Almost-Truthful Mediation Enables Information Exchange between Agents with Opposing Interests

Dmitry Sedov*

*Northwestern University, Evanston, USA

May 12, 2021

Abstract

Information necessary for decision-making is often distributed among agents with opposing interests. In such settings receiving information is desirable for the agents, while revealing it may be privately harmful. This paper constructs a class of almost-truthful mediation protocols that incentivize information exchange in a succinct model capturing such conflicts. The protocols in this class receive signal reports from the agents and send messages back, almost always transmitting the received reports without any distortions. The rare distorted messages, however, are deliberately designed to prevent deviations from truth-telling. Specifically, the mediator's distortions encourage each agent to take the action that is interim-optimal given her private information as reflected by her signal report. Thus, in case of distorted messages, a truthful agent acts on private information only, while the action of a deviating agent is shifted away from the truly interim-optimal one. As a result, the deviating agent is put to a disadvantage when the mediator distorts the signals, which is enough to ensure truthful communication when the misalignment of interests between the agents is sufficiently small. Through its distinctive way of eliciting the truth, almost-truthful mediation highlights the role of information design in facilitating communication.

JEL classification: C72; D82; D83.

Keywords: Communication; Competition; Information; Mediation; Cheap talk; Mechanism design.

E-mail address: dsedov@u.northwestern.edu

Address for correspondence: Department of Economics, Northwestern University, 2211 Campus Drive, Evanston, IL 60208-2600, USA.

Acknowledgements The author thanks Timur Abbiasov, Yingni Guo, Gaston Illanes, Sergei Izmalkov, Riccardo Marchingiglio, Alessandro Pavan, Robert Porter, Dmitry Sorokin, Xavier Vives, Asher Wolinsky, Andrey Zhukov and seminar participants at Northwestern University for helpful comments and discussions. Special thanks to Francisco Poggi and Quitzé Valenzuela-Stookey. The author is also grateful for the reviews received on an earlier version of this paper.

This paper previously circulated under the title "Information Exchange between Competitors". The author gratefully acknowledges financial support received from the Russian Science Foundation [project №15-18-30081] when that version of the paper was in progress.

1 Introduction

Information relevant to decision-making is often distributed among agents with opposing interests. For illustration, consider the following examples. Pharmaceutical companies possess only partial data on drugs under development, but prefer to be the sole owners of the final formula. In knowledge-intensive organizations complementary information is fragmented and held by employees who are frequently motivated by relative evaluation. Intelligence agencies gather incomplete material about the suspects during joint investigations, but are also involved in competition for influence and authority. In these cases, *receiving* information is privately desirable for the agents, while *revealing* it may be privately harmful.

When such a conflict is present, how can mutually advantageous information exchange be organized? The present paper answers this question in a succinct model of distributed information and opposing interests. The model features a one-stage game without monetary transfers, in which agents possess private signals, take actions and receive payoffs that depend on the combination of signals and actions. Agents lack commitment power and may have opposing interests regarding each other's actions. Moreover, the actions are neither strategic complements nor strategic substitutes, which eliminates coordination motives for truthful communication. The usual incentives for information transmission are thus ruled out, direct information exchange cannot be sustained in equilibrium, and information design is isolated as the sole means of enabling communication.

This paper shows that a special class of almost-truthful mediation protocols can facilitate communication between the agents. Protocols in this class are designed to receive private signal reports from the agents and to send messages back so that the response message to a given agent almost always contains the actual signal report of the other agent. With a small probability, however, the response message to a given agent is distorted. The distorted message depends on that agent's own signal report only and does not contain any additional information. The two types of messages (accurate and distorted) take value in the same set, and thus the agents can not distinguish them with certainty. In fact, by the almost-truthful design of the protocol, each agent is almost certain the response message contains the other agent's actual signal report. The undistorted messages lead to complete information revelation. The distorted messages are designed to encourage each agent to take an action that is interim-optimal given her private information as reflected by her signal report. When the mediator uses a distorted message, a non-deviating agent essentially acts on her private information only. A deviating agent, however, shifts her action away from the action that is interim-optimal with respect to her true private signal. As a result, the distorted message hurts the agent more if she deviates rather than not. This ensures that revealing private information truthfully is optimal from the perspective of agent's own actions. While a deviating agent may still benefit from the change in the other party's action caused by the deviation, incentives for truthful communication dominate if the misalignment of interests between the agents is small enough. Thus, an equilibrium with truth-telling exists. Importantly, the ability of the mediation protocols to shift actions, so that the deviating agent is put to a disadvantage, relies on three details. First, the almost-truthful design of the protocols guarantees that agents' posterior beliefs regarding the counterpart's signal place most of the weight on the mediator's message. Second, the deviating and non-deviating behavior need to result in different distorted messages by the mediator, which is guaranteed, if agents' interim beliefs are sufficiently sensitive to their private information. Third, the mediator's messages actually shift agents' decisions, which relies on the assumption that each agent's optimal action is sensitive to the counterpart's information. These three details (a design feature and two assumptions) are crucial for the mediated communication scheme developed in this

paper.
The rest of the paper is organized as follows. Section 2 briefly reviews the relevant literature. Section 3 provides an illustrative example revealing the intuition behind the main result. Section 4 presents the baseline model, proves the non-existence of direct communication, characterizes the class of almost-truthful mediation protocols and demonstrates that such protocols enable information transmission. Section 5 concludes by summarizing the results and discussing applications to the settings of sharing clinical trials data, communication in organizations and exchange of information between intelligence agencies.

2 Literature review

Two branches of research on transmission of unverifiable information are relevant in the context of the present paper.
First, there is a vast literature on communication with informed agents not being able to influence decisions directly. Unmediated cheap talk has been shown to allow information transmission in the seminal paper by Crawford and Sobel (1982). Moreover, important spin-offs have been explored, including, but not limited to, multiple senders by Austen-Smith (1993) and Krishna and Morgan (2001), multiple receivers by Farrell and Gibbons (1989) and Goltsman and Pavlov (2011), multiple rounds of communication by Krishna and Morgan (2004) and Goltsman et al. (2009), unbounded multidimensional state space by Battaglini (2002), bounded multidimensional state space by Ambrus and Takahashi (2008) and communication error by Blume et al. (2007). Communication through a mediator has been explored as well. Among others, Goltsman et al. (2009) characterize the optimal mediated communication in the canonical CS setting. Ivanov (2010) and Ambrus et al. (2013) explore communication via strategic mediators. In the settings listed above truthful information transmission is incentivized by potential actions of the decision-maker. That is, informed agents prefer to tell the truth, since the corresponding action is preferred. These results are based on the existence of some common interest shared by informed agents and decision-makers, i.e. there exist actions over which the preferences of the sender and the receiver coincide. In the present paper the possibility of communication does not rely on the existence of common interests[1]. The only incentive for communication is the higher benefit of information received through the mediation protocol in case of truth-telling (i.e. lower downside of the distorted messages).
Second, communication between partially informed agents, who also take actions, is studied in a number of papers. Galeotti et al. (2013) extend the CS model to the case of multiple decision-makers with private information regarding the state of nature. In their model telling lies is again precluded by the corresponding unfavorable change in other agents' actions. Alonso et al. (2008) model communication between partially informed managers who care about the profits of own divisions and action coordination. Communication via cheap talk is possible in that model, since the managers prefer their actions to be close to each other and thus have an incentive to reveal some private information. In the industrial organization context Goltsman and Pavlov (2014) look into the case of communication between Cournot oligopolists[2], who may share unverifiable private

[1] In fact, the preferences may be completely opposite: the payoff structure, such that an increase in one agent's utility may always lead to the decrease of the other agent's utility, is allowed.

[2] Further examples of the literature on communication in oligopoly include Novshek and Sonnenschein (1982), Vives (1984), Gal-Or (1985), Li (1985), Shapiro (1986), Vives (1990) and Raith (1996). See Kühn and Vives (1995) and Vives (2001) for extensive reviews. This strand of research typically assumes commitment power or verifiable private

information about costs, and show that no information transmission occurs in the cheap-talk game, but information can be transmitted through a neutral third party. The question explored by Goltsman and Pavlov (2014) is similar to that of the present paper, but in their case the competitor's information is relevant, because it affects her action and actions are strategic substitutes. The mediator is able to exploit the coordination motives to achieve communication: some types of agents report truthfully in order to make the opponent less aggressive. The incentives for truth-telling provided by the optimal mediation protocol in the present paper are purely informational: reporting truthfully leads to a higher benefit of information received back. The "secret sharing" game presented as an example in Vida and Forges (2013) has a similar structure to the illustrative example in this paper. However, the example setting in Vida and Forges (2013) features (i) information that is independent across players; (ii) communication between the players relying on the availability of verifiable "signatures" that allow to detect deception of each individual agent; (iii) the possibility of full information transmission. In the present paper, instead, (i) players' types can be correlated; (ii) the mediation protocol solely relies on the information structure and is still able to provide incentives for truth-telling; (iii) only partial information transmission is possible. Kolotilin et al. (2017) explore persuasion of a privately informed receiver. Similarly, in the present paper, after observing a report from one of the agents, the mediation protocol essentially tries to persuade the other one with a caveat that the other agent is informed herself. Kolotilin et al. (2017) show that private persuasion is equivalent to public persuasion, which is not the case in the present paper: the mediation protocol *does* need to condition her recommendation on agents' reports in order to sustain informative communication.

To sum up, the present paper contributes to the literature on communication between agents with misaligned interests by considering a setting that differs from the ones discussed above in terms of assumptions, results and intuition.

3 Illustrative example

This section introduces an example capturing the main intuition of the results in the present paper. In the example players receive binary signals and need to guess the mean of the two signals. While direct communication is not possible, the class of almost-truthful mediation protocols is shown to enable information exchange when (i) signals are correlated, and (ii) the misalignment of interests is sufficiently small.

3.1 Setup

Consider the following game . Each of the two agents $: 2 \{1,2\}$ obtains a binary signal $B_: 2 S = \{0,1\}$ with the following joint distribution c over $S^2 = S \quad S$ parametrized by

information. A notable exception is Ziv (1993), who shows that conveying credible information in the oligopoly setting is also possible if "money-burning" or transfers are allowed. In the present paper agents lack commitment power, information is non-verifiable, and both "money-burning" and transfers are assumed out.

$A \in [1 \bullet 2, 1^{o}$:

	c	
	$s_2 = 0$	$s_2 = 1$
$s_1 = 0$	$\frac{A}{2}$	$\frac{1 \quad A}{2}$
$s_1 = 1$	$\frac{1 \quad A}{2}$	$\frac{A}{2}$

Together, these signals determine the correct action $B = \frac{1}{2} \quad _{:} B_{:} \bullet$ Both agents would like to guess B by choosing an action in the set $A_{:} = \{0, 1 \bullet 2, 1\} \bullet$ The agents have opposing interests and prefer the opponent not to be able to guess the correct action. The payoffs representing such preferences are given by

$$D_{:}(a_{:}, a_{3 \quad :}, B) = \mathbb{1}\{a_{:} = B\} \quad U \quad \mathbb{1}\{a_{3 \quad :} = B\} \tag{1}$$

where $U \in [0, 1]$ parametrizes the degree of interest misalignment between the agents: the higher U the more each player is hurt by the competitor's correct guess.

Notice that under no communication the optimal strategy of each agent is choosing an action that coincides with the observed $O_{:} = B_{:} \bullet$ Also note that the lack of communication is suboptimal: if agents were able to disclose signals to each other, the expected payoffs in the game would go from $A(1 \quad U)$ to $(1 \quad U)$. Proposition A.1 and Proposition A.2 formally establish these two results in Appendix A.

However, the agents are not able to communicate in a game with simultaneous message exchange. If there was a message one agent could send and shift the action of the counterpart, such message would be used to deceive the counterpart. The deceitful agent could benefit from the counterpart's mistake, and would suffer no losses due to the lack of coordination motives and the unchanged counterpart's messaging strategy. See Proposition A.3 of Appendix A for a formal treatment.

3.2 Almost-truthful mediation

While direct communication is impossible, this subsection introduces the notion of almost-truthful mediation, which facilitates information exchange for low enough misalignment of interests.

Mediation setup The almost-truthful mediation protocol receives signal reports from the agents and sends messages back to them. It almost always sends agent $(3 \quad :)$'s signal report to agent $:$ without any distortion. However, with a positive probability Y the mediation protocol returns to agent $:$ a message that coincides with $(3 \quad :)$'s most likely signal given $:$'s own report. Formally, let $\lambda_{:}(B_{:}, B_{3 \quad :})$ be a collection of binary distributions over mediator's messages to agent $:$ when the mediator receives reports $B_{:}, B_{3 \quad :}$ from the agents:

$$\lambda_{:}(B_{:}, B_{3 \quad :}) = \begin{array}{ll} B_{3 \quad :} & \text{with probability} \quad 1 \quad Y \\ B_{:} & \text{with probability} \quad Y \end{array}$$

where

$$Y \in \left[0, \frac{1 \quad A}{A}\right]$$

Truthful equilibrium Such a mediation protocol ensures that there exists an equilibrium in which both agents transmit their information truthfully. Two observations are necessary for this result.

