Information Design and Strategic Communication

Daniel Krähmer*

April 23, 2020

Abstract

I study sender receiver games where the receiver can disclose information to the sender by designing an information structure. I show that by secretely randomizing over information structures, the receiver can virtually attain her complete information payoff even for large conflicts of interest. The key insight is that private knowledge of the information structure induces truthful communication because it allows the receiver to cross check the consistency of the sender's report.

Keywords: sender receiver games, information design, cheap talk
JEL codes: C72, D82, D83

When an expert (he) is consulted by a decision maker (she), the expert frequently relies on information provided by the decision maker to come up with an assessment. If a firm hires a consultancy for advice where to invest, or how to restructure its internal organization, the consultancy's recommendation often relies on the internal data provided by the firm. Financial advisors use information provided by investors about their risk appetite and financial situation. In "expert judgement procedures" (Cooke (1991)), public bodies disseminate information to and collect risk assessments from selected experts.

*University Bonn, Department of Economics, Institute for Microeconomics, Adenauer Allee 24-42, D-53113 Bonn (Germany), kraehmer@hcm.uni-bonn.de. I thank the editor and three referees for excellent suggestions. I thank especially Francesc Dilme and Alex Smolin for insightful comments. I also thank Eugen Kovac, Stephan Lauermann, David Rodina, Rani Spiegler, Roland Strausz, Dezsö Szalay, and Tymon Tatur for useful discussions. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 and Germany's Excellence Strategy EXC 2126/1–390838866 is gratefully acknowledged.

Experts often distort their advice, because their preferred course of action differs from the decision maker's. When the expert's information is exogenously given, this conflict of interest typically impedes communication in equilibrium and prevents a fully informed decision. My main results in this article show that, in contrast, when the decision maker can endogenously design the expert's information, she can essentially attain the same outcome as if she had complete information under fairly weak conditions. This includes cases where the conflict of interest is so large that meaningful communication would be impossible with exogenous information.

The key insight, in a nutshell, is that if the decision maker keeps certain features of the information she provides hidden from the expert, she can induce the expert to provide undistorted advice by threatening to "punish" him should he make a recommendation that is inconsistent with these hidden features. For example, the decision maker could selectively release data of varying precision and, based on this selection, rule out some conclusions that the expert may pretend to have come to. Or, she may build in mistakes in the data, which, if reported as facts, would reveal that the expert is not telling the truth. In other words, the decision maker designs information so as to not only inform but also to monitor the expert.

In the spirit of Bayesian persuasion (Kamenica and Gentzkow (2011)) and information design (Bergemann and Morris (2019)), I model information disclosure by allowing the decision maker to choose any information structure that releases a signal to the expert which is correlated with the underlying state of the world. I capture expertise by assuming that signal realizations are the expert's private information. Expertise may stem from the fact that analyzing and interpreting data requires special knowledge or time-consuming effort. The novelty of my article is that I allow the decision maker to design what I call a *compound information structure*, that is, to secretely randomize over information structures, and then release a signal to the agent according to the realized information structure.

I identify conditions under which there are compound information structures that permit a cheap talk equilibrium in which the expert reports his information truthfully, and in which the true state is (nearly) fully revealed, allowing the decision maker to (nearly) attain her complete information payoff. These conditions are satisfied in many applications, including the classical specification of Crawford and Sobel (1982) for all positive biases; they also allow for state-independent preferences of the expert.

In the first part of the paper, I consider the benchmark scenario in which the decision maker is entirely unrestricted in designing information. This allows for a construction where an information structure is essentially an encryption code: the decision maker reveals the state privately to the expert in encrypted form, and later learns the state by decoding the expert's message using the encryption code. As this takes the decision maker's information design ability to a somewhat unrealistic extreme, in the second part of the paper, I impose the restriction that the signals generated by the information structure are reduced to their natural meaning in the sense that a signal corresponds to the posterior belief it induces. This rules out that the decision maker can inform the expert using an encrypted language and also implies that all decision-relevant information made available by the decision maker becomes the expert's private information.

In both settings, I present a construction so that (1) any information structure over which the decision maker randomizes is (almost) fully informative about the state, and (2) any deviation by the expert from truth-telling is inconsistent with the true information structure for sure. After a deviation, any of her rationalizable actions is therefore a best reply for the decision maker. Hence, a sufficient condition that ensures truth-telling is that the decision maker has a rationalizable action which yields the expert, at any belief possibly induced by the signal, a smaller expected utility, than the equilibrium outcome where the decision maker chooses her favorite action in any state.[1]

Similarities to my construction can be observed in practice in "structured expert judgment" procedures (see, e.g., Cooke (1991), Cooke and Goossens (2000)) where a public body consults experts to obtain probability assessments about a risky decision. The procedure involves an information disclosure part that describes to the experts the context of the decision at hand, possibly including the release of data, and an elicitation part where experts have to state both a probability assessment about variables of interest as well as answers to "seed" (or "calibration") questions. Seed questions are designed by, and their answers are known to the decision maker, and an expert's assessment is discounted if his performance in answering seed questions is poor. In this sense, the choice of seed questions in structured expert judgment serves a monitoring role similar to randomizing over information structures in my construction.

