Hear-No-Evil Equilibrium

aa2336

September 2021

Abstract

A Bayes correlated equilibrium of an incomplete information game (G, S) is any outcome that an omniscient mediator could induce with an appropriate information design policy. In this paper I introduce a refinement of Bayes correlated equilibrium called "hear-no-evil Bayes correlated equilibrium" which further requires that no player can make themself better off by publicly committing to ignore the mediator's message. I provide a simple example in which an agent in a contest strictly prefers not to learn his opponent's type. I show that the set of hear-no-evil equilibria of an incomplete-information game (G, S) is shrinking in S if and only if the set of hear-no-evil equilibria of (G, S) is the same as the set of Bayes correlated equilibrium of (G, S). I provide an example of a BCE which satisifies the HNE condition but is Pareto-dominated by a BCE which fails the condition. I also show that in the parameterized 2x2x2 coordination game of Taneva (2019), all the symmetric BCE satisfy the HNE condition.

1 Introduction

Consider the following example.[1] Two firms must sequentially make a choice between researching technology A and researching technology B. Consumer preferences are not known, and ex ante both firms believe that each technology is equally likely to matter more to consumers. If the two firms invest in different technologies, whichever firm chooses the consumers' preferred technology earns monopoly profits of \$6, and the other firm earns \$0. If the two firms invest in the same technology, they split the market and earn duopoly profits of \$2 each. The two firms are identical, except that one firm must make their investment decision first, and the second firm can observe this decision before making their own. Assume that each firm seeks only to maximize their expected profit. Absent any further information, the first firm will choose a technology at random to invest in, and the second firm will invest in the other technology. This yields an expected profit of \$3 to each firm. Now suppose a consultant makes a public offer to reveal consumer preferences to the first firm and the first firm only. If the first firm accepts, they will invest in the consumers' preferred technology,

[1] A version of this example first appears in Kamien, Tauman, and Zamir (1990)

and the second firm will mimic this choice, guaranteeing each firm profits of \$2. Here, both firms strictly prefer that the first firm is not able to observe the consultant's report before making its investment decision.

Bergemann and Morris (2016) introduce the concept of Bayes correlated equilibrium to discuss the outcomes of an incomplete-information game that can be induced by an omniscient mediator. They show that the mediator can induce players to follow any "obedient" decision rule, in the sense that upon hearing the mediator's recommendation, each player weakly prefers to follow it. Since the set of obedient decision rules is defined by these incentive-compatibility constraints, Bergemann and Morris (2016) show that as the players' baseline level of information increases, the set of obedient decision rules is shrinking.

In the earlier example, if the mediator reveals the state of the world to the first player alone, this induces the obedient decision rule in which the first player correctly guesses the state and the second player mimics him. Implicitly this assumes that the first player has no choice but to hear the mediator's recommendation - while he prefers to follow the recommendation once he hears it, he strictly prefers not to hear it in the first place.

In this paper I introduce the concept of "hear-no-evil" equilibrium, a refinement of Bayes correlated equilibrium in which players not only prefer to follow recommendations they are given, but also would choose to hear the recommendations if given a choice. I show that the set of hear-no-evil equilibria is shrinking in the players' baseline information if and only if every Bayes correlated equilibrium of the game is also a hear-no-evil equilibrium. To illustrate the concept I provide an example of a contest in which one player strictly prefers not to be informed of his opponent's valuation of winning. I also provide an example where a hear-no-evil equilibrium is Pareto-dominated by a Bayes correlated equilibrium that fails the hear-no-evil condition. Finally, I study the parameterized binary coordination game of Taneva (2019) [11], for which I show that every symmetric Bayes correlated equilibrium is also a hear-no-evil equilibrium.

This equilibrium concept is salient for a few reasons. One can think of a group of friends playing a card game or board game around a table. The designer of the game may find it desirable that every BCE of the incomplete information games that arise satisfies the hear-no-evil condition, so that no player has an incentive to convey harmful information to their opponents through table talk. One can also think of a regulator who must decide what censorship policy to impose on some channel. A naive solution might be to censor any information that harms the agent receiving it. However, as I show in section 4, this may be very harmful to social welfare compared to a no-censorship approach.

There is another interpretation of the model presented here. If we know that agents have some baseline level of information S, but we are not sure what information they possess or can acquire beyond that, the set of BCE is the set of possible outcomes that could arise. Thus we can use the set of BCE to make robust predictions about what we can expect to see occur for any information costs the agents may have. This assumes that information acquisition is *covert* - not only do players not know what other players have learned, they also do not know what questions the other players have asked. Here, we can use the hear-no-evil condition to further restrict the set of outcomes that may arise when information acquisition is *overt*. When players do not know what other players have learned, but do know what questions they have asked, the set of outcomes that could arise for arbitrary information costs is exactly the set of HNEBCE. Such situations can arise when information acquisition is capital-intensive or otherwise difficult to hide. For example, the public may not know the findings of Uber's research into self-driving cars, but the public is aware that Uber is conducting this research.