First, if agent $^{1}3\ :^{o}$ reports truthfully, agent $:$ chooses the average between the private signal and the mediator's message as her action irrespective of whether $:$ herself reports truthfully. To see this, note that the posterior probability on the event $B_{3\ :} = <$ where $<$ is the observed mediator's message is greater than $1\bullet2\bullet$ Indeed, if the mediator's message does not coincide with $:$'s report, $< < B_{:}$- then agent $:$ for sure knows that $< = B_{3\ :}$ and $\mathrm{P}_{B_8\ B_{3\ :}=<\mathrm{j}<_{:}{}^{1}B_{:}\text{-}B_{3\ :}{}^{o}=<\text{-}B_{:}} = 1\bullet$ If the mediator's message does coincide with $:$'s report, $< = B_{:}$- then

$$\mathrm{P}_{B_8\ B_{3\ :}=<\mathrm{j}<_{:}{}^{1}B_{:}\text{-}B_{3\ :}{}^{o}=<\text{-}B_{:}} = \frac{\mathrm{P}\ <_{:}{}^{1}B_{:}\text{-}B_{3\ :}{}^{o}=<\mathrm{j}B_{3\ :}=<\text{-}B_{:}\ \mathrm{P}\ B_{3\ :}=<\mathrm{j}B_{:}}{\mathrm{P}\ <_{:}{}^{1}B_{:}\text{-}B_{3\ :}{}^{o}=<\mathrm{j}B_{:}}$$

$$= \begin{cases} \dfrac{A}{A\ ,\ Y^{1}1\quad A^{o}} & \text{if } B_{:}=B_{:} \\ \dfrac{1\quad A}{^{1}1\quad A^{o}\ ,\ YA} & \text{if } B_{:}<B_{:} \end{cases}$$

$$i\ \frac{1}{2}\text{-}$$

since

$$Y\ \ddot{Y}\ \frac{1\quad A}{A}$$

As a result, for low values of Y the agent optimally chooses action $O_{:} = \frac{1}{2}\,{}^{1}B_{:}\ ,\ <_{:}{}^{1}B_{:}\text{-}B_{3\ :}{}^{oo}$, the average of the private signal and mediator's message.

Second, since $:$'s optimal actions conditional on deviating and not are known, the consequences of deviating and not are simple to predict. If the mediator does not distort the information, agent $:$ chooses action $B\ \bullet$ If the mediator distorts the information agent $:$ chooses the interim-optimal action $B_{:}$ in case she doesn't deviate and chooses action $1\bullet2$ in case she does deviate. That is, if $:$ doesn't deviate, she makes a mistake when the mediator distorts and her interim-optimal action is not ex-post optimal, which happens with probability $^{1}1\quad A^{o}Y$. If $:$ deviates, she makes a mistake when the mediator distorts and her interim-optimal action was actually ex-post optimal, which happens with probability AY. Thus the expected *partial payoff from $:$'s own actions* in case of a truthful report equals $+^{)} = 1\quad {}^{1}1\quad A^{o}Y$ and equals $+\quad = 1\quad AY$ in case of a deviation.

It remains to notice that the expected partial payoff in case of reporting truthfully is higher: $+ = +^{)}\quad +\quad = Y^{1}2A\quad 1^{o}\ i\ 0\bullet$ Provided that U is low enough, the positive partial difference $+$ dominates the total payoff difference, and reporting truthfully is optimal for each agent. The existence of a truthful equilibrium is thus shown.

Highlighting incentives and assumptions First, the mediator is able to control agent's beliefs by being sufficiently truthful. Such control provides the mediator with the opportunity to deliberately shift agents' actions when the mediator distorts. Second, when the mediator does distort the information, the action of the deviating player is shifted away from the interim-optimal action, while the action of the non-deviating player is not. As a result, the mediator hurts the deviating player more, when distorting the information: effectively, the deviating player imposes an inefficient action upon herself and is put to a disadvantage.

This last point depends on the assumption of $A\ i\ 1\bullet2$- which imposes positive correlation between the agents' signals. If $A = 1\bullet2$ (signals are uncorrelated), then the mediator can not shift the action away from the interim-optimal action[3]. Thus the mediator's distortion is equally harmful for the agent irrespective of whether she reports truthfully or not, and communication breaks down. In fact $A = 1\bullet2$ is problematic in two ways: (i) non-uniqueness of interim-optimal action, (ii)

[3]Which is not unique: both the private signal $B_{:}$ and $1\bullet2$ are optimal actions for every agent.

different types have same beliefs about the counterpart's signals. While there is no distinction between the two in the illustrative example, the main result of this paper will separately impose the assumptions of uniqueness (see Assumption 4.1 and Assumption 4.2) and sensitivity of beliefs to private information (see Assumption 4.4). Jointly these assumptions guarantee that the mediator's messages can shift agents' actions.

Additionally, the payoff structure in (1) ensures that the agent's optimal action is sensitive to the counterpart's information. If this was not the case, agents would have little incentive to report their signals truthfully as getting additional information from the counterpart would be worthless.

4 Main result

This section generalizes the illustrative example by constructing a model that allows for arbitrary joint distributions over states of nature. When agents' preferences are separable in actions, the almost-truthful mediation enables communication, while direct communication is impossible. Sufficient conditions for the almost-truthful mediation protocol allowing truthful information exchange are established. These conditions are: (i) action sensitivity to counterpart's information, (ii) sufficient variation in interim beliefs across agent types, and (iii) sufficiently weak misalignment of interests.

4.1 Model

The model consists of two agents who are endowed with private information regarding the state of nature and have to take an action. Each action affects the payoffs of both agents. A form of additive separability is assumed: the action's effect on the other party's payoff only depends on the action itself and the state of nature, but not on that other party's action.

Baseline Formally, consider a 2-agents setup with a finite state space $\Omega = \Omega_1 \times \Omega_2$. Each agent i learns the realization of $\theta_i \in \Omega_i$ (the signal), but does not learn θ_{-i}. Let $p(\cdot)$ be the common prior over Ω. Also, let $p_i(\cdot) = p(\cdot \mid \theta_i)$ and $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot \mid \theta_i]$ be agent i's posterior and expectation operator upon learning θ_i respectively. Each agent i chooses an action a_i from a finite action space A_i. It is assumed that the agents' preferences over action profiles are represented by a state-dependent utility function

$$u_i(a, \theta) = v_i(a_i, \theta) + \lambda \cdot v_{-i}(a_{-i}, \theta)$$

$v_i(a_i, \theta)$ captures agent i's value from taking action a_i in a given state $\theta \in \Omega$. Without loss of generality assume $v_i(a_i, \theta) > 0$ for every i, a_i and θ. The preferences of agent i with respect to $(-i)$'s actions in a given state are captured by the cost function $\lambda \cdot v_{-i}(a_{-i}, \theta)$. The cost function v_{-i} captures the idea of agents' opposing interests: as in the illustrative example above, a change in agent $(-i)$'s action that is beneficial for agent $(-i)$ is allowed to increase the cost function v_{-i} (and thus to be harmful) for agent i. The cost component of the utility function is further parametrized by the parameter $\lambda \geq 0$ that captures the degree of the misalignment of interests[4]. It

[4]This paper is primarily motivated by the situations where agents have opposing interests, but the model formally allows for aligned agents' preferences as well (i.e. a change in one player's action being beneficial for both players). Exploiting the specifics of such an alignment of interests can, in principle, be used to facilitate communication, but constructing such setting-specific schemes is beyond the scope of the present paper. The mediation protocol class introduced further in the paper can enable information exchange in the situations of aligned interests (under a set of assumptions), but does not depend on such an alignment.

is also worth noting that the separability of agents' utility functions in each other's actions implies that actions are neither strategic complements nor strategic substitutes. This eliminates the option to reveal information about the counterpart's action as a potential leverage that the mediator can use to elicit the truth (this leverage is used, for example, in Goltsman and Pavlov (2014)).

Definitions The definitions that simplify the notation in the remaining part of the paper are now introduced. First, in a fixed state B in S each agent i can maximize her payoff by taking the same action for all actions of $\neg i$. That is, due to the assumption of separability of agents' preferences with respect to each other's actions, agent i can choose the action that maximizes the E_i-component of her utility. Definition 4.1 below introduces formal notation for such state-specific *correct actions*:

Definition 4.1. For agent i in state $B \in \mathsf{S}$ let $O_i(B)$ be the set of state-specific *correct actions*. That is, $O_i(B) = \arg\max_{O_i} E_i(O_i, O_{-i}, B)$.

Similarly, for every privately observed state $B_i \in \mathsf{S}_i$ the *interim-correct actions* are defined:

Definition 4.2. For each agent i and signal $B_i \in \mathsf{S}_i$ let $O(B_i)$ be the set of *interim-correct actions*. That is, $O_i(B_i) = \arg\max_{O_i} \mathsf{E}_{\cdot} \left[E(O_i, O_{-i}, B) \mid B_i \right]$.

Definition 4.2 captures the notion of the best possible actions in autarky. Such actions would be taken by each agent in the absence of any information exchange. Definition 4.3 below introduces the notion of *interim-correct counterpart signals* that links correct actions and interim-correct actions.

Definition 4.3. For each agent i and signal $B_i \in \mathsf{S}_i$ let $T_i(B_i)$ be the set of *interim-correct counterpart signals*. That is, $T_i(B_i) = \{ B_{-i} \mid O_i(B_i) = O_i(B_i, B_{-i}) \}$.

Specifically, if agent $\neg i$'s signal was revealed to belong to the set $T_i(B_i)$ then agent i endowed with signal B_i would have no incentive to take an action other than the interim-correct one. At this point, $T_i(B_i)$ can be an empty set, Assumption 4.2 below ensures it is non-empty.

A preliminary result The following lemma establishes a natural result that will be useful throughout the rest of the paper. It states that an action that is correct for a given signal of the counterpart will be chosen for high enough belief on this signal.

Lemma 4.1. *Let σ_i be agent i's belief over S_{-i}. There exists a $\bar{X}_i$ such that for all B_{-i} if $\sigma_i(B_{-i}) > \bar{X}_i$ then* $\arg\max_{O_i} \mathsf{E}_{\sigma_i} \left[E_i(O_i, O_{-i}, B) \right] = O_i(B_i, B_{-i})$.

Proof. Notice that for $\sigma_i = 1$ and every $O \in O_i(B_i, B_{-i})$

$$\begin{aligned}
\mathsf{E}_{\sigma_i} \left[E_i(O, O_{-i}, B) \right] &= \sum_{C_{-i}} \sigma_i(C_{-i}) E_i(O, O_{-i}, B) \\
&= E_i(O, O_{-i}, B) \\
&> E_i(O', O_{-i}, B) \\
&= \sum_{C_{-i}} \sigma_i(C_{-i}) E_i(O', O_{-i}, B) \\
&= \mathsf{E}_{\sigma_i} \left[E_i(O', O_{-i}, B) \right]
\end{aligned}$$

for every $O' \notin O_i(B_i, B_{-i})$ by Definition 4.1. Thus by continuity of $\sum_{C_{-i}} \sigma_i(C_{-i}) E_i(O, O_{-i}, B)$ with respect to $\sigma_i(C_{-i})$ there exists a $\bar{X}_i(B_{-i})$ such that the same strict inequality holds for all σ_i such that $\sigma_i(B_{-i}) > \bar{X}_i(B_{-i})$. The proof of the lemma is completed by defining $\bar{X}_i = \max_{B_{-i}} \bar{X}_i(B_{-i})$.

4.2 Assumptions

The additional assumptions stated below limit the scope of the results of the present paper to settings in which the prior has full support and the action space is of intermediate coarseness. As will be clear from the main result of the paper, intermediate coarseness guarantees two things. First, actions can be shifted by additional information. Second, concealment of information can appear like provision of additional information. It should be noted that while the assumptions below impose restrictions on endogenous objects, such a presentation leads to a succinct description of the setup's features that are necessary for the main result of the current paper.

First, an assumption regarding the structure of the correct action set is made.

ASSUMPTION 4.1. (i) $O_i(B)$ is a singleton for every i and $B \in S$. (ii) For every agent i and signal B_i if $B'_{3\,i} < B_{3\,i}$ then $O_i(B_i, B'_{3\,i}) < O_i(B_i, B_{3\,i})$.

Part (i) precludes the existence of actions that lead to the same consequences in a given state and can be interpreted as a "no redundant actions" requirement. Part (ii) is a sensitivity assumption which ensures that the action space is rich enough so that the choice of action can be adjusted for alternative states of nature.

Then, an assumption regarding the structure of the interim-correct set is made.

ASSUMPTION 4.2. (i) $O_i(B_i)$ is a singleton for every i and $B_i \in S_i$. (ii) For every agent i and signal B_i there exists $B_{3\,i}$ such that $O_i(B_i) = O_i(B_i, B_{3\,i})$.

Part (i) of Assumption 4.2 is a joint assumption on the action space and the information structure. It ensures no indifferences on the interim stage. Under Assumption 4.1 this assumption is guaranteed to hold when the posteriors c_i are close enough to degenerate ones: $c_i(|B_i) > 0$, for the true signal $B_{3\,i}$ only[5]. The results of the present paper are thus guaranteed to hold when agents' signals exhibit a sufficient degree of dependence.

Part (ii) of Assumption 4.2 is a coarseness assumption which guarantees that the interim action can not be perfectly adjusted to the non-degenerate interim information of the agents. It is guaranteed to hold if there are no "redundant" actions in the action set ($A_i = \{O_i(B) | B \in \Omega\}$ for every i, and if the E_i-component is monotone with respect to own-signal (for every agent i, signals $B_i, B_{3\,i}$ and $B'_{3\,i}$, $E_i(O_i(B_i, B'_{3\,i}), O_i(B_i, B_{3\,i})) > E_i(O_i(B'_i, B'_{3\,i}), O_i(B_i, B_{3\,i}))$). Note that Assumption 4.1 and Assumption 4.2 jointly ensure that the set of interim-correct counterpart signals $\tau_i(B_i)$ is a singleton for all agents i and private states B_i.