My article contributes to the literature that studies information disclosure by an uninformed decision maker to a (biased) expert by introducing the idea of randomizing over information structures. Ivanov (2010) and Deimen and Szalay (2019) show that when the decision maker

[1] In the benchmark case, the sufficient condition is also necessary.

cannot mix over information structures there is a trade-off between providing more information and an increased loss of control. This trade-off is eliminated with my approach. In Wei (2018) and Lipnowski, Mathevet and Wei (2020) the decision maker provides a signal, and the expert can observe any noisy version of this signal at a cost. While I abstract from moral hazard on the expert's side, in these papers the decision maker is not allowed to mix between signals.[2] The use of randomizing over information structures is explored in mechanism design settings by Zhu (2018) and Krähmer (2020) in which agents have additional, exogenously given private information. A key difference is that in a mechanism design setting, punishments after deviations do not have to be sequentially rational. The truth-telling logic in my benchmark case exploits that the expert is "confused", as his private information is useless by itself. This is reminiscent of Watson (1996) who makes this point in a setting with exogenous information. My work also shares similarities with Rahman and Obara (2010) and Rahman (2012) where, in the context of moral hazard in teams, randomized secret effort recommendations to team member A help to extract private information from team member B, because this allows for checking the consistency of B's report with the effort recommended to A.

The paper is organized as follows. Section I presents an example. Section II presents the general setup. Sections III and IV contain the main results. Section V concludes. All proofs are in the appendix.

I. Example

A decision maker, referred to as receiver, R, has to decide between three actions $a \in \{0, 1/2, 1\}$. There are two equally likely states of the world $\omega \in \{0, 1\}$. Before making the decision, R can hire an expert, referred to as sender, S, for advice. The payoffs from action a in state ω for R and S are

(1) $$u_R(a, \omega) = -(a - \omega)^2;$$

(2) $$u_S(a, \omega) = u_S(a) \quad \text{with} \quad u_S(0) = 0, \quad u_S(1/2) = \chi, \qquad u_S(1) = 1,$$

[2]In Bloedel and Segal (2018), it is the expert who provides a signal, and the decision maker has "reading costs". Mixing between signals by the expert is then pointless. In Lipnowski, Doron and Shishkin (2019), the expert both designs the information and can with an exogenous probability secretly manipulate the signal ex post. In my case, it is the decision maker who secretely manipulates the signal with an endogenous probability.

where $\chi < 1$. That is, R wants to match the state if she is sufficiently certain about the state, and to choose the moderate action $a = 1/2$ for intermediate beliefs $Pr(\omega = 1) \in [1/4, 3/4]$, while S's preferences are state-independent, and he prefers $a = 1$ in any state.

Suppose S relies on information provided by R to come up with informed advice (such as internal sales data if R is a firm, or information about her wealth if R is an investor). In contrast to S, R lacks the expertise to analyze and interpret the data and assess its relevance for the decision.[3] After S has assessed the information, he recommends an action in the form of a (cheap talk) message to R.

In the spirit of information design, suppose R can inform S by choosing any information structure, specifying a set of signals s and conditional probabilities $\pi_\omega(s)$ with which signal s is disclosed to S, conditional on ω.

Given S prefers R to always choose $a = 1$, there seems to be no point for R to provide information: Whatever signal S observes, given his biased interests and his lack of commitment to communicate truthfully, he will always recommend $a = 1$. Hence, it seems R should simply ignore the advice and choose her best uninformed decision.

I now show how, nevertheless, information provision can improve R's decision. The idea is that R may secretely manipulate the information in a way that allows her to cross-check the consistency of S's report with the information provided. Specifically, suppose R has a data set that reveals the true state, captured by an information structure π with two signals $s \in \{s_0, s_1\}$, and

$$\pi_0(s_0) = \pi_1(s_1) = 1. \tag{3}$$

Moreover, R has a second, "faulty" data set that only nearly reveals the state, captured by an information structure $\hat{\pi}$ with two signals $s \in \{s_0, \hat{s}_1\}$, and

$$\hat{\pi}_0(s_0) = 1 - \lambda, \quad \hat{\pi}_1(\hat{s}_1) = 1, \tag{4}$$

where $\lambda > 0$ is small. Figure 1 shows the beliefs induced by π and $\hat{\pi}$ and R's optimal action as a function of beliefs.[4]

R then secretely flips a coin to decide which data she supplies to S. S then privately observes the signal s without knowing, however, by which data set it is generated. Subsequently, S reports a message $m \in \{s_0, \hat{s}_1, s_1\}$ to R, and R chooses an action.

[3] Alternatively, analyzing the data may require costly effort by S where I abstract from moral hazard issues.

[4] Alternatively, a signal s can be identified with the posterior belief $\tilde{s} = Pr(\omega = 1 \mid s)$ it induces. Signal s_1 then

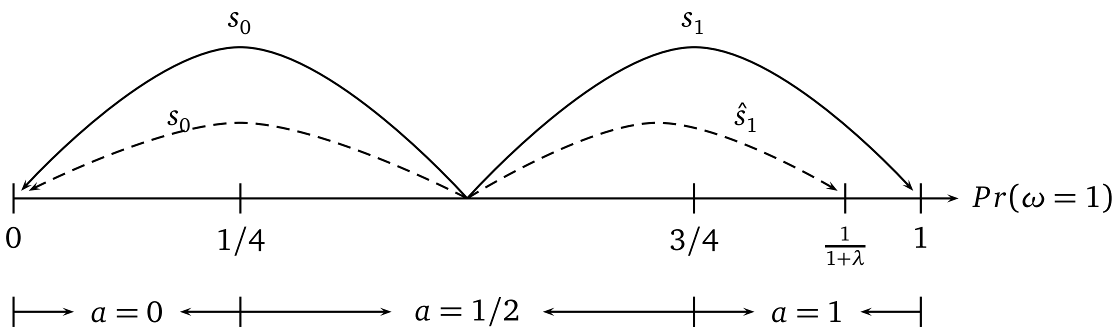

Figure 1: Posteriors induced by π (solid) and $\hat{\pi}$ (dashed), and R's optimal actions