The rest of this paper is structured as follows. The remainder of this section provides a brief review of related literature. Section 2 defines the sets of Bayes correlated equilibria and hear-no-evil equilibria, and includes the main result. Section 3 provides an example to illustrate the concept of hear-no-evil equilibrium, where I state the constraints defining the set of BCE, as well as the additional constraint required for the set of HNEBCE. Section 4 describes a counterexample showing that restricting attention to the set of HNEBCE can harm agent welfare. Section 5 considers a parameterized 2x2x2 coordination game, and shows that all symmetric BCE satisfy the hear-no-evil condition. Section 6 concludes.

1.1 Related Literature

The seminal works of Blackwell (1951, 1953) [5] [4] show that in a standard decision problem under uncertainty (i.e. a single player Bayesian game), giving an agent access to more information is equivalent to expanding the set of expected payoff vectors over which the agent can choose. In this case an expected-utility maximizer always prefers more information to less.

However, in environments with more than one player, having access to more information need not make agents better off. Perhaps the first example of this in the literature comes from Hirshleifer (1971) [7]. In a pure exchange economy with two states of the world and risk-averse traders, where some agents are disproportionately endowed with claims that pay out in state a and others disproportionately endowed with claims that pay out in state b, absent information agents will trade contingent claims to smooth their consumption between the two states of the world. However, if the state of the world is revealed before trade can take place, the ability to smooth consumption through trade vanishes. Here, such a society of risk-averse agents would collectively be willing to pay to suppress the revelation of this information until after trade has taken place. While Hirshleifer focuses on the social cost of information in this example, we can similarly think of a single risk-averse agent who prefers not to be informed of the state of nature, knowing that if he is informed, no other agent will be willing to trade with him, as in Akerlof (1970) [1].

Neyman (1991) [8] points out that learning some information can only make a player worse off insofar as other players are aware of it: If there's a chance that I may receive some private information and everybody is aware of this, I am better off in the realizations where I do receive the private information than those in which I do not. With this in mind, when I talk about an agent preferring not to receive some information, I mean that the agent prefers to play a version of the game in which he never receives the information.

Bassan, Scarsini, and Zamir (1997) [3] provide a number of examples of games where the value of information may be negative.

Bassan et al (2003) [2] finds a condition under which refining the information structure of a game leads to a Pareto-improvement of equilibrium outcomes. Namely, if a game has a unique Pareto-optimal outcome, then that outcome is supported in a Nash equilibrium, and that outcome Pareto-dominates any Nash equilibrium that arise from a coarsening of the information structure of the game. Additionally, if the Pareto frontier of a game is not a singleton, there exists a coarsening of the game's information structure with a Nash equilibrium that at least one agent strictly prefers to any equilibrium of the original game.

Giving players the ability to commit to ignoring their information has a parallel in giving players the ability to commit to ruling out certain actions, as in the commitment games of Renou (2009) [9].

Recently it has come up in the information design literature that a receiver may be made worse off by refining his private information. Kolotilin (2018) includes a school-employer example in which the school chooses a more rigorous grading scheme when the employer's interview is less informative, to the extent that the employer prefers to have access only to the less informative interview, instead of the more informative interview. Roesler and Balaz-Szentes (2017) [10] present an example of bilateral trade in which the buyer prefers to observe a garbling of his true valuation, rather than learning his true valuation, of the seller's good.

2 Model

2.1 Bayes Correlated Equilibrium

The set of players is finite and given by $1, 2, ..., I$ with typical player i. The set of states Θ is finite with typical element θ.

A *basic game* $G = ((A_i, u_i)_{i=1}^I, \psi)$ contains for each player i a finite action set A_i and a utility function $u_i : A_1, \times ... \times A_I \times \Theta \to \mathbf{R}$, as well as a full-support prior belief $\psi \in \Delta_{++}(\Theta)$ shared by all agents, according to which $\theta \in \Theta$ is distributed.[2]

An *information structure* $S = ((T_i)_{i=1}^I, \pi)$ consists of a finite set T_i of types for each player i and a signal distribution $\pi : \Theta \to \Delta(T_1 \times ... \times T_I)$. Each "type" of agent is distinguished by their beliefs about the state of the world and higher-order beliefs about the types of other players. We can refer interchangeably to an agent i as being of type t_i, and as having observed signal realization t_i.