Finally, the assumption regarding the information structure is made.

ASSUMPTION 4.3. (i) $|S_1| = |S_2|$. (ii) $c(B) > 0$ for all $B \in S$.

Jointly, these assumptions can be interpreted as similarity of agents' private signal quality. Part (i) of Assumption 4.3 limits the scope of the paper to settings where each agent can receive the same number of different signals. It will help ensuring that different types of each agent hold sufficiently different beliefs about the counterpart's signals[6]. Part (ii) of Assumption 4.3 prevents complete elimination of uncertainty and thus each type of every agent is *not* entirely informed about the state of nature upon realization of the private signal.

[5] To see this formally, one can utilize Lemma 4.1.

[6] The setup under part (i) of Assumption 4.3 can be interpreted as follows. Agents agree on the possible states of the world: $S_1 = S_2 = \Omega$, and receive noisy signals about the true state with the signal space coinciding with the states of the world set.

4.3 Impossibility of direct communication

Can the two agents help each other directly by simultaneously sending messages that are at least partially informative? Intuitively, this should not be the case. Conditional on receiving a message from the counterpart and the corresponding posterior over the counterpart's signal (which governs the action choice), each agent has no incentives to send truthful information. This section establishes the impossibility of direct communication.

Proposition 4.1. *There exist cost functions* $2_i(\cdot)$ *such that for any direct communication extension* W_i *of the baseline setup only interim-optimal actions are taken.*

Proof. Consider the case of "extreme" conflict: $2_i(\theta_3) = B^0 = E_{3\,i}(\theta_3) = B^0$. Consider an extended game in which agents can simultaneously send each other a message from a predetermined set W_i upon observing the private signal. Suppose that there is an equilibrium in which agent $(3\,i)^0$s expected value of the own-choice component $E_{3\,i}$ is strictly greater than under no communication. Then there exists a type of agent i and a message F_i sent in equilibrium such that some type of agent $(3\,i)^0$ is expected to shift her action away from the interim-optimal one conditional on receiving such a message. Then consider the following deviation: send F_i irrespective of own type and follow the same action plan (possibly contingent on received messages) as in equilibrium. Such a deviation is profitable, as it ensures that the expectation of $E_{3\,i}$ is strictly lower than under no communication, and thus the expectation of $U2_i = UE_{3\,i}$ is higher. The expectation of $E_{3\,i}$ is decreased under such a deviation since at least one type of $(3\,i)^0$ shifts the action away from the interim-optimal irrespective of i's true information.

4.4 Almost-truthful interim-biased mediation

This subsection introduces interim-biased and almost-truthful interim-biased protocol classes. Such protocols take signal reports from both players and send private messages to each agent. The set of messages M_i to agent i coincides with the set $S_{3\,i}$. This captures the notion that the mediator may share some of $(3\,i)^0$'s private information with agent i.

Interim-biased mediation protocols introduced in Definition 4.4 below are randomized for each pair of signal reports. Such protocols transmit agent $(3\,i)^0$'s report to agent i with probability $1 - \gamma_i$ and send a message with i's interim-correct counterpart signal with the complementary probability γ_i.

Definition 4.4. Let an *interim-biased mediation protocol* τ be a collection of random variables $\tau^1_{i\,i=1,2}$ with

$$\tau^1_i(\theta_i, \theta_{3\,i}) = \begin{cases} \theta_{3\,i} & \text{with probability } 1 - \gamma_i \\ f_i(\theta_i)^0 & \text{with probability } \gamma_i \end{cases} \tag{2}$$

for some $\gamma_i \in [0, 1/4]$.

The following lemma establishes that there exist positive probabilities γ_i such that, conditional on agent $(3\,i)^0$ reporting truthfully, agent i believes that the mediator's message coincides with $(3\,i)^0$'s report irrespective of i's own report. Formally,

Lemma 4.2. *For every agent* i *and* $\theta_i \neq \theta'_i$ *there exists an* $\gamma_i(\theta_i)^0 > 0$ *such that if an interim-biased mediation protocol satisfies* $\gamma_i < \gamma_i(\theta_i)^0$ *then* $P(\theta_{3\,i} = \tau_i(\theta_i, \tau^1_i(\theta_i, \theta_{3\,i})^0) = \tau(\theta'_i, \tau^1_i(\theta_i, \theta_{3\,i})^0) = \tau(\theta'_i, \theta_{3\,i}) \mid \theta_i)$ *for all* $\theta_i, \theta'_i \in S_i$ *and* $\tau \in S_{3\,i}$.

Proof. Consider the beliefs of agent i upon reporting signal s_i and receiving message m from the mediator.

1. If $m < f_i(s_i)$ then the posterior probability that agent $j \neq i$'s signal report is equal to t can be computed as follows.

$$P(s_j = t \mid s_i, m = f_i(s_i)) = \frac{P(m = f_i(s_i) \mid s_i, s_j = t)\, P(s_j = t \mid s_i)}{P(m = f_i(s_i) \mid s_i)} = \frac{(1 - \varepsilon_i)\, P(s_j = t \mid s_i)}{P(m = f_i(s_i) \mid s_i)}$$

Notice that given Definition 4.4, $P(m = f_i(s_i) \mid s_i, s_j = t) = (1 - \varepsilon_i)$ and thus $P(s_j = t \mid s_i, m = f_i(s_i)) = 1 > \lambda$. Consequently, for the lemma to be true only the case of $m = f_i(s_i)$ needs to be considered.

2. If $m = f_i(s_i)$ then

$$P(s_j = t \mid s_i, m = f_i(s_i)) = \frac{1}{1 + \sum_{\theta_j \neq t} \frac{1}{\gamma_i(\theta_j \mid s_i, m)}},$$

where

$$\gamma_i(\theta_j \mid s_i, m) = \frac{\pi(s_j = \theta_j \mid s_i)}{\pi(s_j = t \mid s_i)}$$

Note that $\gamma_i(\theta_j \mid s_i, m)$ is a positive finite number for every m and θ_j under Assumption 4.3. Thus if ε_i satisfies

$$\varepsilon_i \leq \lambda_i = \min_{s_i \in S_i} \left\{ \frac{1 - \lambda}{\lambda} \left(\sum_{\theta_j \neq t} \frac{1}{\gamma_i(\theta_j \mid s_i, m)} \right)^{-1} \right\} = \frac{1 - \lambda}{\lambda} \min_{s_i \in S_i} \frac{\pi(s_j = t \mid s_i)}{1 - \pi(s_j = t \mid s_i)}, \tag{3}$$

then $P(s_j = t \mid s_i, m = f_i(s_i)) > \lambda$ for every $s_i \in S_i$.

The proof is completed by observing that $\lambda_i \in (0, 1)$.

Now the class of *almost-truthful interim-biased mediation protocols* is introduced. Such protocols belong to the interim-biased mediation class and share a defining common feature: the probability $1 - \varepsilon_i$ of truthful transmission of each agent's report is close to 1. Formally,

Definition 4.5. Let an *almost-truthful interim-biased mediation protocol* σ^0 be a collection of random variables $\{\sigma^0_i\}_{i=1,2}$ with $\sigma^0_i(s_i, s_j) = \sigma^1_i(s_i, s_j)$ for some $\varepsilon_i \in (0, \bar{\varepsilon}_i)$.

Notice that due to Lemma 4.1 and Lemma 4.2, each almost-truthful interim-biased mediation ensures agents i's posterior belief has a high enough weight on t when the mediator's message is t. Thus agents take action $a_i(s_i) = \theta_i$ upon receiving t from the mediator. Importantly, this fact only depends on agent $j \neq i$ (but not agent i) reporting truthfully. Lemma 4.3 below states this result formally.

Lemma 4.3. *For each agent i's signal s_i, signal report s_i and realization t of mediator's message $\sigma^1_i(s_i, s_j)$, agent i's optimal action coincides with the correct action in state (s_i, θ_j):* $\arg\max_{a_i} E_{\theta_j}\, u_i(a_i(s_i), \theta_j \mid s_i, \sigma^0_i(s_i, s_j) = t) = a_i(s_i, \theta_j)$.

Proof. By Definition 4.5 agent i's posterior belief over agent $j \neq i$'s signal places a higher than $\bar{\varepsilon}_i$ weight on t. By Lemma 4.2, $\arg\max_{a_i} E_{\theta_j}\, u_i(a_i(s_i), \theta_j \mid s_i, \sigma^0_i(s_i, s_j) = t) = a_i(s_i, \theta_j)$.

4.5 Equilibrium with information exchange

This subsection establishes that protocols from the almost-truthful interim-biased class allow truthful information exchange between the agents. The additional required assumption ensures enough variation in intermediate beliefs across different types of agents.

First, the additional assumption is stated.

ASSUMPTION 4.4. For each agent : and pair of signals B⁰: < B:- Ť:¹B⁰:º < Ť:¹B:º•

This assumption is interpreted as follows: the interim-correct actions are rationalized by different counterpart's signals. Notice that Assumption 4.4 implies jS₁j = jS₂j- a feature of the setting that was assumed earlier under Assumption 4.3. Assumption 4.4 is guaranteed to hold if receiving different signals leads to sufficiently different beliefs about the opponents' signals. For example, if the posterior weight on the most likely opponent's signal is sufficiently strong and under different observed signals different counterpart's signals are most likely, then the result of Lemma 4.1 and Assumption 4.4 hold.

Next, the E:-component of the utility is shown to be strictly maximized by truthful reporting. Let +<0:¹B:-B:º be the expectation of agent :'s E:-component of the utility conditional on reporting B: to an almost-truthful mediation protocol <0: when the true signal is B: and agent ¹3 :º reports truthfully. Notice that by Lemma 4.3,

+<0:¹B:-B:º = Õ_{C3 :} c:¹C3 :jB:º ¹1 Y:º E:¹0:¹B:-C3 :º-¹B:-C3 :ºº
¸ Õ_{C3 :} c:¹C3 :jB:º Y: E:¹0:¹B:-Ť:¹B:ºº-¹B:-C3 :ºº

The following lemma establishes that truthful reporting maximizes the E:-component of the utility:

LEMMA 4.4. *Suppose that Assumption 4.4 holds. For each agent :- signals B: < B: and mediation protocol <0- +<0:¹B:-B:º ¡ +<0:¹B:-B:º•*

Proof. Define +<0:¹B:-B:º = +<0:¹B:-B:º +<0:¹B:-B:º• Notice that

+<0:¹B:-B:º =
= Õ_{C3 :} c:¹C3 :jB:º E:¹0:¹B:-C3 :º-¹B:-C3 :ºº »Y: Y:¼
¸ Õ_{C3 :} c:¹C3 :jB:º Y: E:¹0:¹B:-Ť:¹B:ºº-¹B:-C3 :ºº E:¹0:¹B:-Ť:¹B:ºº-¹B:-C3 :ºº
= Y: Õ_{C3 :} c:¹C3 :jB:º E:¹0:¹B:-Ť:¹B:ºº-¹B:-C3 :ºº E:¹0:¹B:-Ť:¹B:ºº-¹B:-C3 :ºº

Rearranging,

+<0:¹B:-B:º = Y: Õ_{C3 :} c:¹C3 :jB:º E:¹0:¹B:-Ť:¹B:ºº-¹B:-C3 :ºº
Y: Õ_{C3 :} c:¹C3 :jB:º E:¹0:¹B:-Ť:¹B:ºº-¹B:-C3 :ºº

First notice that 0:¹B:-Ť:¹B:ºº = 0:¹B:º by Definition 4.2 and part (i) of Assumption 4.2. Next, by Assumption 4.4 different intermediate actions are rationalized by different counterpart's signals and thus Ť:¹B:º < Ť:¹B:º• Also, by Assumption 4.1 the optimal action varies for different signals of the counterpart and thus 0:¹B:-Ť:¹B:ºº < 0:¹B:-Ť:¹B:ºº. Since by Assumption 4.2 there

is a unique interim-correct action, $O_i(\tau_i(B_i), \tau_{-i}(B_{-i})) < O_i(\tau_i(B_i), B_{-i}) = O_i(\tau_i(B_i), \tau_{-i}(B_{-i}))$ Combining, for $O_i^0 = O_i(\tau_i(B_i), \tau_{-i}(B_{-i}))$,

$$+_{\sigma_i}(\tau_i(B_i), B_{-i}) = \underbrace{Y_i}_{\geq 0} \underbrace{\mathbb{E}_{c_i}\left[E_i(O_i(\tau_i(B_i), B_{-i}), B)\right] - \mathbb{E}_{c_i}\left[E_i(O_i^0, B)\right]}_{\geq 0} \geq 0$$

where the first inequality $Y_i \geq 0$ is ensured by Definition 4.5 of almost-truthful mediation protocols, and the second inequality $\mathbb{E}_{c_i}\left[E_i(O_i(\tau_i(B_i), B_{-i}), B)\right] - \mathbb{E}_{c_i}\left[E_i(O_i^0, B)\right] \geq 0$ is ensured by Definition 4.2 of the interim-optimal action.

Consider now an extended game in which agents simultaneously send reports to an almost-truthful mediation protocol, observing the private signal, receive the mediator's messages back and then simultaneously choose actions. Let $*_{\sigma_i}(\tau_i(B_i), B_i)$ be the expectation of agent i's utility conditional on reporting B_i to an almost-truthful mediation protocol σ_i^0 when the true signal is B_i. Let $_{\sigma_i}(\tau_i(B_i), B_i)$ be the expectation of agent i's 2_i-component of the utility conditional on reporting B_i to an almost-truthful mediation protocol σ_i^0 when the true signal is B_i.