To check whether there is an equilibrium in which S reports truthfully, (5) depicts a best reply by R in such an equilibrium. Note that a strategy by R specifies an action for any message and *any* information structure. In particular, if S reports truthfully, then under information structure π, a message s_0 (resp. s_1) reveals that the state is $\omega = 0$ (resp. $\omega = 1$), and R optimally chooses $a = 0$ (resp. $a = 1$). Crucially, however, in a truth-telling equilibrium, message $\hat{s}_1$ is never sent under, and is thus inconsistent with, information structure π. Assume that if R were nevertheless to receive message $\hat{s}_1$ under π, she would simply maintain her prior belief and thus choose $a = 1/2$.[5]

$$(5) \qquad a^*(\pi, m) = \begin{cases} 0 & \text{if} \quad m = s_0 \\ 1 & \text{if} \quad m = s_1 \\ 1/2 & \text{if} \quad m = \hat{s}_1 \end{cases}, \quad a^*(\hat{\pi}, m) = \begin{cases} 0 & \text{if} \quad m = s_0 \\ 1 & \text{if} \quad m = \hat{s}_1 \\ 1/2 & \text{if} \quad m = s_1 \end{cases}.$$

Next, consider S's optimal reporting strategy against (5). Truth-telling is optimal for S when having observed $\hat{s}_1$ and s_1, because this gets S his most preferred action $a = 1$. The key trade-off for S occurs when having observed s_0. In this case, S knows that the state is $\omega = 0$, and when he reports truthfully, $a = 0$ is chosen for sure, resulting in a payoff $U^* = 0$ for S.

On the other hand, when S "lies" and reports $\hat{s}_1$, this induces R to choose $a = 1$ under $\hat{\pi}$, but under π, the lie is detected and "punished" with action $a = 1/2$. More precisely, conditional on s_0, S believes with probability $1/(2-\lambda)$ that π has occurred and with probability $(1-\lambda)/(2-\lambda)$

corresponds to $\tilde{s} = 0$, signal $\hat{s}_1$ to $\tilde{s} = 1/(1+\lambda)$, and s_1 to $\tilde{s} = 1$.

[5]The best reply under $\hat{\pi}$ can be derived analogously. If $\hat{s}_1$ is reported under $\hat{\pi}$, R believes that $\omega = 1$ with probability $1/(1+\lambda)$ and her best reply is $a = 1$ if λ is sufficiently small.

that $\hat{\pi}$ has occurred.[6] Thus, if S reports $\hat{s}_1$, this is detected as a lie with probability $1/(2-\lambda)$, and his expected utility is

$$U(\hat{s}_1 \mid s_0) = \frac{1}{2-\lambda} \cdot \chi + \frac{1-\lambda}{2-\lambda} \cdot 1. \tag{6}$$

Similarly, if S reports s_1, his expected utility is

$$U(s_1 \mid s_0) = \frac{1-\lambda}{2-\lambda} \cdot \chi + \frac{1}{2-\lambda} \cdot 1. \tag{7}$$

Therefore, since $U^* = 0$, truth-telling is optimal if χ is sufficiently negative.

In a truth-telling equilibrium, the outcome differs from R's optimal complete information outcome only in the event that the state is $\omega = 1$, the information structure is $\hat{\pi}$, and signal s_0 is released. Since this event has probability $1/4 \cdot \lambda$, R's expected utility converges to her complete information payoff as λ goes to 0.

What if χ is not sufficiently negative for (6) and (7) to be negative? My results below show that even then the decision maker can attain her complete information payoff almost fully. The idea is to randomize over many information structures of the form of $\hat{\pi}$, each with a different, small λ. This will increase the probability to detect a deviation while maintaining that a truth-telling equilibrium is almost fully informative.

II. Setting

There is a receiver (she) and a sender (he). There is a finite set of states of the world Ω and a finite set of actions A.[7] Let $\mu_0 \in \Delta(\Omega)$ be the commonly known prior. The receiver and sender have preferences $v : A \times \Omega \to \mathbb{R}$ and $u : A \times \Omega \to \mathbb{R}$, respectively. For belief $\mu \in \Delta(\Omega)$, let the receiver's set of optimal actions be[8]

$$A_R(\mu) = \arg\max_{a \in A} E_\mu[v(a, \tilde{\omega})]. \tag{8}$$

[6] By Bayes' rule:

$$Pr(\pi \mid s_0) = \frac{[\pi_0(s_0) \cdot 1/2 + \pi_1(s_0) \cdot 1/2] \cdot 1/2}{[(\pi_0(s_0) \cdot 1/2 + \pi_1(s_0) \cdot 1/2] \cdot 1/2 + [(\hat{\pi}_0(s_0) \cdot 1/2 + \hat{\pi}_1(s_0) \cdot 1/2] \cdot 1/2} = \frac{1}{2-\lambda}.$$

[7] Allowing only for finite state and action spaces is to avoid technicalities. With continuous actions, Proposition 1 below holds unchanged, and Proposition 2 holds under standard conditions on preferences. With continuous states, approximation results can be obtained by discretizing the state space and applying Proposition 1 and 2 to the discretization. I omit the details.

[8] The subindex μ indicates the probability distribution with which the expectation is taken, a tilde within the expectation indicates the integration variable.

I make the (generically satisfied) assumption that the receiver has a unique optimal action if he knows that the state is ω and denote it by $a_R(\omega)$. Let $A_R = \cup_{\mu \in \Delta(\Omega)} A_R(\mu)$ be the set of the receiver's rationalizable actions.

