}(\theta|x=0) \) is continuous in \( \mu \) since \( H \) is twice continuously differentiable, and so \( x'(\mu) \) is uniformly continuous on any closed ball \( B \) around \( \mu_0 \) such that \( B \) is in the interior of the simplex. As shown above, for sufficiently large \( N \), \( \tau_N(\mu \in B) \to 1 \), so \( \tau_N \to \delta_{\mu_0} \); by (12), \( |\tau_N - \delta_{\mu_0}| \to 0 \) in the weak* topology, where \( \delta_{\mu_0} \) is the Dirac measure that places probability 1 on \( \mu_0 \). By the triangle inequality from (i), this means that \( \mu_N \to \mu_0 \).

(iii) Fix function \( t(x) \). Since \( E_{\mu(\cdot|x)}[\theta] \) is strictly increasing in \( x \) by (24), \( t(x) \) will be a strictly convex function by (18). Hence by Jensen’s inequality, for any \( \sigma \) that does not satisfy the properties of Lemma B, there exists \( \hat{\sigma} \) that satisfies the properties in Lemma B such that \( \int_0^1 t(x)d\hat{\sigma}(x) > \int_0^1 t(x)d\sigma(x) \). As in the proof of Proposition 3, any \( \sigma \) that satisfies these properties can be implemented by a second-price auction with reserve price \( r = t(x^*) \) by the revenue equivalence theorem of Myerson (1981).

Next, for any fixed \( t \), the distribution \( \sigma \) satisfying the properties in Lemma B that maximizes \( \int_0^1 t(x)d\sigma(x) \) is that which sets \( x^* = 0 \), as for any other value, the distribution over \( x \in [x^*,1] \) would remain unchanged by setting \( x^* \) instead. Since \( t \) is a strictly increasing function and the new distribution first-order stochastically dominates the old one, this increases \( \int_0^1 t(x)d\sigma(x) \). Thus, for fixed \( t(\cdot) \), a second-price auction with a reserve price of 0 is optimal.

I now show that in the limit as \( N \to \infty \), there is a unique limit value \( t(x) \) of any implementable sequence of \( \{ t_N(x) \}_{N=1}^\infty \), and so one will be able to invoke the above result to conclude that this form of auction is optimal. First, consider the sequence of distributions \( \{ \tau_N \} \) and their pushforward measures \( \{ \sigma_N \} \). For sufficiently high \( N \), there exists Bayes-plausible \( \hat{\tau}_N \) such that its pushforward measure \( \hat{\sigma}_N \) satisfies the properties in Lemma B and is a mean-preserving spread of \( \sigma_N \), with some corresponding value of \( x^* \). To see this, by Coddington and Levinson, Theorem 7.6, for any \( \epsilon > 0 \) there exists \( \delta > 0 \) such that if \( \mu \in B_\delta(\mu_0) \) (the closed ball of radius \( \delta \) around \( \mu_0 \) in the simplex), then the solutions for \( (t(x),\mu(\cdot|x)) \) under \( \mu = \mu \) differ from those under \( \mu = \mu_0 \) by
at most \(\epsilon\) in the Euclidean topology. Consider the function

\[
\phi_N(\mu) = \mu + \frac{1}{2}[\mu_0 - \int_0^1 \mu(\cdot|x)d\hat{\tau}_N(\mu(\cdot|x))]
\]

Clearly, \(\phi_N(\mu) = \mu\) if and only if \(\int_0^1 \mu(\cdot|x)d\hat{\tau}_N(\mu(\cdot|x)) = \mu_0\). As \(\mu(\cdot|x)\) is uniformly continuous in \(\mu \in \bar{B}_\delta(\mu_0)\), it follows that for \(N\) large enough, \(\|\mu - \int_0^1 \mu(\cdot|x)d\hat{\tau}_N(\mu(\cdot|x))\| < \delta\) by (12) and (22) for all \(\mu \in B_\delta(\mu_0)\), as \(\tau\) converges to the Dirac measure on \(\mu\) by (ii). Hence, by the triangle inequality,