For low enough conflict of interest, as parametrized by U- reporting truthfully is strictly optimal in such an extended game:

Theorem 4.1. *Suppose that Assumption 4.4 holds. There exists an $\bar{U}$ such that for all $U \in [0, \bar{U})$- for each agent i- signal $B_i \in \mathcal{B}_i$ and almost-truthful mediation protocol σ_i^0- $*_{\sigma_i}(\tau_i(B_i), B_i) \geq *_{\sigma_i}(\tau_i(B_i), B_i)$.*

Proof. Define $*_{\sigma_i}(\tau_i(B_i), B_i)$ and $_{\sigma_i}(\tau_i(B_i), B_i)$ analogously to $+_{\sigma_i}(\tau_i(B_i), B_i)$. Notice that

$$*_{\sigma_i}(\tau_i(B_i), B_i) = +_{\sigma_i}(\tau_i(B_i), B_i) - U \cdot {}_{\sigma_i}(\tau_i(B_i), B_i)$$

Let

$$\underline{+}_{\sigma_i} = \min_{B_i \in \mathcal{B}_i} \left[+_{\sigma_i}(\tau_i(B_i), B_i)\right]$$

and

$$\overline{}_{\sigma_i} = \max\left[0, \max_{B_i \in \mathcal{B}_i} {}_{\sigma_i}(\tau_i(B_i), B_i)\right]$$

Notice that $\underline{+}_{\sigma_i} \geq 0$ by Lemma 4.4 and $\overline{}_{\sigma_i} > 0$ by construction. Define

$$\bar{U} = \begin{cases} \infty & \text{if } \forall i: \overline{}_{\sigma_i} = 0 \\ \min_i \left[\dfrac{\underline{+}_{\sigma_i}}{\overline{}_{\sigma_i}}\right] & \text{if } \exists i: \overline{}_{\sigma_i} < 0 \end{cases}$$

It remains to notice that for $U \in [0, \bar{U})$- $*_{\sigma_i}(\tau_i(B_i), B_i) \geq 0$ for every i and $B_i \in \mathcal{B}_i$- which completes the proof.

Lemma 4.4 and Theorem 4.1 share the main intuition with the illustrative example in Section 3. Lemma 4.4 establishes each agent can optimize her own action by reporting truthfully. The reasons for that are: (i) the agents trust the mediator's message since it is almost always truthful, (ii) the mediators distortions harm a deviating agent more by shifting her action away from the interim-optimal action. Theorem 4.1 establishes that when the misalignment of interests is small enough, the incentives to report truthfully dominate as optimizing own action is relatively more important than benefiting from a shift in the counterpart's actions.

On information structure and almost-truthful mediation As stated in Section 4.2, the assumptions imposed on the setting in order to show that almost-truthful mediation enables information exchange require that the action space is of intermediate coarseness. A similar observation can be made regarding the prior distribution c• Intuitively, c should exhibit an intermediate degree of dependence across its two dimensions, S_1 and S_2- in order for a given almost-truthful mediation protocol to work.

To see this, first fix an U and an almost-truthful mediation protocol characterized by Y that allows for information exchange. Consider a small increase in $c_{:}{}^{1}f_{:}{}^{1}B_{:}{}^{o}jB_{:}{}^{o}$ for all $B_{:}$• Notice that $Y_{:}$ defined in (3) necessarily goes down. There is thus no guarantee that the Y-protocol ensures a high enough posterior belief weight on its message among the agents. Consequently, there is no guarantee that the Y-protocol still allows information exchange under the new prior c• On the other hand, consider the case when almost-truthful mediation is impossible due to the failure of Assumption 4.4. Increasing the conditional beliefs $c_{:}{}^{1}\,jB_{:}{}^{o}$ on the most likely counterpart's signal given $B_{:}$ can restore Assumption 4.4 and thus lead to the possibility of communication through a given almost-truthful mediation protocol.

Increasing the degree of dependence between the two dimensions of c can thus make a given protocol less trustworthy, but can also make the interim beliefs more sensitive to signals. Informally, this makes intermediate degree of dependence across c's two dimensions most suitable for almost-truthful mediation.

4.6 Almost-truthful mediation can be optimized with a linear program

The almost-truthful mediation can be generalized, allowing the probability $Y_{:}$ of information distortion to depend on the report profile submitted by the agents:

DEFINITION 4.6. Let a *generalized almost-truthful interim-biased mediation protocol* be a collection of random variables

$$<_{:}^{6}{}^{1}B_{:}\text{-}B_{3\ :}{}^{o} = \begin{array}{ll} B_{3\ :} & \text{with probability } 1\quad Y_{:}{}^{1}B_{:}\text{-}B_{3\ :}{}^{o} \\ f_{:}{}^{1}B_{:}{}^{o} & \text{with probability } Y_{:}{}^{1}B_{:}\text{-}B_{3\ :}{}^{o}\text{-} \end{array}$$

such that $Y_{:}{}^{1}B_{:}\text{-}B_{3\ :}{}^{o}\ 2\ \ 0\text{-}\bar{Y}_{:}{}^{1}\bar{X}_{:}{}^{o}$ for all $B_{:}\text{-}B_{3\ :}$•

Proposition 4.2. *The Pareto-optimal generalized almost-truthful interim-biased mediation protocol can be found in polynomial time with a linear programming problem.*

Proof. Similarly to previous cases, let $*_{<_{:}^{6}{}^{1}B_{:}\text{-}B_{:}{}^{o}}$ be the expectation of agent $:$'s utility conditional on reporting $B_{:}$ to a generalized almost-truthful mediation protocol $<_{:}^{6}$ when the true signal is $B_{:}$ and agent $^{1}3\quad :{}^{o}$ reports truthfully. Notice that

$$\begin{aligned} *_{<_{:}^{6}{}^{1}B_{:}\text{-}B_{:}{}^{o}} = & \sum_{C_{3\ :}} c_{:}{}^{1}C_{3\ :}jB_{:}{}^{o}\quad {}^{1}1\quad Y_{:}{}^{1}B_{:}\text{-}C_{3\ :}{}^{oo}\quad E_{:}{}^{1}O_{:}{}^{1}B_{:}\text{-}C_{3\ :}{}^{o}\text{-}{}^{1}B_{:}\text{-}C_{3\ :}{}^{oo} \\ & + \sum_{C_{3\ :}} c_{:}{}^{1}C_{3\ :}jB_{:}{}^{o}\quad Y_{:}{}^{1}B_{:}\text{-}C_{3\ :}{}^{o}\quad E_{:}{}^{1}O_{:}{}^{1}B_{:}\text{-}f_{:}{}^{1}B_{:}{}^{oo}\text{-}{}^{1}B_{:}\text{-}C_{3\ :}{}^{oo} \\ & + \sum_{C_{3\ :}} c_{:}{}^{1}C_{3\ :}jB_{:}{}^{o}\quad {}^{1}1\quad Y_{3\ :}{}^{1}C_{3\ :}\text{-}B_{:}{}^{oo}\quad 2_{:}{}^{1}O_{:}{}^{1}B_{:}\text{-}C_{3\ :}{}^{o}\text{-}{}^{1}B_{:}\text{-}C_{3\ :}{}^{oo} \\ & + \sum_{C_{3\ :}} c_{:}{}^{1}C_{3\ :}jB_{:}{}^{o}\quad Y_{3\ :}{}^{1}C_{3\ :}\text{-}B_{:}{}^{o}\quad 2_{:}{}^{1}O_{:}{}^{1}f_{3\ :}{}^{1}C_{3\ :}{}^{o}\text{-}C_{3\ :}{}^{o}\text{-}{}^{1}B_{:}\text{-}C_{3\ :}{}^{oo} \end{aligned}$$

is linear in the Y-parameters. Thus the Pareto-optimization problem with weights _:-

$$\max_{fY:^{1}-^{o}g} \; \_:C:^{1}B:^{o} * <6^{1}B:-B:^{o}$$
$$\text{subject to} \quad * <6^{1}B:-B:^{o} \; 6 \; * <6^{1}B:-B:^{o}$$

is a linear program and can be solved by the standard methods (and in polynomial time), see Vanderbei (2014) and references therein.

Remark 4.1. *Appendix B finds the optimal mediation protocol for the illustrative example of Section 3 and demonstrates that it belongs to the class of generalized almost-truthful interim-biased mediation protocols. The optimal mediation protocol sends distorted messages relatively more often when the signals reported by the agents are jointly unlikely. This feature leads to stricter informational punishments for deviations, while permitting more accurate information transmission when the agents report truthfully.*

5 Conclusion

This section concludes the paper by discussing the applications, providing a summary of the results and outlining directions for future research.

5.1 Applications

The mediation protocol introduced in Definition 4.5 suggests potentially useful insights for arranging information exchange in settings where competition between different parties is impossible (or too costly) to avoid by changing an organizational structure altogether.

Consider the case of pharmaceutical companies mentioned in the introduction. While in the recent years there has been some movement both by regulators and private entities in the direction of sharing clinical trials data[7], it is still recognized that there are significant challenges to such a practice, in particular, there exist risks of commercial misuse of the shared data and the proper policies to mitigate such risks are still to be developed. For example, a report by the Institute of Medicine (2015, p. 142)[8] acknowledges that (i) "sponsors of clinical trials have serious concerns about competitors copying data packages that lack strong regulatory data protection" and (ii) "what types of policies might be implemented to protect data from "unfair commercial use" is a difficult question", listing registration-only or online-only access to shared information and data request reviews "by an independent third party" as potential solutions. The intuition developed in this paper suggests that a central authority gathering the data from competing companies and distributing the results with a small bias to each company's private report may provide appropriate motivation to share data in addition to the policies recommended by the Institute of Medicine (2015).

Many organizations try to encourage knowledge sharing among their employees[9]. Appropriately

[7] See, for example, the Wall Street Journal article by Burton (2016) on the expansion of requirements for publication of clinical studies' results and the New York Times publication by Thomas (2012) on GlaxoSmithKline's effort to open up its drug research, which resulted in the clinical trials data request system being set up at `www.clinicalstudydatarequest.com`.

[8] The Institute of Medicine was renamed to the National Academy of Medicine in 2015.

[9] See, for example, the Harvard Business Review discussion of knowledge sharing by Myers (2015) and the Wall Street Journal article on knowledge hiding by Deal (2018).

structured compensation schemes, along with trust, corporate culture and management support have been recognized as important forces that drive information exchange in knowledge-intensive enterprises, see Wang and Noe (2010) for an extensive review. The mediation protocol designed in the present paper indicates that organizational leaders can also act as key communication intermediaries for their subordinates, who possess some complementary knowledge and engage in competition, and hints at a strategy of information dissemination. In particular, transmitting full information in most cases, and sometimes encouraging the employees to act on private information may work as the proper mediation tactic. While some information is transferred, the rare encouragement to act on private information only can ensure that the employees who engage in knowledge hiding are put to a disadvantage. Such implicit penalty can thus prevent improper communication, importantly, even if the organizational leader is unable to establish the quality of the reports directly. Finally, consider the case of competing intelligence agencies. An example of such a situation is described by a think-tank Council on Foreign Relations (2006) which claims that "fundamental cultural differences and turf wars have long hindered cooperation between the two agencies [FBI and CIA]". This conflict may have contributed to the fact that "agencies did not adequately share relevant counter-terrorism information, prior to September 11", see Finding 9 of the Joint Inquiry into Intelligence Community Activities (2002). As a response to this concern, the Intelligence Reform and Terrorism Prevention Act of 2004 established the position of Director of National Intelligence, the responsibilities of whose Office include overseeing the Information Sharing Environment that facilitates exchange of intelligence across various governmental agencies. The concern about the difficulty of providing proper incentives for sharing information across competing agencies is still relevant, see the discussions by Garicano and Posner (2005, p. 161) and the current head of the U.S. State Department's Bureau of Counterterrorism, Sales (2010, p. 281). Modest implications of the results in this paper point to (i) the disincentives to share complete intelligence information in the presence of inter-agency competition; and (ii) Information Sharing Environment's potential use as an imperfect information transmission filter, establishing proper incentives for knowledge sharing. When considering these applications, one should keep in mind that the mediation protocol class constructed in this paper relies on the assumption of the separability of agents' utility functions in each other's actions. Thus the insights apply best to settings where agents can not mitigate each other's actions directly but are affected by such actions (companies operating in separate markets, employees working on different projects, intelligence agencies investigating parallel cases). At the same time, one should also keep in mind that the construction presented in this paper is primarily theoretical and demonstrates that appropriately designing an information environment can facilitate communication.

5.2 Concluding remarks

This paper studies communication between partially informed agents with opposing interests. For such agents receiving information is desirable, while revealing it may be privately harmful. The paper (i) offers a simple model that captures the main attributes of such a tradeoff; (ii) characterizes the class of almost-truthful mediation protocols that enable communication provided that the misalignment of interests between the agents is sufficiently small, while beliefs and actions are sufficiently sensitive to information; (iii) highlights the main leverage that allows communication: almost-truthful information transmission provides the mediator with the opportunity to distort actions in a deliberate manner so that a deviating agent is put to a disadvantage; (iv) suggests applications of the results to the settings of clinical trials data exchange, organizational knowledge hiding and sharing of intelligence information.