Both parties are initially uninformed, but the receiver can, without observing herself, disclose information to the sender by choosing any information structure which provides the sender with a signal. A "simple" information structure is a joint probability distribution $\pi \in \Delta(\Omega \times S)$ over Ω and a set of signals S with the property that the marginal distribution over states is equal to the prior: $\pi(\{\omega\} \times S) = \mu_0(\omega)$. I denote by $\pi_\omega \in \Delta(S)$ the signal distribution, conditional on ω. Let $\mathscr{I}$ be the set of simple information structures.

The point of this paper is to show the power of a disclosure policy that randomizes between information structures. A "compound" information structure $\Pi \in \Delta(\mathscr{I})$ is a distribution over simple information structures.[9] The timing is as follows:

1. The receiver designs a compound information structure Π and a message set M, and Π and M become common knowledge.

2. A simple information structure π is drawn according to Π, and the receiver privately observes π.

3. The sender privately observes the signal s generated by π.

4. The sender sends a cheap talk message $m \in M$.

5. The receiver observes m and chooses an action.

My objective is to study conditions under which there is Π and M so that the cheap talk game induced at stage 4 has a Perfect Bayesian Equilibrium (PBE) in which the receiver exactly, or nearly, attains her complete information payoff. Given Π and M, a sender strategy σ specifies for all s a distribution $\sigma_s \in \Delta(M)$ over messages, and a receiver strategy τ specifies for all π and m a distribution $\tau_{(\pi,m)} \in \Delta(A)$ over actions. A belief system β for the receiver specifies for all π and m a belief $\beta_{(\pi,m)} \in \Delta(\Omega)$ over states.

A PBE is a combination $(\sigma, (\tau, \beta))$ where σ is a best reply for the sender given τ, τ specifies an optimal action for the receiver given β, and β is consistent with σ and π, that is, it is derived from σ by Bayes' rule for all information sets that can be reached under σ, given π.

[9] To keep notation simple, I take S to be the same for all π. The support of signals may, however, differ across π.

Given Π and M and a PBE, the receiver's expected utility in state ω is:

$$V(\omega) = \int_{\mathscr{I}} \int_S \int_M \int_A v(a,\omega)\, d\tau_{(\pi,m)}(a)\, d\sigma_s(m)\, d\pi_\omega(s)\, d\Pi(\pi). \tag{9}$$

I say the first-best is implementable if there is Π and M and a PBE so that

$$V(\omega) = v(a_R(\omega),\omega) \quad \text{for all } \omega \in \Omega. \tag{10}$$

I say the first-best is virtually implementable if for all $\epsilon > 0$, there is Π and M and a PBE so that

$$V(\omega) > v(a_R(\omega),\omega) - \epsilon \quad \text{for all } \omega \in \Omega. \tag{11}$$

III. Unrestricted information structures

In this section, I consider the benchmark in which the receiver is completely unrestricted in designing information structures.

Proposition 1. *The first-best is implementable if and only if there is $\underline{a} \in A_R$ so that*

$$E_{\mu_0}[u(\underline{a},\tilde{\omega})] \leq E_{\mu_0}[u(a_R(\tilde{\omega}),\tilde{\omega})]. \tag{12}$$

Condition (12) is fairly weak and satisfied in many applications. For example, it is satisfied if there is an action in A_R which is the worst action for the sender in all states (such as in the example of Section I). Condition (12) is met in the discrete state space version of Crawford and Sobel (1982)[10], and it also allows for state-independent (strict) preferences of the sender in which case, with exogenous information, only babbling equilibria would exist.

The sufficiency part of the proof employs the following compound information structure.[11] First, the receiver secretely draws $\lambda = (\lambda_1, \ldots, \lambda_{|\Omega|})$ from the uniform (product) distribution on $[0,1]^{|\Omega|}$. Second, conditional on ω and λ, the signal $s = \lambda_\omega$ is disclosed to the sender with probability 1. Note, since λ is uniformly distributed, the signal s by itself is entirely uninformative about the true ω, but jointly, λ and s fully reveal the state.

[10] In particular, let $\Omega = \{\omega_1, \ldots, \omega_N\}$, $0 \leq \omega_1 < \ldots < \omega_N \leq 1$, and $v(a,\omega) = -(a-\omega)^2$ and $u(a,\omega) = -(a - b(\omega) - \omega)^2$ with positive bias $b(\omega) > 0$. Note that $A_R = [\omega_1, \omega_N]$. Condition (12) is then satisfied, since the left hand side of (12), attained at $\underline{a} = \omega_1$, is equal to $-E_{\mu_0}[(\omega_1 - \tilde{\omega} - b(\tilde{\omega}))^2]$ which is smaller than $-E_{\mu_0}[b(\tilde{\omega})^2]$ on the right hand side.

[11] The same information structure is used in Krähmer (2020).

Consider now the cheap talk game with $M = [0,1]$. In a candidate truth-telling equilibrium, because the receiver knows λ, any message $m \notin \{\lambda_1, \ldots, \lambda_{|\Omega|}\}$ is "off-path" and detected as a lie by the receiver. In a PBE, the receiver may choose any rationalizable action after such a message, especially $\underline{a}$. Further, because the sender observes only $s = \lambda_\omega$, and λ is uniformly distributed, he believes that a deviation $m \neq s$ is off-path and detected as a lie with probability 1. Therefore, truth-telling is optimal if the receiver's off-path action gives the sender a smaller ex ante expected utility than when the equilibrium action $a_R(\omega)$ is chosen in any state. Because s is entirely uninformative, this is satisfied under (12) for the off-path action $\underline{a}$. Finally, in a truth-telling equilibrium, the receiver obtains the first-best, because jointly, λ and s fully reveal the state.