\[
|\mu_0 - \phi_N(\mu)| \leq \frac{1}{2}|\mu_0 - \mu| + \frac{1}{2} \left| \int_0^1 \mu(\cdot|x)d\hat{\tau}_N(\mu(\cdot|x)) - \mu \right|
\]

\[
\leq \frac{1}{2}\delta + \frac{1}{2}\delta = \delta
\]

and so \(\phi_N(\mu) \in \bar{B}_\delta(\mu_0)\). Since \(\phi_N(\mu)\) is continuous, by the Brouwer fixed point theorem there exists \(\mu \in B_\delta(\mu_0)\) such that \(\phi_N(\mu) = \mu\), which implies that \(\int_0^1 \mu(\cdot|x)d\hat{\tau}_N(\mu(\cdot|x)) = \mu_0\) as required. Thus, given \(\tau_N\) and \(\sigma_N\), there exist such \(\hat{\tau}_N\) and \(\hat{\sigma}_N\), respectively, for high enough \(N\).

Let \(t_N\) and \(\hat{t}_N\) be the corresponding transfer functions. Consider any subsequence such that \(\sigma_N \rightarrow \sigma\) and \(\hat{\sigma}_N \rightarrow \hat{\sigma}\) in the weak* topology. For any \(y\), by the Portmanteau theorem,

\[
\int_0^y \sigma([0,x])dx \leq \liminf \int_0^y \sigma_N([0,x])dx \leq \liminf \int_0^y \hat{\sigma}_N([0,x])dx = \int_0^y \hat{\sigma}([0,x])dx
\]

where the last holds with equality because either \(\hat{\sigma}\) is absolutely continuous (if \(x^* = 0\)) or \(\hat{\sigma}([0,x^*]) = \hat{\sigma}(x = 0)\). Thus, \(\hat{\sigma}\) is a mean-preserving spread of \(\sigma\). Moreover, by the Lipschitz continuity of \(H\), both \(t_N \rightarrow t_{\mu_0}\) and \(\hat{t}_N \rightarrow t_{\mu_0}\) uniformly on \([0,1]\), where \(t\) is defined for the contour starting at \(\mu = \mu_0\) (Coddington and Levinson, Theorem 7.1). Since \(t\) is also continuous, by the Portmanteau theorem and the dominated convergence theorem,

\[
\lim_{N \to \infty} \int_0^1 Nt_{\mu_0}(x)d\sigma(x) = \lim_{N \to \infty} \int_0^1 Nt_N(x)d\sigma_N(x)
\]
assuming that \( \lim_{N \to \infty} \int_0^1 N t_{\mu_0}(x) d\sigma_N(x) \) is finite. Differentiating (29) when it holds with equality at \( x \) yields

\[
x = \left[ \hat{\sigma}_N((0, x)) \right]^{N-1}
\]

\[
\Rightarrow \frac{d\hat{\sigma}_N}{dx}(x) = \frac{(x)^{\frac{2-N}{N-1}}}{N-1} \leq \frac{2}{N x}
\]

Indeed,

\[
\lim_{N \to \infty} N \frac{d\hat{\sigma}_N}{dx}(x) = \frac{1}{x}
\]

Since, by (18),

\[
x \cdot \min\{\theta \in \Theta\} \leq t(x) \leq x \cdot \max\{\theta \in \Theta\}
\]

by the dominated convergence theorem we have (even for \( x^* = 0 \), by defining for each \( N \) at the limit as \( x^* \to 0 \))

\[
\int_{x^*}^1 N t_{\mu_0}(x) d\sigma_N(x) \leq \int_{x^*}^1 2 \max\{\theta \in \Theta\} d\sigma(x)
\]

\[
\Rightarrow \lim_{N \to \infty} \int_{x^*}^1 N t_{\mu_0}(x) d\sigma_N(x) = \int_{x^*}^1 \frac{t_{\mu_0}(x)}{x} dx
\]

As observed earlier, for fixed \( t(\cdot) \), setting \( x^* = 0 \) is optimal. Therefore, any mechanism in the limit is dominated by a second-price auction with reserve price 0, which yields the revenue as given in (13). \( \square \)