Three generalizations of this paper's results can be interesting. First, relaxing the separability of preferences in agents' actions may be desirable. Provided that the component of utility related to the interplay of agents' actions is sufficiently small compared to the $E_{\cdot}{}^{1}O_{\cdot}{}^{-B^{0}}$ action-state component of the utility, a similar argument to Theorem 4.1 seems feasible. Next, increasing the number of agents in the model would lead to relaxed truth-telling incentives as individual deviations are easier to detect under correlated signals. However, it is interesting, whether any adjustments to the mediation design of the present paper (that does not rely on detecting deviations) are needed to extend the possibility result to multi-player settings. Finally, generalizing the class of almost-truthful mediation protocol to a continuous state and action space with an appropriate payoff structure is another potential direction for further research. This would require ensuring that (i) the agents cannot distinguish between accurate and distorted messages of the mediator (e.g. the distorted message can't be a non-random function of the agent's report); (ii) the deviating agent is shifted away from the interim-optimal action *more*[10] than the truthful agent when distorted messages are sent. While formalizing these arguments is interesting, it is outside of the scope of the present paper, which provides a possibility result and highlights intuition.

References

Alonso, R., W. Dessein, and N. Matouschek (2008). When Does Coordination Require Centralization? *American Economic Review 98*(1), 145–79.

Ambrus, A., E. M. Azevedo, and Y. Kamada (2013). Hierarchical Cheap Talk. *Theoretical Economics 8*(1), 233 – 261.

Ambrus, A. and S. Takahashi (2008). Multi-Sender Cheap Talk with Restricted State Spaces. *Theoretical Economics 3*(1), 1–27.

Austen-Smith, D. (1993). Interested Experts and Policy Advice: Multiple Referrals under Open Rule. *Games and Economic Behavior 5*(1), 3 – 43.

Battaglini, M. (2002). Multiple Referrals and Multidimensional Cheap Talk. *Econometrica 70*(4), 1379 – 1401.

Bergemann, D. and S. Morris (2018). Information Design: A Unified Perspective. *Journal of Economic Literature*. Forthcoming.

Blume, A., O. J. Board, and K. Kawamura (2007). Noisy Talk. *Theoretical Economics 2*(4), 395 – 440.

Burton, T. M. (2016, September). New U.S. Rule to Expand Requirements for Publication of Clinical Trials. *The Wall Street Journal*. Accessed: 2018-09-26. Archived at https://web.archive.org/web/20171113163510/https://www.wsj.com/articles/new-u-s-rule-to-expand-requirements-for-publication-of-clinical-trials-1474040304.

Council on Foreign Relations (2006, January). FBI and Law Enforcement. Accessed: 2017-11-19. Archived at https://web.archive.org/web/20171119172108/https://www.cfr.org/backgrounder/fbi-and-law-enforcement.

Crawford, V. P. and J. Sobel (1982). Strategic Information Transmission. *Econometrica 50*(6), 1431 – 1451.

Deal, J. (2018, August). How Leaders Can Stop Employees from Deliberately Hiding Information. *The Wall Street Journal*. Accessed: 2018-10-14. Archived at https://web.archive.org/web/20181015152315/https://blogs.wsj.com/experts/2018/08/14/how-leaders-can-stop-employees-from-deliberately-hiding-information/.

Farrell, J. and R. Gibbons (1989). Cheap Talk with Two Audiences. *The American Economic Review 79*(5), 1214 – 1223.

[10] In the continuous action space both agents' optimal actions conditional on the available information will reflect the possibility that the mediator's message was distorted.

Forges, F. (1986). An Approach to Communication Equilibria. *Econometrica 54*(6), 1375–1385.

Gal-Or, E. (1985). Information Sharing in Oligopoly. *Econometrica 53*(2), 329–343.

Galeotti, A., C. Ghiglino, and F. Squintani (2013). Strategic information transmission networks. *Journal of Economic Theory 148*(5), 1751 – 1769.

Garicano, L. and R. A. Posner (2005). Intelligence Failures: An Organizational Economics Perspective. *Journal of Economic Perspectives 19*(4), 151–170.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009). Mediation, Arbitration and Negotiation. *Journal of Economic Theory 144*(4), 1397 – 1420.

Goltsman, M. and G. Pavlov (2011). How to Talk to Multiple Audiences. *Games and Economic Behavior 72*(1), 100 – 122.

Goltsman, M. and G. Pavlov (2014). Communication in Cournot Oligopoly. *Journal of Economic Theory 153*, 152 – 176.

Institute of Medicine (2015). *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. Washington, D.C.: The National Academies Press.

Ivanov, M. (2010). Communication Via a Strategic Mediator. *Journal of Economic Theory 145*(2), 869 – 884.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017). Persuasion of a Privately Informed Receiver. *Econometrica 85*(6), 1949–1964.

Krishna, V. and J. Morgan (2001). A Model of Expertise. *The Quarterly Journal of Economics 116*(2), 747 – 775.

Krishna, V. and J. Morgan (2004). The Art of Conversation: Eliciting Information from Experts through Multi-Stage Communication. *Journal of Economic Theory 117*(2), 147 – 179.

Kühn, K.-U. and X. Vives (1995). *Information Exchange Among Firms and Their Impact On Competition*. Luxembourg: Office for Official Publications of the European Communities.

Li, L. (1985). Cournot Oligopoly with Information Sharing. *The RAND Journal of Economics 16*(4), 521–536.

Myers, C. G. (2015, November). Is Your Company Encouraging Employees to Share What They Know? *Harvard Business Review*. Accessed: 2018-10-14. Archived at https://web.archive.org/web/20181015154019/https://hbr.org/2015/11/is-your-company-encouraging-employees-to-share-what-they-know.

Myerson, R. B. (1982). Optimal Coordination mechanisms in Generalized Principal-Agent Problems. *Journal of Mathematical Economics 10*(1), 67 – 81.

Myerson, R. B. (1986). Multistage Games with Communication. *Econometrica 54*(2), 323–358.

Novshek, W. and H. Sonnenschein (1982). Fulfilled Expectations Cournot Duopoly with Information Acquisition and Release. *The Bell Journal of Economics 13*(1), 214–218.

Raith, M. (1996). A General Model of Information Sharing in Oligopoly. *Journal of Economic Theory 71*(1), 260–288.

Sales, N. A. (2010). Share and Share Alike: Intelligence Agencies and Information Sharing. *Geo. Wash. L. Rev. 78*(2), 279–352.

Shapiro, C. (1986). Exchange of Cost Information in Oligopoly. *The Review of Economic Studies 53*(3), 433–446.

Thomas, K. (2012, October). Glaxo Opens Door to Data on Research . *The New York Times*. Accessed: 2018-09-27. Archived at https://web.archive.org/web/20180928103418/https://www.nytimes.com/2012/10/11/business/glaxo-opens-door-to-data-on-its-research.html.

U.S. Congress (2004). *Intelligence Reform and Terrorism Prevention Act of 2004*. 108th Cong., 2d sess., Public Law 108-458. Washington, D.C.: U.S. G.P.O.

U.S. Congress, Senate, Select Committee on Intelligence. U.S. Congress, House, Permanent Select Committee on Intelligence (2002). *Joint Inquiry into Intelligence Community Activities before and after the Terrorist Attacks of September 11, 2001: Report of the U.S. Senate Select Committee on Intelligence and*

U.S. House Permanent Select Committee on Intelligence together with Additional Views. 107th Cong., 2d sess., S. Rep. 107-351, H. Rep. 107-792. Washington, D.C.: U.S. G.P.O.

Vanderbei, R. (2014). *Linear programming: Foundations and Extensions*. New York: Springer.

Vida, P. and F. Forges (2013). Implementation of Communication Equilibria by Correlated Cheap Talk: The Two-Player Case. *Theoretical Economics 8*(1), 95–123.

Vives, X. (1984). Duopoly Information Equilibrium: Cournot and Bertrand. *Journal of Economic Theory 34*(1), 71–94.

Vives, X. (1990). Trade Association Disclosure Rules, Incentives to Share Information, and Welfare. *The RAND Journal of Economics 21*(3), 409–430.

Vives, X. (2001). *Oligopoly Pricing: Old Ideas and New Tools*. Cambridge, Mass. ; London, England: MIT Press, 2001.

Wang, S. and R. A. Noe (2010). Knowledge Sharing: A Review and Directions for Future Research. *Human Resource Management Review 20*(2), 115 – 131.

Ziv, A. (1993). Information Sharing in Oligopoly: The Truth-Telling Problem. *The RAND Journal of Economics 24*(3), 455–465.

A Proofs for Section 3

Proposition A.1. *The only Bayesian Nash Equilibrium of the game consists of a strategy profile* $0_{:}{}^{1\text{o}}\,{}^{2}_{8=1}$ *with*

$$0_{:}{}^{1}B_{:}{}^{\text{o}} = B_{:}$$

Each player's ex-ante expected equilibrium payoff of the game is

$$c \;\; = A^{1}1 \quad U^{\text{o}}$$

Proof. Note that agent $:$'s best response to each strategy of agent ${}^{1}3 \quad :^{\text{o}}$ requires selecting the most likely correct action of nature given $:$'s signal. Since

$$\mathrm{P} \;\; B \;\; = B_{:}\mathrm{j}B_{:} \;\; = A\text{-}$$

while

$$\mathrm{P} \;\; B \;\; = 1\bullet 2\mathrm{j}B_{:} \;\; = 1 \quad A$$

and then since $A \;\; ¡ \;\; 1\bullet 2\text{-}$ it must be that in equilibrium each agent selects an action that coincides with her signal and thus indeed $0_{:}{}^{1}B_{:}{}^{\text{o}} = B_{:}\bullet$

Given the equilibrium strategies, each agent guesses the the correct action with probability A and thus the ex-ante equilibrium payoff of each player is $c \;\; = A^{1}1 \quad U^{\text{o}}\bullet$

Proposition A.2. *Consider a benevolent third party that observes both signals, cares equally about the agents and solves for the first-best*

$$c^{\text{-}} \;\; = \max_{\text{f}0_{:}{}^{1}B^{\text{o}}\text{g}^{2}_{8=1}} \mathrm{E}_{B} \;\; " \;\; \tilde{\text{O}}_{:} \; D_{:}{}^{1}0_{:}\text{-}0_{3} \;\; {}_{:}\text{-}B^{\text{o}} \;\; \#$$

This problem is solved by $0_{:}{}^{1}B^{\text{o}} = B$ *and accordingly* $c^{\text{-}} \;\; = 1 \quad U\bullet$

Proof. Note that for each choice of $\{\theta_i(\omega_i)\}_{\omega=1}^{2}$

$$\mathbb{E}_{\omega}\left[\sum_i \pi_i(\theta_i(\omega_i), \theta_{-i}(\omega_{-i}), \omega)\right] = \sum_i \sigma_i(\theta_i(\omega_i), \omega) + \sum_i \rho(\theta_i(\omega_i), \omega) \tag{4}$$

where σ_i is the unconditional probability of player i guessing the correct action ω given $\theta_i(\omega_i)$. When $\alpha \in (1/2, 1)$ and $\rho \in (0, 1)$ the expression in (4) is maximized by $\sigma_i = 1$ for each i. The only way to achieve $\sigma_i = 1$ is by setting $\theta_i(\omega_i) = \omega$. The corresponding expected payoff of each agent is $\gamma - \beta = 1 - \rho$.

Proposition A.3. *Consider the game Γ extended with a finite set M_i of messages that agent i can send to agent $-i$ upon observing her private signal. Let Γ^{*} denote the extended game. In every weak Perfect Bayesian Equilibrium of Γ^{*} for each message $m_{-i} \in M_{-i}$ player i chooses the action according to her private signal only: $\theta_i(\omega_i) = \omega_i$.*

Proof. Note first that according to Proposition A.1, each agent chooses the same action as her signal in the absence of any additional information. Expected equilibrium payoffs of the players in the game with communication take the form

$$u_i = \sigma_i + \rho\sigma_{-i}$$
$$u_{-i} = \sigma_{-i} + \rho\sigma_i$$

where σ_i is the probability of player i guessing the correct action. Since each player can guarantee herself a correct guess with probability α based on the private information only, it must be that $\sigma_i + \sigma_{-i} > \alpha$.
Now suppose that there exists a message m_{-i} sent by type ω_{-i} of agent $-i$ in equilibrium, such that some type ω_i of agent i chooses action $1-\omega_i$.
There are two cases: either (i) the type ω_i continues to choose action ω_i upon observing message m_{-i} or (ii) the type ω_i chooses action $1-\omega_i$ upon observing message m_{-i}. In both of these cases player $-i$ can deceive agent i to choose the correct action with probability less than α.

(i) If type ω_i continues to choose action ω_i upon observing message m_{-i} then by sending message m_{-i} irrespective of her own signal, agent $-i$ induces type ω_i of agent i to do action $1-\omega_i$ both in case of coinciding or non-coinciding signals, while type $1-\omega_i$ continues to act on private information only. Thus in case of $-i$ always sending message m_{-i}, the probability of i guessing the correct action is $\sigma_i = 1-\alpha \neq \alpha \geq \sigma_i$. Since for a fixed strategy of i the probability of $-i$ guessing the correct action σ_{-i} remains constant, agent $-i$ has a profitable deviation.

(ii) If type ω_i chooses action $1-\omega_i$ upon observing message m_{-i} then by sending message m_{-i} irrespective of her signal, agent $-i$ induces all types of agent i to do action $1-\omega_i$ and the probability of i guessing the correct action in case of such a deviation is $\sigma_i = 1 - \alpha \neq \alpha \geq \sigma_i$. Again, since for a fixed strategy of i the probability of $-i$ guessing the correct action σ_{-i} remains constant, agent $-i$ has a profitable deviation.