IV. Information structures with natural signals

One way to think of the previous construction is that the receiver designs a fully informative (simple) information structure (with signals $s = \omega$ say) and then reveals to the sender the resulting signal in encrypted form (using a private, random code λ). This allows the receiver both to detect lies and to identify the state by decoding the sender's true message. This takes the receiver's information design ability to a somewhat unrealistic extreme. Moreover, the compound information structure leaves the sender totally "confused" and endows the receiver with private information to interpret the sender's signal (similarly to Watson (1996)). This appears at odds with the intuitive notion that experts are hired because they possess superior skills in analyzing and interpreting information.

To address these concerns, I now impose the restriction that the receiver can inform the sender only using a non-encrypted language where signals are reduced to their natural meaning and correspond to the posterior beliefs they induce. This implies that once the receiver has designed a simple information structure, she cannot further encrypt the signals to confuse the sender. Formally, let $F^\pi \in \Delta(S)$ be the marginal over signals induced by π. I say a compound information structure Π has *natural signals* if a signal is a belief over Ω, that is, $S = \Delta(\Omega)$, and the posterior induced by a signal is equal to the signal:

$$(13) \qquad Pr(\omega \mid s, \pi) = s(\omega) \quad \text{for all } \omega \in \Omega, \pi \in \text{supp}\Pi, s \in \text{supp}F^\pi.$$

Next to ruling out encrypted signals, a compound information structure with natural signals has the property that observing the signal alone is sufficient to form beliefs about the state, and additional knowledge of the true simple information structure does not alter one's beliefs. This captures a situation in which all decision-relevant information made available by the receiver becomes the sender's private information.

Moreover, I say the first-best is (virtually) implementable with natural signals if in the definition of (virtual) implementation, the compound information structure is required to be one with natural signals.

The example in Section I can be rephrased in terms of natural signals (see footnote 4). Moreover, the first-best is generally at most virtually, yet not exactly, implementable with natural signals. Indeed, with natural signals, the first-best requires that the state is perfectly revealed to the sender, and that the sender perfectly reveals it to the receiver. This is only possible when the sender and receiver have the same preferred action in any state. The main result of this section is:

Proposition 2. *The first-best is virtually implementable with natural signals if there is $\underline{a} \in A_R$ so that*

$$u(\underline{a}, \omega) < u(a_R(\omega), \omega) \quad \text{for all } \omega \in \Omega, \tag{14}$$

or if there is a state $\hat{\omega}$ so that

$$u(a_R(\hat{\omega}), \omega) < u(a_R(\omega), \omega) \quad \text{for all } \omega \neq \hat{\omega}. \tag{15}$$

Condition (14) is satisfied in the example in Section I whenever $\chi < 0$. Condition (15) is satisfied in the example in Section 2 whenever $\chi < 1$, as well as in the discrete state space version of Crawford and Sobel (1982).[12] Both conditions allow for state-independent preferences of the sender.

I illustrate the construction behind Proposition 2 for the two states case $\omega \in \{1, 2\}$. For $\lambda = (\lambda_1, \lambda_2) \in (0, 1)^2$, define the simple information structure π^λ with the two possible signals $s \in \{\lambda_1, \lambda_2\}$ so that signals correspond to the posteriors they induce: $Pr(\omega = 2 \mid s = \lambda_i, \pi^\lambda) = \lambda_i$.[13]

The compound information structure is defined as follows: First, the receiver secretely draws $\lambda = (\lambda_1, \lambda_2)$ from the uniform (product) distribution on $[0, \bar{\lambda}] \times [1 - \bar{\lambda}, 1]$ for some small $\bar{\lambda}$.

[12] Recall the notation in footnote 10. For $\hat{\omega} = \omega_1$, we have $a_R(\hat{\omega}) = \omega_1$, and the left hand side of (15) is equal to $-(\omega_1 - \omega - b(\omega))^2$ which is smaller than $-b(\omega)^2$ on the right hand side for all $\omega \neq \omega_1$.

[13] π^λ is well-defined if the prior is in the convex hull of λ_1 and λ_2.

Second, signal s is disclosed to the sender according to π^λ. By construction, this compound information structure displays natural signals.

Consider now the cheap talk game where the sender reports a message from the set of possible posteriors $M = [0, \bar{\lambda}] \cup [1 - \bar{\lambda}, 1]$. If the sender reports truthfully, the receiver's utility gets arbitrarily close to her complete information payoff if $\bar{\lambda}$ is sufficiently small, because then any π^λ induces posteriors at which the state is known with near certainty.

To see that truth-telling is an equilibrium for small $\bar{\lambda}$ under (14) and (15), observe that because the receiver knows λ, any message $m \notin \{\lambda_1, \lambda_2\}$ is "off-path" and detected as a lie by the receiver. In a PBE, the receiver may therefore choose any rationalizable action after such message. Further, because the sender observes only $s = \lambda_\omega$, and λ is uniformly distributed, he believes that a deviation $m \neq s$ is off-path and detected as a lie with probability 1. Since $\bar{\lambda}$ is small, the sender, moreover, essentially knows the state. Therefore, truth-telling is optimal if in any state ω, the receiver's off-path action gives the sender a smaller utility than the equilibrium action $a_R(\omega)$ in that state.[14] Under (14), this is satisfied for the off-path action $\underline{a}$, and under (15) for the off-path action $a_R(\hat{\omega})$.

In the constructions behind Propositions 1 and 2, any deviation from truth-telling by the sender is "off path", and therefore, in a PBE, off-path beliefs can be assigned so as to maximally punish the sender. This raises the concern that these off-path beliefs may not survive standard refinements. I now illustrate that for any compound information structure that (possibly virtually) implements the first-best, there is a modified compound information structure which virtually implements the first-best in a PBE in which all messages are on path. The idea is to disclose to the sender a noisy full support signal with a small probability.