As in Bergemann and Morris (2016), among others, let the pair of basic game G and information structure S define a game of incomplete information (G, S). Write $A = A_1 \times ... \times A_I$ and $T = T_1 \times ... \times T_I$.

[2] Here $\Delta_{++}(\Theta)$ is the set of probability distributions over Θ for which all elements are assigned strictly positive probability.

Given a game (G, S), a *decision rule* $\sigma : T \times \Theta \to \Delta(A)$ can be thought of as a strategy for a mediator who observes the realization of the state $\theta \in \Theta$ and player types $t \in T$, who then sends an action recommendation $a_i \in A_i$ to each player i.

A decision rule σ is *obedient* if, for each agent i, given any realization of t_i and action recommendation a_i, the agent is weakly better off choosing action a_i than any other action $a_i' \in A_i$.

A decision rule σ is defined to be a Bayes Correlated Equilibrium for (G, S) if it is obedient for (G, S). Let $BCE(G, S)$ denote the set of Bayes correlated equilibria of the game of incomplete information (G, S).

One of the main results of Bergemann and Morris (2016) links the set of Bayes correlated equilibria of (G, S) to the set of outcomes that can arise when players may view other information structures in addition to S. In order to state and discuss their results, I will first establish a few definitions about combinations of information structures.

If an agent observes two signals, this can be written equivalently as that agent observing a single "combined" signal. Define $S^* = (T^*, \pi^*)$ to be a *combination* of two information structures $S^1 = (T^1, \pi^1), S^2 = (T^2, \pi^2)$ if $T_i^* = T_i^1 \times T_i^2$ for all i, and the marginal distributions over T_1 and T_2 are preserved, that is,

$$\sum_{t^2 \in T^2} \pi^*(t^1, t^2|\theta) = \pi^*(t^1|\theta) \quad \forall t^1 \in T^1, \theta \in \Theta,$$

$$\sum_{t^1 \in T^1} \pi^*(t^1, t^2|\theta) = \pi^*(t^2|\theta) \quad \forall t^2 \in T^2, \theta \in \Theta.$$

Because no restrictions are placed on the correlation between S^1 and S^2, there can be many possible information structures resulting from the combination of the two.

Say that S^* is an *expansion* of S^1 if there is some other information structure S^2 such that S^* is a combination of S^1 and S^2. Note that every information structure S is an expansion of itself - specifically it is the combination of itself and a completely uninformative information structure S'. For an example of such an uninformative information structure, consider $S' = ((T_i')_{i=1}^I, \pi')$ where $|T_i| = 1$ for all i.

Say that S^* is *individually sufficient* for S if and only if there exists an expansion of S that has the same canonical representation as S^*.[3] Where this is the case I will write $S^* \succ S$.

Bergemann and Morris (2016) prove two main results about Bayes correlated equilibria. First, they show that if σ is a BCE of (G, S), then there is an expansion S^* of S such that there is a Bayes Nash equilibrium of (G, S^*) in which agents act as if according to σ. Second, they prove that the set of BCE

[3] Note that Bergemann and Morris (2016) start by defining individual sufficiency in a different manner, but then show that their definition is equivalent to this one. The concept of a "canonical representation" of an information structure comes from Mertens and Zamir (1985). This is the representation such that no two types share both the same beliefs about the state and higher-order beliefs about the beliefs of other players.

of (G, S^*) is a subset of the set of BCE of (G, S) for every basic game G if and only if S^* is individually sufficient for S.

2.2 "Hear-No-Evil" Equilibrium

Fix a basic game G and two information structures $S = (T, \pi)$, $S^* = (T^*, \pi^*)$ such that S^* is an expansion of S. This means there is some other information structure $S' = (T', \pi')$ such that S^* is a combination of S and S', so we can write $T^* = T \times T'$. As we know, it is not necessarily the case that every agent will want to observe the information bestowed upon him by S' if given a choice. To capture this idea, I will define a modified information structure $\hat{S}^j$ which is identical to S^* for all agents except agent j, for whom it is identical to S.

Formally, define $\hat{T}^j = (T_1^*, ..., T_j, ..., T_I^*)$ and $\hat{\pi}^j : \Theta \to \Delta(T_1^*, ..., T_j, ..., T_I^*)$ by

$$\hat{\pi}^j((t_1, t_1'), ..., t_j, ...(t_I, t_I')|\theta) = \sum_{t_j' \in T_j'} \pi^*((t_1, t_1'), ..., (t_j, t_j'), ...(t_I, t_I')|\theta)$$

for all $((t_1, t_1'), ..., t_j, ...(t_I, t_I')) \in \hat{T}^j$ and all $\theta \in \Theta$.