Thus if there exists a message m_{-i} that induces some type of agent i to choose action $1-\omega_i$, agent $-i$ necessarily has a profitable deviation. Therefore, it must be that in every every weak Perfect Bayesian Equilibrium of the game Γ^{*} with direct communication, each agent acts on her private information only.

B Optimal mediation for the illustrative example

This appendix section finds the optimal mediation protocol for the illustrative example of Section 3. Recall that the baseline case of the example consists of the game Γ. In the game, each agent $i \in \{1, 2\}$ obtains

a binary signals $B_i \in S = \{0,1\}$ with the following joint distribution P over $S^2 = S \times S$ parametrized by $A \in [1/2, 1]$:

	$s_2 = 0$	$s_2 = 1$
$s_1 = 0$	$\frac{A}{2}$	$\frac{1-A}{2}$
$s_1 = 1$	$\frac{1-A}{2}$	$\frac{A}{2}$

Together, these signals determine the correct action

$$B = \frac{1}{2} \sum_i B_i \bmod 2$$

Both agents would like to guess B by choosing an action in the set $A_i = \{0,1\}$. The agents have a conflict of interest and prefer the opponent not to be able to guess the correct action. The payoffs representing such preferences are given by

$$u_i(a_i, a_{-i}, B) = \mathbb{1}\{a_i = B\} - U \cdot \mathbb{1}\{a_{-i} = B\} \tag{5}$$

where $U \in [0,1]$.

Consider now a mediated game. That is, a mediation protocol is introduced that receives reports from the agents and sends messages back to the agents. For each possible report profile received from the agents the protocol specifies a distribution on the messages sent back to the agents.

Due to the revelation principle (see Myerson (1982, 1986) and Forges (1986)), attention can be restricted to *direct revelation* mediation protocols that take the type-reports from the agents and send back action recommendations to each player. A direct revelation mediation protocol π is defined as a function from the product type space into the joint distributions over the action recommendations $\pi : S \to \Delta(A)$. That is, each agent is asked to submit her signal received and conditional on the pair of reports is advised on an action, possibly in a random way. The mediation protocol should be such that the agents find it optimal to report their true types and follow the recommended action conditional on the other player reporting truthfully and following recommendations[11].

Let M be the set of all *incentive-compatible* mediation protocols. Let c_i^{π} be the ex-ante expected equilibrium payoff of agent i in game π with two agents, signals jointly distributed according to P in (B), the correct action B determined as in (B), payoff-relevant actions be $A_i = \{0,1\}$ and the mediation protocol $\pi \in M$. Consider now the problem of optimal mediation protocol design faced by competitors

$$\max_{\pi \in M} \left[\sum_i c_i^{\pi} \right] \tag{6}$$

The following sequence of lemmas first simplifies the problem in (6) and leads to Theorem B.1 which presents the optimal mediation protocol.

Lemma B.1 below shows that, when solving (6), without loss of generality one can consider only mediation protocols that generate independent recommendations conditional on the pair of reports.

Lemma B.1. *For every* $\pi \in M$ *there exists* $\pi' \in M$ *such that*

(i) $\pi' : S \to \Delta(A_1) \times \Delta(A_2)$

[11] The solution concept corresponds to *information design with elicitation* in terminology of Bergemann and Morris (2018).

(ii) $\pi^{1}(B)$ *and* $\pi^{01}(B)$ *have the same marginal distributions for every* B

(iii) $c_i^{\pi^0} = c_i^{\pi}$ *for every* i

Proof. Consider an incentive-compatible mediation protocol $\pi \in M$. Define π^0 to be a mapping from S to $\Delta(A \times A)$ such that for each $B \in S$ - for each vector of recommendations a

$$\pi^{01}(B)(a) = \prod_{i \in I} \sum_{a_{-i} \in A^{I \setminus \{i\}}} \pi^{1}(B)(a_i, a_{-i})$$

That is, π^0 is defined to be the product of the marginal distributions of π for each pair of reports B. By construction, (i) $\pi^0 \in \Delta(A \times A)$ and (ii) π and π^0 have the same marginal distributions.
Now, since $\pi \in M$ - it must also be that $\pi^0 \in M$. To see this observe first that for type t_i of agent i that submitted report t_i and received a recommendation a_i - agent i's posterior over the signal of agent $\{-i\}$ [and thus also i's preferred action] is pinned down by the marginal of recommendation distributions. Thus for each report t_i of type t_i of agent i under π^0 - the resulting distribution of i's actions is the same as under π and consequently the probability of type t_i agent i making a correct guess is the same under π and π^0. Similarly, for each report t_i of agent i under π^0 - the resulting distribution of $\{-i\}$'s actions is the same as under π. Since the distribution of both agents' actions pin down the expected payoff of agent i - type t_i of agent i has the same expected payoff for each report t_i under π and π^0. By assumption there were no profitable deviations from truth-telling under π - thus so is the case under π^0. It is therefore proved that $\pi^0 \in M$ and (iii) $c_i^{\pi^0} = c_i^{\pi}$.

Exploiting Lemma B.1, one can restrict attention to mediation protocols with independent action recommendations when solving for the optimal communication protocol. Thus from now on M is redefined to be the set of IC mediation protocols with independent action recommendations.

Next, Lemma B.2 shows that none of the IC mediation protocols recommend action $1 - t_i$ to agent i who reported type t_i. That is, an action that is known to be incorrect by a particular agent is never recommended to this agent.

LEMMA B.2. *For every* $\pi \in M$ *it must be that* $\pi^{1}(B) \in \Delta(A \times A \setminus \{(1-t_i, t_i)\})$.

Proof. Suppose that for some $\pi \in M$ action $1 - t_i$ is recommended to agent with a truthful report t_i. Note that following such a recommendation yields a probability 0 of agent i guessing B. Obviously, agent i can deviate from following the recommendation, act on her private information only and guarantee herself at least a probability A of guessing B.

Every $\pi \in M$ is now completely summarized by a vector in $[0,1]^8$ with typical values $\alpha_{s_i} \in [0,1]$ presented in the following table:

π

	$\hat{s}_2 = 0$	$\hat{s}_2 = 1$
$\hat{s}_1 = 0$	$\alpha^1_{00}, \alpha^2_{00}$	$\alpha^1_{01}, \alpha^2_{10}$
$\hat{s}_1 = 1$	$\alpha^1_{10}, \alpha^2_{01}$	$\alpha^1_{11}, \alpha^2_{11}$

In the table $<^{\cdot}_{\theta\text{-}\vartheta}$ is the probability of recommending the correct action to agent $\cdot$- when agent $\cdot$ reported θ and agent $^{1}3\ \ \cdot^{o}$ reported ϑ [12]. The vector $<$ consisting of $<^{\cdot}_{\theta\text{-}\vartheta}\ 2\ \text{»}0\text{-}1¼$ is *feasible* if the corresponding $''$ belongs to $M\bullet$

Lemma B.3 below simplifies the problem in (6) even further, stating that there is no loss in optimizing the weighted sum of payoffs with respect to just two variables: the probability of recommending a correct action in case of (i) coinciding and (ii) non-coinciding reports.

To prove Lemma B.3 the following claim is first established.

Claim: A mediation protocol $''$ defined by a vector $<^{\cdot}_{\theta\text{-}\vartheta\ \theta\text{-}\vartheta=0\text{-}0\ \cdot=1}{}^{n\quad o_{1\text{-}1}\quad 2}$ belongs to the set of IC mediation protocols M if and only if the following two sets of condition hold. First, for each $\cdot$ and θ

$$<^{\cdot}_{\theta\theta}A > {}^{1}1\quad <^{\cdot}_{\theta\text{-}1\ \theta}{}^{o}{}^{1}1\quad A^{o} \tag{7}$$

$$<^{\cdot}_{\theta\text{-}1\ \theta}{}^{1}1\quad A^{o} > {}^{1}1\quad <^{\cdot}_{\theta\text{-}\theta}{}^{o}A \tag{8}$$

Second, for each $\cdot$ and θ

$$\begin{array}{l} {}^{1}<^{\cdot}_{\theta\theta}\quad U<^{3}_{\theta\theta}{}^{\cdot o}A\ \text{¸}\ {}^{1}<^{\cdot}_{\theta\text{-}1\ \theta}\quad U<^{3}_{1\ \theta\text{-}\theta}{}^{o}{}^{1}1\quad A^{o} > \\ \quad A^{1}1\quad U^{1}1\quad <^{3}_{\theta\text{-}1}{}^{\cdot}_{\theta}{}^{oo}\ \text{¸}\ {}^{1}1\quad A^{o}{}^{1}0\quad U^{1}1\quad <^{3}_{1}{}^{\cdot}_{\theta\text{-}1\ \theta}{}^{oo} \qquad \text{if } {}^{1}1\quad <^{\cdot}_{1\ \theta\text{-}\theta}{}^{o}A\ i\ <^{\cdot}_{1\ \theta\text{-}1\ \theta}{}^{1}1\quad A^{o} \\ \quad A^{1}<^{\cdot}_{1\ \theta\text{-}\theta}\quad U^{1}1\quad <^{3}_{\theta\text{-}1}{}^{\cdot}_{\theta}{}^{oo}\ \text{¸}\ {}^{1}1\quad A^{o}{}^{1}<^{\cdot}_{1\ \theta\text{-}1\ \theta}\quad U^{1}1\quad <^{3}_{1\ \theta\text{-}1}{}^{\cdot}_{\theta}{}^{oo} \qquad \text{if } {}^{1}1\quad <^{\cdot}_{1\ \theta\text{-}\theta}{}^{o}A\ 6\ <^{\cdot}_{1\ \theta\text{-}1\ \theta}{}^{1}1\quad A^{o} \end{array} \tag{9}$$

Conditions (7)-(8) ensure that each agent finds it profitable to follow the mediation protocol's recommendation. Conditions (9) ensure that there are no profitable deviations from reporting truthfully to the mediation protocol.

Proof: The two sets of IC conditions (7)-(8) and (9) are established separately.

"Following recommendation" conditions Under an IC direct revelation mediation protocol, agent $\cdot$ finds it profitable to follow the mediation protocol's recommendation. In the problem at hand, one needs to establish conditions under which this is the case for every signal and recommendation obtained by agent $\cdot$.

- Suppose agent $\cdot$ has signal θ and obtained a recommendation $'$ to do θ. Such a recommendation can occur when the reported pair of types is either $^{1}\theta\text{-}\theta^{o}$ or $^{1}\theta\text{-}1\quad \theta^{o}\bullet$ The posterior probability that the state is $^{1}\theta\text{-}\theta^{o}$ (and thus that the correct action is indeed θ) is equal to

$$\begin{aligned} P\ \ B = {}^{1}\theta\text{-}\theta^{o}\mathrm{j}\ ' = \theta\text{-}B_{\cdot} = \theta\ \ &= P\ \ B_{3\ \cdot} = \theta\mathrm{j}\ ' = \theta\text{-}B_{\cdot} = \theta \\ &= \frac{P\ \ B_{\cdot} = \theta\text{-}B_{3\ \cdot} = \theta\text{-}\ ' = \theta}{P\ \ B_{\cdot} = \theta\text{-}\ ' = \theta} \\ &= \frac{<^{\cdot}_{\theta\theta}\ \frac{A}{2}}{<^{\cdot}_{\theta\theta}\ \frac{A}{2}\ \text{¸}\ {}^{1}1\quad <^{\cdot}_{\theta\text{-}1\ \theta}{}^{o}\ \frac{1\ A}{2}} \end{aligned}$$

  Agent $\cdot$ following recommendation means that $B = {}^{1}\theta\text{-}\theta^{o}$ is more likely than $B = {}^{1}\theta\text{-}1\quad \theta^{o}$, which yields the first IC constraint

$$<^{\cdot}_{\theta\theta}A > {}^{1}1\quad <^{\cdot}_{\theta\text{-}1\ \theta}{}^{o}{}^{1}1\quad A^{o}$$

  and condition (7) is established.

[12] The correct action recommendation to agent $\cdot$ in this case is $^{1}\theta\ \text{¸}\ \vartheta^{o}\bullet 2$ in case of truthful reporting, while the incorrect action recommendation to agent $\cdot$ is $^{1\,1}\theta\ \text{¸}\ {}^{1}1\quad \vartheta^{oo}\bullet 2\bullet$

- Suppose now agent i has signal s and obtained a recommendation r to do 1•2• Such a recommendation can occur when the reported pair of types is either (s, s^0) or $(s, 1 - s^0)$• The posterior probability that the state is $(s, 1 - s^0)$ (and thus the correct action is indeed 1•2) is equal to

$$\begin{aligned} P(\theta = (s, 1-s^0) \mid r = 1\bullet 2\text{-}\theta_i = s) &= P(\theta_{-i} = 1-s \mid r = 1\bullet 2\text{-}\theta_i = s) \\ &= \frac{P(\theta_i = s\text{-}\theta_{-i} = 1-s \text{-} r = 1\bullet 2)}{P(\theta_i = s\text{-} r = 1\bullet 2)} \\ &= \frac{\pi(s, 1-s) \frac{1}{2}}{\pi(s, s) \frac{1}{2} + \pi(1-s, 1-s^0) \frac{1}{2}} \end{aligned}$$

  Agent 1 following recommendation means that $(s, 1 - s^0)$ is more likely than (s, s^0), which yields the second IC constraint

$$\pi(s, 1-s) > \pi(s, s^0)$$

  and condition (8) is established.

The "following recommendation" conditions are thus established.