Consider $\Omega = \{1, 2\}$ and suppose Π implements the first-best with a message set $M = S$ and a PBE in which the sender tells the truth. Define the modified compound information structure $\tilde{\Pi}$ as follows.

- Step 1: π is drawn according Π.
- Step 2: Conditional on π and ω, the signal is drawn according to the simple information structure $\tilde{\pi}$, defined as follows:

[14]Because A is finite and $a_R(\omega)$ is uniquely optimal in ω, $a_R(\omega)$ is also optimal if the receiver's belief that the state is ω is almost 1.

– with probability ϵ_ω, the signal is "noisy", that is, s is uniformly distributed on S.

– with probability $1-\epsilon_\omega$, the signal s is distributed with π_ω.

Consider the cheap talk game induced by $\tilde{\Pi}$ with $M = S$. Because for any $\tilde{\pi}$ the support of signals is S, in the candidate truth-telling equilibrium, all messages in M are in the support of equilibrium messages, given $\tilde{\pi}$, and thus on path.

Moreover, after a message $m \notin \cup_\omega \mathrm{supp}\pi_\omega$, the receiver infers that the signal must have been noisy and updates her beliefs to

$$Pr(\omega \mid m) = \frac{\epsilon_\omega \mu_0(\{\omega\})}{\epsilon_1 \mu_0(\{1\}) + \epsilon_2 \mu_0(\{2\})}. \tag{16}$$

By choosing ϵ_1 and ϵ_2 appropriately, any belief can be induced after such a message. In particular, by inducing the beliefs that are the off-path beliefs in the cheap talk game under the original Π, the receiver's best reply remains unchanged. Therefore, truth-telling remains a PBE under $\tilde{\Pi}$.

Finally, by letting ϵ_1 and ϵ_2 go to zero, the receiver's payoff under $\tilde{\Pi}$ converges to her payoff under Π. Consequently, $\tilde{\Pi}$ virtually implements the first-best.

V. Conclusion

The paper studies sender receiver games where the receiver can disclose information to the sender by designing an information structure. By secretely randomizing over information structures, the receiver can virtually attain her complete information payoff even for large conflicts of interest. The key insight is that private knowledge of the information structure allows the receiver to cross-check the consistency of the sender's report. Hence, information design serves not only to inform but also to monitor the sender.

A Appendix

Proof of Proposition 1 *Sufficiency*: The following two steps define the compound information structure:

1. $\lambda \in [0,1]^{|\Omega|}$ is drawn according to the uniform (product) distribution on $[0,1]^{|\Omega|}$, and the receiver observes the outcome. (Any λ corresponds to a simple information structure.)

2. Conditional on λ and ω, the sender observes $s = \lambda_\omega$ with probability 1.

Note that given λ and $s = \lambda_\omega$, the state is ω with probability 1.[15]

Consider the cheap talk game with $M = [0,1]$ and a candidate truth-telling equilibrium ($m = s$). Given λ, if $m = \lambda_\omega$ for some ω, the receiver infers that the state is ω, and her optimal action is $a_R(\omega)$. Moreover, any message $m \notin \{\lambda_1, \ldots, \lambda_{|\Omega|}\}$ is off path given λ, and in a PBE may induce a receiver belief which rationalizes $\underline{a} \in A_R$. Hence, a best reply for the receiver to truth-telling by the sender is:

$$a^*(\lambda, m) = \begin{cases} a_R(\omega) & \text{if} \quad m = \lambda_\omega \quad \text{for some } \omega \\ \underline{a} & \text{else} \end{cases}. \tag{A.1}$$

I now argue that the sender's best reply to (A.1) is truth-telling. Indeed, since the signal is entirely uninformative for the sender, his belief about the state is unchanged. Moreover, if the sender reports truthfully, the receiver chooses $a_R(\omega)$ in state ω, thus yielding the sender expected utility (conditional on s) $E_{\mu_0}[u(a_R(\tilde{\omega}), \tilde{\omega})]$.

On the other hand, if the sender deviates by reporting $m \neq s$, then since he does not observe λ_θ, $\theta \neq \omega$, and because λ is uniformly distributed, he assigns probability 0 to the event that m coincides with some λ_θ, $\theta \neq \omega$. Accordingly, the sender expects $\underline{a}$ to be chosen with probability 1, yielding expected utility $E_{\mu_0}[u(\underline{a}, \tilde{\omega})]$. By (12), truth-telling is thus a best reply to (A.1).

Finally, because the state is fully revealed in a truth-telling equilibrium, the first-best is implemented.

Necessity: Towards a contradiction, suppose (12) is violated:

$$E_{\mu_0}[u(a, \tilde{\omega})] > E_{\mu_0}[u(a_R(\tilde{\omega}), \tilde{\omega})] \quad \text{for all } a \in A_R. \tag{A.2}$$

Moreover, suppose there is S and $\Pi \in \Delta(\mathscr{I})$ and M and a PBE $(\sigma, (\tau, \beta))$ of the induced cheap talk game in which the receiver gets her complete information payoff. In a PBE, we have

$$a \in A_R \quad \text{for all } a \in \text{supp}(\tau_{(\pi,m)}), \pi \in \text{supp}(\Pi), m \in M. \tag{A.3}$$

To establish the contradiction, I argue that there is $s \in S$ and $\hat{m} \in M$ so that the sender is better off by deviating to $\hat{m}$ rather than using an equilibrium message $m^*(s) \in \text{supp}(\sigma_s)$.

Indeed, because the receiver attains her complete information payoff, the sender's utility from reporting $m^*(s) \in \text{supp}(\sigma_s)$ if he has observed s is

$$U(m^*(s) \mid s) = E[u(a_R(\tilde{\omega}), \tilde{\omega}) \mid s]. \tag{A.4}$$

[15] Because λ is jointly uniformly distributed, the event that $\lambda_\omega = \lambda_{\hat{\omega}}$ for some $\omega \neq \hat{\omega}$ has probability 0.

On the other hand, if he reports $\hat{m}$, his expected utility is

$$U(\hat{m} \mid s) = \int E\left[\int u(a,\tilde{\omega})\, d\tau_{(\pi,\hat{m})}(a) \mid \pi, s\right] d\Pi(\pi \mid s). \tag{A.5}$$

Integration over s with respect to the marginal signal distribution induced by Π, denoted $F \in \Delta(S)$, delivers:

$$\int U(\hat{m} \mid s)\, dF(s) = \int E\left[\int u(a,\tilde{\omega})\, d\tau_{(\pi,\hat{m})}(a) \mid \pi\right] d\Pi(\pi) \tag{A.6}$$

$$= \int E_{\mu_0}\left[\int u(a,\tilde{\omega})\, d\tau_{(\pi,\hat{m})}(a)\right] d\Pi(\pi) \tag{A.7}$$

$$> E_{\mu_0}[u(a_R(\tilde{\omega}),\tilde{\omega})] \tag{A.8}$$

$$= \int U(m^*(s) \mid s)\, dF(s), \tag{A.9}$$

where the second line is due to π and ω being stochastically independent, the third line follows from (A.2) and (A.3), and the final line from (A.4).

Hence, there is a signal s so that $U(\hat{m} \mid s) > U(m^*(s) \mid s)$, a contradiction. □

Proof of Proposition 2 I first assume that (14) is true. Let $\epsilon > 0$. Define

$$\eta = \frac{\epsilon}{2 \cdot \max_{\omega\in\Omega}[v(a_R(\omega),\omega) - \min_{a\in A} v(a,\omega)]}. \tag{A.10}$$

I next define the simple information structures over which randomization will take place. For $\bar{\lambda} \in (0,1)$, and $\lambda = (\lambda_1, \ldots, \lambda_{|\Omega|}) \in [0,\bar{\lambda}]^{|\Omega|}$, define for every ω the belief $\mu_\omega^\lambda \in \Delta(\Omega)$ via

$$\mu_\omega^\lambda(\theta) = \begin{cases} 1-\lambda_\omega & if \quad \theta = \omega \\ \frac{\lambda_\omega}{|\Omega|-1} & if \quad \theta \neq \omega \end{cases}. \tag{A.11}$$

Let $\delta_\omega \in \Delta(\Omega)$ be the belief which puts probability 1 on ω. Then for all ω:

$$\mu_\omega^\lambda \to \delta_\omega \quad as \quad \bar{\lambda} \to 0. \tag{A.12}$$

Hence, for sufficiently small $\bar{\lambda}$:

$$\mu_0 \in \text{Conv}\{\mu_1^\lambda, \ldots, \mu_{|\Omega|}^\lambda\}. \tag{A.13}$$

Define π^λ as the simple information structure with signal space $S = \Delta(\Omega)$ which induces a distribution of posterior beliefs with support $\{\mu_1^\lambda, \ldots, \mu_{|\Omega|}^\lambda\}$. It is well known that such an information structure exists under (A.13). Moreover, (A.12) implies that conditional on ω, the signal $s = \mu^\lambda_\omega$ is released with probability close to 1 as $\bar{\lambda}$ gets close to 0.[16] Hence, there is $\bar{\lambda}_A$ so that:

$$\text{(A.15)} \qquad \pi^\lambda_\omega(\mu^\lambda_\omega) > 1-\eta, \quad \text{and} \quad \pi^\lambda_\omega(\mu^\lambda_\theta) < \frac{\eta}{|\Omega|-1} \quad \text{for all } \lambda \in [0,\bar{\lambda}_A]^{|\Omega|}, \theta \neq \omega.$$

Because the action space is finite and $a_R(\omega)$ is the receiver's uniquely optimal action at belief δ_ω, (A.12) implies that there is $\bar{\lambda}_B \leq \bar{\lambda}_A$ so that for all ω:

$$\text{(A.16)} \qquad A_R(\mu^\lambda_\omega) = \{a_R(\omega)\} \quad \text{for all } \lambda \in [0,\bar{\lambda}_B]^{|\Omega|}.$$

Let

$$\text{(A.17)} \qquad T(\omega) \;=\; \sum_{\theta\neq\omega} u(a_R(\omega),\theta) - u(\underline{a},\theta).$$

Then by (14), there is $\bar{\lambda}_C \leq \bar{\lambda}_B$ so that for all ω:

$$\text{(A.18)} \qquad (1-\lambda_\omega)(u(a_R(\omega),\omega)-u(\underline{a},\omega)) + \frac{\lambda_\omega}{|\Omega|-1}T(\omega) > 0 \quad \text{for all } \lambda \in [0,\bar{\lambda}_C]^{|\Omega|}.$$

Finally, given $\epsilon, \eta, \bar{\lambda}_C$, the compound information structure Π is defined in two steps.

1. λ is drawn according to the uniform (product) distribution on $[0,\bar{\lambda}_C]^{|\Omega|}$.

2. Second, conditional on λ, the signal s is drawn according to π^λ.

By construction, Π has natural signals.