"Truthful reporting" conditions Under an IC direct revelation mediation protocol, agent i finds it profitable to report truthfully to the mediation protocol and follow the recommendation rather than misreporting and doing some other action upon receiving a recommendation. In the problem at hand, one needs to establish conditions under which this is the case for every signal obtained by agent i.

- Suppose agent i received a signal s. If she reports [T]ruthfully and follows the recommendation by the mediation protocol (and so does agent $3 - i$), the expected payoff is

$$\begin{aligned} v_i^T &= E(v_i \mid \theta_i = s) = E(v_i \mid \theta_i = s\text{-}\theta_{-i} = s) P(\theta_{-i} = s \mid \theta_i = s) \\ &\quad + E(v_i \mid \theta_i = s\text{-}\theta_{-i} = 1-s) P(\theta_{-i} = 1-s \mid \theta_i = s) \\ &= \pi(s, s) u(a^0, s) + \pi(s, 1-s) u(a^1, 1-s) \end{aligned}$$

- If agent i misreports and sends $1 - s$ to the mediation protocol instead of s, it is possible to hear two recommendations in response: $1 - s$ or 1•2• The optimal actions in each of these cases are established below:

  – What is the optimal action if $1 - s$ is recommended back by the mediation protocol? The conditional probability of agent $3 - i$ having a signal s is equal to

$$\begin{aligned} P^*(\theta_{-i} = s \mid r = 1-s\text{-}\theta_i = s) &= \frac{P^*(\theta_i = s\text{-}\theta_{-i} = s\text{-} r = 1-s)}{P^*(r = 1-s\text{-}\theta_i = s)} \\ &= \frac{\pi(1-s, s) \frac{1}{2}}{\pi(1-s, s) \frac{1}{2} + \pi(s, 1-s) \frac{1}{2}} \end{aligned}$$

  where P^* stands for the updated probabilities given an [U]ntruthful report.

  Similarly, the conditional probability of agent $3 - i$ having a signal $1 - s$ is equal to

$$P^*(\theta_{-i} = 1-s \mid r = 1-s\text{-}\theta_i = s) = \frac{\pi(s, 1-s) \frac{1}{2}}{\pi(1-s, s) \frac{1}{2} + \pi(s, 1-s) \frac{1}{2}}$$

  Consequently, if $\pi(1-s, s^0) > \pi(s, 1-s)$ agent i will do action s upon receiving signal $1 - s$ and if $\pi(1-s, s^0) < \pi(s, 1-s)$ agent i will do action 1•2 upon receiving signal $1 - s$•

– What is the optimal action if 1•2 is recommended back by the mediation protocol? The conditional probability of agent ¹3 :º having a signal 8 is equal to

$$P_* \; B_3 \; _: = 8 \mathrm{j} \; ' = 1•2- B_: = 8 \; = \frac{P_* \; B_3 \; _: = 8- B_: = 8- \; ' = 1•2}{P_* \; ' = 1•2- B_: = 8}$$

$$= \frac{<^{:}_{1\; 8-8} \; \frac{A}{2}}{<^{:}_{1\; 8-8} \; \frac{A}{2} \; ¸ \; ¹1 \; <^{1}_{1\; 8-1\; 8}º \; \frac{1\; A}{2}}-$$

while the conditional probability of agent ¹3 :º having a signal 1 8 is equal to

$$P_* \; B_3 \; _: = 1 \; 8 \mathrm{j} \; ' = 1•2- B_: = 8 \; = \frac{¹1 \; <^{:}_{1\; 8-1\; 8}º \; \frac{1\; A}{2}}{<^{:}_{1\; 8-8} \; \frac{A}{2} \; ¸ \; ¹1 \; <^{:}_{1\; 8-1\; 8}º \; \frac{1\; A}{2}}-$$

Note that the IC constraint (8) and Assumption 1 imply that $<^{:}_{1\; 8-8}A$ ¡ ¹1 $<^{:}_{1\; 8-1\; 8}$º¹1 Aº and thus agent : prefers to do action 8 upon hearing a recommendation of 1•2 from the mediation protocol.

- Now [after trivially calculating the probabilities of agent ¹3 :º making the correct guess], the expected payoff of agent : in case of misreporting and sending 1 8 instead of 8 can be computed. If ¹1 $<^{:}_{1\; 8-8}$ºA ¡ $<^{:}_{1\; 8-1\; 8}$¹1 Aº- agent : does 8 in any case and gets an expected utility of

$$c^{*-¡}_{:} = A¹1 \; U¹1 \; <^{3}_{8-1}{}^{:}_{8}ºº \; ¸ \; ¹1 \; Aº¹0 \; U¹1 \; <^{3}_{1}{}^{:}_{8-1\; 8}ºº$$

If ¹1 $<^{:}_{1\; 8-8}$ºA 6 $<^{:}_{1\; 8-1\; 8}$¹1 Aº- agent : does 1/2 in case of hearing a recommendation of 1 8 and 8 in case of recommendation of 1•2 and gets an expected utility of

$$\begin{aligned} c^{*-6}_{:} &= A¹<^{:}_{1\; 8-8} \; 1 \; ¸ \; ¹1 \; <^{:}_{1\; 8-8}º \; 0 \; U¹1 \; <^{:}_{8-1\; 8}ºº \\ &¸ \; ¹1 \; Aº¹<^{:}_{1\; 8-1\; 8} \; 1 \; ¸ \; ¹1 \; <^{:}_{1\; 8-1\; 8}º \; 0 \; U¹1 \; <^{3}_{1}{}^{:}_{8-1\; 8}ºº \\ &= A¹<^{:}_{1\; 8-8} \; U¹1 \; <^{3}_{8-1}{}^{:}_{8}ºº \; ¸ \; ¹1 \; Aº¹<^{:}_{1\; 8-1\; 8} \; U¹1 \; <^{3}_{1}{}^{:}_{8-1\; 8}ºº \end{aligned}$$

Thus the IC constraint for agent : reporting truthfully upon observing signal 8 is

$$\begin{aligned} &¹<^{:}_{88} \; U<^{3}_{88}{}^{:}ºA \; ¸ \; ¹<^{:}_{8-1\; 8} \; U<^{3}_{1}{}^{:}_{8-8}º¹1 \; Aº > \\ &\quad A¹1 \; U¹1 \; <^{3}_{8-1}{}^{:}_{8}ºº \; ¸ \; ¹1 \; Aº¹0 \; U¹1 \; <^{3}_{1}{}^{:}_{8-1\; 8}ºº \qquad \text{if } ¹1 \; <^{:}_{1\; 8-8}ºA \; ¡ \; <^{:}_{1\; 8-1\; 8}¹1 \; Aº \\ &\quad A¹<^{:}_{1\; 8-8} \; U¹1 \; <^{3}_{8-1}{}^{:}_{8}ºº \; ¸ \; ¹1 \; Aº¹<^{:}_{1\; 8-1\; 8} \; U¹1 \; <^{3}_{1}{}^{:}_{8-1\; 8}ºº \qquad \text{if } ¹1 \; <^{:}_{1\; 8-8}ºA \; 6 \; <^{:}_{1\; 8-1\; 8}¹1 \; Aº \end{aligned}$$

and "truthful reporting" conditions (9) are established.
The proof of the claim is now completed. ž

LEMMA B.3. *There exists a solution to the optimal mediation protocol design problem in eq.* (6) *such that* 8:-8 $<^{:}_{8-8}$ = ? 2 »0- 1¼ *and* $<^{:}_{8-1\; 8}$ = @ 2 »0- 1¼•

Proof: Note first that the optimal mediation protocol design problem

$$\max_{" \, 2 \mathrm{M}} \; " \; \tilde{O}_{8} \; c_8^{"} \; \# \tag{10}$$

can be written more explicitly as

$$\max_{<^{:}_{8-1\; 8}} \; " \; \frac{A}{2} \; \tilde{O}_{8} \; \tilde{O}_{:} \; <^{:}_{88} \; U<^{3}_{88}{}^{:} \; ¸ \; \frac{1\; A}{2} \; \tilde{O}_{8} \; \tilde{O}_{:} \; <^{:}_{8-1\; 8} \; U<^{3}_{8-1}{}^{:}_{8} \; \# \tag{11}$$

s. t. (7)-(9)

To prove the lemma, one needs to notice that (i) the constraint set defined by IC conditions is convex; (ii) the objective and the constraint set are symmetric with respect to players and states.

Convexity of the constraint set (7)-(9) To establish (i) one can show first that each IC condition defines a convex set. This is obvious for conditions (7)–(8) as these a linear in parameters. Now consider the typical truthful-reporting IC constraint

$$¹<^{;}_{88} \quad U<^{3}_{88}{}^{;}ºA \, ¸ \, ¹<^{;}_{8-1\ 8} \quad U<^{3}_{1\ 8-8}{}^{;}º¹1 \quad Aº >$$
$$A¹1 \quad U¹1 \quad <^{3}_{8-1\ 8}{}^{;}ºº \, ¸ \, ¹1 \quad Aº¹0 \quad U¹1 \quad <^{3}_{1\ 8-1\ 8}{}^{;}ºº \qquad \text{if } ¹1 \quad <^{;}_{1\ 8-8}ºA \ ¡ \ <^{;}_{1\ 8-1\ 8}¹1 \quad Aº$$
$$A¹<^{;}_{1\ 8-8} \quad U¹1 \quad <^{3}_{8-1\ 8}{}^{;}ºº \, ¸ \, ¹1 \quad Aº¹<^{;}_{1\ 8-1\ 8} \quad U¹1 \quad <^{3}_{1\ 8-1\ 8}{}^{;}ºº \qquad \text{if } ¹1 \quad <^{;}_{1\ 8-8}ºA \ 6 \ <^{;}_{1\ 8-1\ 8}¹1 \quad Aº \tag{12}$$

and consider two vectors C and B that satisfy those constraints.

If both C and B are such that $¹1 \quad C^{;}_{1\ 8-8}ºA \ ¡ \ ;^{;}_{1\ 8-1\ 8}¹1 \quad Aº$ and $¹1 \quad B^{;}_{1\ 8-8}ºA \ ¡ \ B^{;}_{1\ 8-1\ 8}¹1 \quad Aº$ or $¹1 \quad C^{;}_{1\ 8-8}ºA \ 6 \ ;^{;}_{1\ 8-1\ 8}¹1 \quad Aº$ and $¹1 \quad B^{;}_{1\ 8-8}ºA \ 6 \ B^{;}_{1\ 8-1\ 8}¹1 \quad Aº$ - then the convex combination $D = VC \, ¸ \, ¹1 \quad VºB- V \ 2 \ »0- 1¼$ also satisfies the constraint (12) with $¹1 \quad D^{;}_{1\ 8-8}ºA \ ¡ \ D^{;}_{1\ 8-1\ 8}¹1 \quad Aº$ or $¹1 \quad D^{;}_{1\ 8-8}ºA \ 6 \ D^{;}_{1\ 8-1\ 8}¹1 \quad Aº$ respectively by linearity of both RHS and LHS of the inequality in (12).

Now suppose that

$$¹1 \quad C^{;}_{1\ 8-8}ºA \ ¡ \ ;^{;}_{1\ 8-1\ 8}¹1 \quad Aº$$
$$¹C^{;}_{88} \quad UC^{3}_{88}{}^{;}ºA \, ¸ \, ¹C^{;}_{8-1\ 8} \quad UC^{3}_{1\ 8-8}{}^{;}º¹1 \quad Aº > A¹1 \quad U¹1 \quad C^{3}_{8-1\ 8}{}^{;}ºº \, ¸ \, ¹1 \quad Aº¹0 \quad U¹1 \quad C^{3}_{1\ 8-1\ 8}{}^{;}ºº$$
$$¹1 \quad B^{;}_{1\ 8-8}ºA \ 6 \ B^{;}_{1\ 8-1\ 8}¹1 \quad Aº$$
$$¹B^{;}_{88} \quad UB^{3}_{88}{}^{;}ºA \, ¸ \, ¹B^{;}_{8-1\ 8} \quad UB^{3}_{1\ 8-8}{}^{;}º¹1 \quad Aº > A¹B^{;}_{1\ 8-8} \quad U¹1 \quad B^{3}_{8-1\ 8}{}^{;}ºº \, ¸ \, ¹1 \quad Aº¹B^{;}_{1\ 8-1\ 8} \quad U¹1 \quad B^{3}_{1\ 8-1\ 8}{}^{;}ºº$$

and let $D = VC \, ¸ \, ¹1 \quad VºB$ with $V \ 2 \ »0- 1¼•$

First, either $¹1 \quad D^{;}_{1\ 8-8}ºA \ ¡ \ D^{;}_{1\ 8-1\ 8}¹1 \quad Aº$ or $¹1 \quad D^{;}_{1\ 8-8}ºA \ 6 \ D^{;}_{1\ 8-1\ 8}¹1 \quad Aº•$ Suppose $¹1 \quad D^{;}_{1\ 8-8}ºA \ ¡ \ D^{;}_{1\ 8-1\ 8}¹1 \quad Aº-$ then to show convexity of the set defined by constraint (12), one needs to establish that

$$¹B^{;}_{88} \quad UB^{3}_{88}{}^{;}ºA \, ¸ \, ¹B^{;}_{8-1\ 8} \quad UB^{3}_{1\ 8-8}{}^{;}º¹1 \quad Aº > A¹1 \quad U¹1 \quad B^{3}_{8-1\ 8}{}^{;}ºº \, ¸ \, ¹1 \quad Aº¹0 \quad U¹1 \quad B^{3}_{1\ 8-1\ 8}{}^{;}ºº \tag{13}$$