Consider now the cheap talk game where the sender reports a belief, $M = \Delta(\Omega)$, and a candidate truth-telling equilibrium ($m = s$). Given π^λ, if $m = \mu^\lambda_\omega$ for some ω, the receiver's optimal action is $a_R(\omega)$ by (A.16). Moreover, any message $m \notin \{\mu^\lambda_1,\ldots,\mu^\lambda_{|\Omega|}\}$ is off path given π^λ, and in a PBE may induce a receiver belief which rationalizes $\underline{a} \in A_R$. Hence, a best reply for the receiver to truth-telling by the sender is:

$$\text{(A.19)} \qquad a^*(\pi^\lambda,m) = \begin{cases} a_R(\omega) & \text{if} \quad m = \mu^\lambda_\omega \quad \text{for some } \omega \\ \underline{a} & \text{else} \end{cases}.$$

[16] In fact, by Bayes' rule we have for all ω,θ:

$$\text{(A.14)} \qquad \pi^\lambda_\omega(\mu^\lambda_\theta) = \frac{\mu^\lambda_\theta(\omega)}{1-\mu^\lambda_\theta(\omega)} \frac{\sum_{\tilde{\omega}\neq\omega}\pi^\lambda_\omega(\mu^\lambda_\theta)\mu_0(\tilde{\omega})}{\mu_0(\omega)}.$$

Hence, as $\mu^\lambda_\theta(\omega) \to 0$ for all $\theta \neq \omega$, this implies that $\pi^\lambda_\omega(\mu^\lambda_\omega) \to 1$.

Next, I argue that truth-telling is a best reply to (A.19) for the sender. Indeed, if the sender has observed μ_ω^λ and reports truthfully, he obtains

$$U^*(\omega) = (1-\lambda_\omega)u(a_R(\omega),\omega) + \sum_{\theta\neq\omega} \frac{\lambda_\omega}{|\Omega|-1} u(a_R(\omega),\theta). \tag{A.20}$$

If he deviates, $m \neq \mu_\omega^\lambda$, then since he does not observe μ_θ^λ, $\theta \neq \omega$, and since λ is uniformly distributed, he assigns probability 0 to the even that m coincides with some μ_θ^λ, $\theta \neq \omega$. Accordingly, the sender expects $\underline{a}$ to be chosen with probability 1 after a deviation, yielding

$$U^d(\omega) = (1-\lambda_\omega)u(\underline{a},\omega) + \sum_{\theta\neq\omega} \frac{\lambda_\omega}{|\Omega|-1} u(\underline{a},\theta). \tag{A.21}$$

Thus, truth-telling is indeed a best reply, because by (A.18):

$$U^*(\omega) - U^d(\omega) = (1-\lambda_\omega)(u(a_R(\omega),\omega) - u(\underline{a},\omega)) + \frac{\lambda_\omega}{|\Omega|-1} T(\omega) > 0. \tag{A.22}$$

The receiver's utility in this PBE is

$$\begin{aligned}
V(\omega) &= \sum_\theta \pi_\omega^\lambda(\mu_\theta^\lambda) v(a_R(\theta),\omega) && \text{(A.23)}\\
&\geq v(a_R(\omega),\omega) + \sum_{\theta\neq\omega} \pi_\omega^\lambda(\mu_\theta^\lambda)[\min_{a\in A} v(a,\omega) - v(a_R(\omega),\omega)] && \text{(A.24)}\\
&\geq v(a_R(\omega),\omega) + \eta[\min_{a\in A} v(a,\omega) + v(a_R(\omega),\omega)] && \text{(A.25)}\\
&> v(a_R(\omega),\omega) - \epsilon, && \text{(A.26)}
\end{aligned}$$

where (A.24) is straightforward, (A.25) follows from (A.15), and (A.26) from (A.10). Consequently, the first-best is virtually implemented.

The proof for case (15) is identical with $\underline{a}$ replaced by $a_R(\hat{\omega})$. □

References

Bergemann, Dirk, and Stephen Morris. 2019. "Information Design: A Unified Perspective." *Journal of Economic Literature*, 57: 44–95.

Bloedel, Alexander W., and Ilya Segal. 2018. "Persuasion with Rational Inattention." *Working paper, Stanford University*.

Cooke, Roger M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford University Press on Demand.

Cooke, Roger M., and Louis H. J. Goossens. 2000. “Procedures Guide for Structured Expert Judgement.” *European Commission. Report EUR 18820*.

Crawford, Vincent P., and Joel Sobel. 1982. “Strategic Information Transmission.” *Econometrica*, 50: 1431–1451.

Deimen, Inga, and Dezsö Szalay. 2019. “Information and Communication in Organizations.” *American Economic Review p&p*, 109: 545–549.

Ivanov, Maxim. 2010. “Informational Control and Organizational Design.” *Journal of Economic Theory*, 145: 721–751.

Kamenica, Emir, and Matthew Gentzkow. 2011. “Bayesian Persuasion.” *American Economic Review*, 101: 2590–2615.

Krähmer, Daniel. 2020. “Information Disclosure and Full Surplus Extraction in Mechanism Design.” *Journal of Economic Theory*, 187. https://doi.org/10.1016/j.jet.2020.105020.

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei. 2020. “Attention Management.” *American Economic Review Insights*, 2: 17–23.

Lipnowski, Elliot, Ravid Doron, and Denis Shishkin. 2019. “Persuasion Via Weak Institutions.” *Working paper, Columbia University*.

Rahman, David. 2012. “But Who Will Monitor The Monitor?” *American Economic Review*, 102: 2767–2797.

Rahman, David, and Ichiro Obara. 2010. “Mediated Partnerships.” *Econometrica*, 78: 285–308.

Watson, Joel. 1996. “Information Transmission when the Informed Party Is Confused.” *Games and Economic Behavior*, 12: 143–161.

Wei, Dong. 2018. “Persuasion Under Costly Learning.” *Working paper, Berkeley University*.

Zhu, Shuguang. 2018. “Private Disclosure with Multiple Agents.” *Working paper, Toulouse School of Economics*.