Indeed, if this is the case, then also

$$¹D^{;}_{88} \quad UD^{3}_{88}{}^{;}ºA \, ¸ \, ¹D^{;}_{8-1\ 8} \quad UD^{3}_{1\ 8-8}{}^{;}º¹1 \quad Aº > A¹1 \quad U¹1 \quad D^{3}_{8-1\ 8}{}^{;}ºº \, ¸ \, ¹1 \quad Aº¹0 \quad U¹1 \quad D^{3}_{1\ 8-1\ 8}{}^{;}ºº$$

since D is a convex combination of C and $B•$

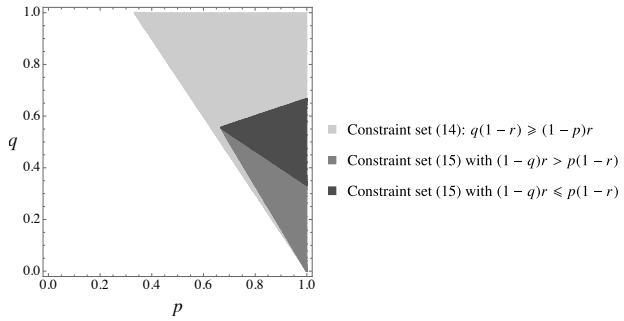
To show (13) note that

$$¹B^{;}_{88} \quad UB^{3}_{88}{}^{;}ºA \, ¸ \, ¹B^{;}_{8-1\ 8} \quad UB^{3}_{1\ 8-8}{}^{;}º¹1 \quad Aº > A¹B^{;}_{1\ 8-8} \quad U¹1 \quad B^{3}_{8-1\ 8}{}^{;}ºº \, ¸ \, ¹1 \quad Aº¹B^{;}_{1\ 8-1\ 8} \quad U¹1 \quad B^{3}_{1\ 8-1\ 8}{}^{;}ºº$$
$$> A¹1 \quad U¹1 \quad B^{3}_{8-1\ 8}{}^{;}ºº \, ¸ \, ¹1 \quad Aº¹0 \quad U¹1 \quad B^{3}_{1\ 8-1\ 8}{}^{;}ºº-$$

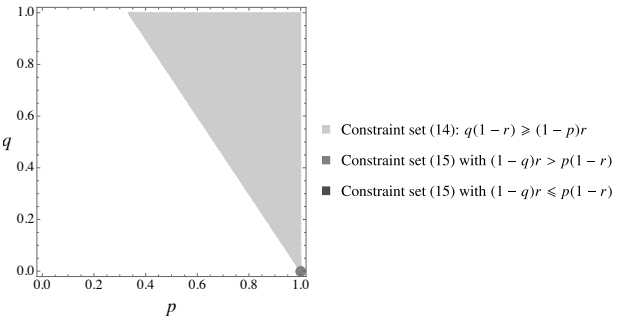
where the last inequality follows from $¹1 \quad B^{;}_{1\ 8-8}ºA \ 6 \ B^{;}_{1\ 8-1\ 8}¹1 \quad Aº•$ The case of $¹1 \quad D^{;}_{1\ 8-8}ºA \ 6 \ D^{;}_{1\ 8-1\ 8}¹1 \quad Aº$ is similar.

Since the intersection of convex sets is convex, the constraint set is convex itself.

Symmetry of the objective (11) *and the constraint set* (7)-(9) Note that if the maximization problem is solved by some vector $< \, = \, ¹<^{1}- <^{2}º$ (where $<^{8}$ denotes the subvector of probabilities related to agent 8), then vector $<^{0} = ¹<^{2}- <^{1}º$ leads to the same value of the objective function and constraints are satisfied at $<^{0}$ by symmetry. Thus the same value of the objective is achieved at the average of $<- <^{0}$ and one can restrict attention to maximizing with 4-element vector $<_{00}- <_{01}- <_{10}- <_{11}•$ Again, swapping $<_{00}$ with $<_{11}$



(a) A typical constraint set with U 6 1 A•



(b) A typical constraint set with U ¡ 1 A•

Figure 1: Typical constraint sets.

and $<_{01}$ with $<_{10}$ leads to the same value of the modified objective function and constraints being satisfied by symmetry. Thus one can restrict attention to maximizing with respect to 2-element vector ¹?-@º with $? = <_{88}$ and $@ = <_{8\text{-}1\ 8}$ and the proof of the lemma is now completed. J

Utilizing the results from the preceding lemmas, Theorem B.1 provides an explicit solution to the problem of designing the optimal mediation protocol.

Theorem B.1. *The optimal mediation protocol design problem in eq.* (6) *is solved by* " 2 M *such that* 8:-8 ¹< º$_{8-8}$ = ? *and* ¹< º$_{8-1\ 8}$ = @ *with*

$$
\begin{aligned}
? &= 1\text{-} \quad @ = \frac{2A \quad 1}{2A \quad 1 \ , \ U} \text{-} \quad \text{if } U \ 6 \ 1 \quad A \\
¹? &= 1\text{-} \quad @ = 0º\text{-} \quad \text{if } U \ ¡ \ 1 \quad A
\end{aligned}
$$

Proof. Due to Lemma B.3 the optimal mediation protocol design problem is reduced to

$$\max_{?\text{-}@2»0\text{-}1¼2} »?A \ , \ @¹1 \quad Aº¼$$

subject to

$$?A > ¹1 \quad @º¹1 \quad Aº \tag{14}$$

$$@¹1 \quad Aº > ¹1 \quad ?ºA \tag{15}$$

$$¹? \quad U?ºA \ , \ ¹@ \quad U@º¹1 \quad Aº > \begin{array}{llllllll} A¹1 & U¹1 & @ºº \ , \ ¹1 & Aº¹0 & U¹1 & ?ºº & \text{if } ¹1 \quad @ºA \ ¡ \ ?¹1 & Aº \\ A¹@ & U¹1 & @ºº \ , \ ¹1 & Aº¹? & U¹1 & ?ºº & \text{if } ¹1 \quad @ºA \ 6 \ ?¹1 & Aº \end{array} \tag{16}$$

For a moment, ignore the first two constraints (14)-(15). It will be verified later that the solution of the relaxed maximization problem still satisfies these two constraints.

- Note that the value of the objective grows with $?$ along the line ¹1 @ºA = ?¹1 Aº under Assumption 1. Indeed, substituting

  $$@ = 1 \quad ?\frac{1 \quad A}{A}$$

  into the objective yields coefficient equal to 2 1•A on the variable $?$ and thus the objective grows in ?• Having observed this, it is easy to see that the value of the objective in the region with ¹1 @ºA ¡ ?¹1 Aº is not higher than at the point point on its boundary with the highest value of p, which is 1-$\frac{2A \quad 1}{A}$ •

- Now also note that the objective grows with $?$ along the line

  $$¹? \quad U?ºA \ , \ ¹@ \quad U@º¹1 \quad Aº = A¹@ \quad U¹1 \quad @ºº \ , \ ¹1 \quad Aº¹? \quad U¹1 \quad ?ºº$$

Indeed, substituting

$$@ = ?\frac{1 \quad 2A \; , \; U}{1 \quad 2A \quad U} \qquad \frac{U}{1 \quad 2A \quad U}$$

into the objective yields coefficient $^1U \; , \; 1^o{}^1 2A \quad 1^o \bullet {}^1U \; , \; 2A \quad 1^o \; ¡ \; 0$ on $?\bullet$ Thus if the point $1\text{-}\frac{2A \quad 1}{2A \; , \; U \quad 1}$ satisfies $^1 1 \quad @^o A \; 6 \; ?^1 1 \quad A^o$, it has the highest value of the objective in the region with $^1 1 \quad @^o A \; 6 \; ?^1 1 \quad A^o \bullet$

- Note that $U \; 6 \; 1 \quad A$ simultaneously guarantees that $1\text{-}\frac{2A \quad 1}{2A \; , \; U \quad 1}$ satisfies $^1 1 \quad @^o A \; 6 \; ?^1 1 \quad A^o$ and has a higher value of the objective than $1\text{-}\frac{2A \quad 1}{A}$ • Moreover, the point $1\text{-}\frac{2A \quad 1}{2A \; , \; U \quad 1}$ satisfies the two omitted constraints (14)-(15) of the maximization problem. Thus the point that maximizes the objective for $U \; 6 \; 1 \quad A$ is $1\text{-}\frac{2A \quad 1}{2A \; , \; U \quad 1}$ •
- If in turn $U \; ¡ \; 1 \quad A$- then there are no points that satisfy the constraint with $^1 1 \quad @^o A \; Ÿ \; ?^1 1 \quad A^o$: the set defined by $^1 ? \quad U ?^o A \; , \; ^1 @ \quad U @^o{}^1 1 \quad A^o > A^1 @ \quad U^1 1 \quad @^{oo} \; , \; ^1 1 \quad A^o{}^1 ? \quad U^1 1 \quad ?^{oo}$ has no intersection with the set defined by $^1 1 \quad @^o A \; Ÿ \; ?^1 1 \quad A^o$- which is easy to verify by comparing the values of the linear constraints at the boundary points of the constraint set $? = 0$ and $? = 1$. Moreover, the only point satisfying the constraint with $^1 1 \quad @^o A > ?^1 1 \quad A^o$ is $^1 1\text{-}0^o$- which remains the only candidate for the optimal mediation protocol when $U \; ¡ \; 1 \quad A\bullet$ This point obviously satisfies the omitted constraints (14)-(15).
- For a graphical treatment, two typical constraint sets are shown in Appendix B.

The search for the optimal mediation protocol is thus completed

$$\begin{array}{lll} ? = 1\text{-} & @ = \dfrac{2A \quad 1}{2A \quad 1 \; , \; U} \text{-} & \text{if } U \; 6 \; 1 \quad A \\ {}^1 ? = 1\text{-} & @ = 0^o\text{-} & \text{if } U \; ¡ \; 1 \quad A \end{array}$$

and Theorem B.1 is proved.

Remark B.1. *" is the unique solution of the optimal mediation protocol design problem in the set* M *of protocols with independent action recommendations.*

Proof. Table 1 below explicitly presents " together with its gradient and the gradients of the 8 constraints that bind at the maximum found in Theorem B.1. The binding constraints are (i) $<^{\cdot}_{88} \; 6 \; 1 \; \overset{1}{\underset{8=0}{\overset{\eta}{}}} \; \overset{0_2}{\underset{:=1}{}}$; (ii) the "truthful reporting" constraints in (9).

$<^{\cdot}_{89}$	"	r of obj.	r of constraints							
			$<^{\cdot}_{88} \; 6 \; 1$				"truthful reporting"			
$<^1_{00}$	?	$A^1 1 \quad U^o$	1	0	0	0	A	UA	$1 \quad A$	$U^1 1 \quad A^o$
$<^1_{01}$	@	$^1 1 \quad A^o{}^1 1 \quad U^o$	0	0	0	0	$^1 1 \quad A^o$	UA	A	$U^1 1 \quad A^o$
$<^1_{10}$	@	$^1 1 \quad A^o{}^1 1 \quad U^o$	0	0	0	0	A	$U^1 1 \quad A^o$	$^1 1 \quad A^o$	UA
$<^1_{11}$	?	$A^1 1 \quad U^o$	0	1	0	0	$1 \quad A$	$U^1 1 \quad A^o$	A	UA
$<^2_{00}$	?	$A^1 1 \quad U^o$	0	0	1	0	UA	A	$U^1 1 \quad A^o$	$1 \quad A$
$<^2_{01}$	@	$^1 1 \quad A^o{}^1 1 \quad U^o$	0	0	0	0	UA	$^1 1 \quad A^o$	$U^1 1 \quad A^o$	A
$<^2_{10}$	@	$^1 1 \quad A^o{}^1 1 \quad U^o$	0	0	0	0	$U^1 1 \quad A^o$	A	UA	$^1 1 \quad A^o$
$<^2_{11}$	?	$A^1 1 \quad U^o$	0	0	0	1	$U^1 1 \quad A^o$	$1 \quad A$	UA	A

Table 1: Value of objective, gradients of objective and constraints.

Now note that the gradient of the objective is a linear combination of the gradients of the constraints with weights

$$\lambda = \begin{pmatrix} \frac{(2r-1)(\alpha^2-1)}{2r+\alpha-1} \\ \frac{(2r-1)(\alpha^2-1)}{2r+\alpha-1} \\ \frac{(2r-1)(\alpha^2-1)}{2r+\alpha-1} \\ \frac{(2r-1)(\alpha^2-1)}{2r+\alpha-1} \\ \frac{(r-1)(\alpha-1)}{2r+\alpha-1} \\ \frac{(r-1)(\alpha-1)}{2r+\alpha-1} \\ \frac{(r-1)(\alpha-1)}{2r+\alpha-1} \\ \frac{(r-1)(\alpha-1)}{2r+\alpha-1} \end{pmatrix}, \tag{17}$$

that are strictly positive when $r \in (1/2, 1)$ and $\alpha \in (0, 1)$. Thus σ^* is locally the unique solution to the maximization problem that defines the optimal mediation protocol. Due to the convexity of the constraint set and the linearity of the objective, σ^* is also the unique global solution in the set $\mathcal{M}$ of mediation protocols with independent action recommendations.

Remark B.2. *It is easy to see that the mediation protocol of Theorem B.1 can be implemented by a generalized almost-truthful interim-biased mediation protocol with*

$$y_i(s_i, s_{-i}) = \begin{cases} 0, & \text{if } s_i = s_{-i}; \\ \frac{\alpha}{2r-1+\alpha}, & \text{if } s_i \neq s_{-i}. \end{cases}$$