Say that S^* satisfies the *hear-no-evil condition* for player j with respect to (G, S) if there is some Bayes Nash equilibrium of (G, S^*) that player j weakly prefers to some Bayes Nash equilibrium of $(G, \hat{S}^j)$. I take this as a necessary condition for player j to not *unilaterally* commit to rejecting his signal, if he knew no other player would do so. There are other reasonable conditions that can be posed - for example, requiring that j to prefer his least favorite equilibrium of (G, S^*) to his least favorite equilibrium of $(G, \hat{S}^j)$ - but for now I will focus on this weakest possible condition.

A decision rule σ constitutes a *hear-no-evil Bayes correlated equilibrium* of (G, S) if there is an expansion S^* of S such that (i) a BNE of (G, S^*) induces σ and (ii) S^* satisfies the hear-no-evil condition with respect to (G, S) for all players. Let $HNE(G, S)$ denote the set of hear-no-evil Bayes correlated equilibria of the incomplete information game (G, S). Note that since S vacuously satisfies the hear-no-evil condition with respect to itself for all players and games, all Bayes Nash equilibria of (G, S) are also hear-no-evil Bayes correlated equilibria of (G, S)

2.3 Result

As shown by Bergemann and Morris (2016), for two information structures S and S^*, S^* is individually sufficient for S if and only if $BCE(G, S^*) \subseteq BCE(G, S)$ for all basic games G. Here I pose a similar question about the set of hear-no-evil equilibria. Specifically, for a fixed game G, I show that $HNE(G, S^*) \subseteq HNE(G, S)$ for all S^* individually sufficient for S if and only if the hear-no-evil condition has no bite. The proof relies on S^* vacuously satisfying the hear-no-evil condition with regards to itself. Thus if we assume that $HNE(G, S^*) \subseteq HNE(G, S)$ for all S^* individually sufficient for S, then any expansion of S must satisfy the hear-no-evil condition for each player.

Theorem 2.1 *For a fixed game G and information structure S, $HNE(G, S^*) \subseteq HNE(G, S)$ for all $S^* \succ S$ if and only if $HNE(G, S) = BCE(G, S)$.*

Proof ($\rightarrow$) Suppose $HNE(G, S^*) \subseteq HNE(G, S)$ for all S^* with $S^* \succ S$. Since S^* satisfies the HNE condition with respect to itself for all agents, the equilibrium associated with S^* is trivially found in $HNE(G, S^*)$. Then this equilibrium must also be in $HNE(G, S)$, implying that S^* satisfies the HNE condition for all i with respect to S. Thus for all $S^* \succ S$, S^* satisfies the HNE condition with respect to S for all i, implying the HNE condition has no bite for any expansion of S. Thus we have $HNE(G, S) = BCE(G, S)$.

($\leftarrow$) Assume that $HNE(G, S) = BCE(G, S)$. From Theorem 2 of Bergemann and Morris (2016), we know that $BCE(G, S^*) \subseteq BCE(G, S)$ for any $S^* \succ S$. Since $HNE(G, S^*) \subseteq BCE(G, S^*)$, we must have $HNE(G, S^*) \subseteq HNE(G, S)$ as well.

Note that this holds for alternative definitions of the hear-no-evil condition that have S satisfying the hear-no-evil condition with respect to itself for all S. For example, it would still apply if the hear-no-evil condition required each player to weakly prefer all the equilibria of (G, S^*) to all the equilibria of (G, S). However, it would not hold if we assume the hear-no-evil condition requires agents to *strictly* prefer some equilibria of (G, S^*) to some equilibria of (G, S).

3 Example

The following example is a special case of the one-sided incomplete information contest examined in Denter, Morgan, and Sisak (2014)[6], which also appears in Zhang and Zhou (2014)[12].

There is a contest between two parties A and B, who simultaneously choose how much effort to exert. The cost to party i of exerting effort x_i is normalized to x_i. The probability that party i wins the contest when effort levels are (x_A, x_B) is given by $p_i(x_A, x_B) = \frac{x_i}{x_A + x_B}$. The value to party i of winning the contest is given by v_i. Both players are expected utility maximizers.

Suppose further that $v_A = 10$, and v_B is randomly distributed with $Pr(v_B = 5) = .5 = Pr(v_B = 15)$. Suppose B knows the value of v_B, but A does not. As shown by Denter, Morgan, and Sisak (2014) in the general case, and which I will demonstrate in the specific case below, A will strictly prefer to remain uninformed of the value of v_B, assuming that B is aware of whether or not A knows v_B. One can imagine B posting the value of v_B to his LinkedIn profile, which will notify him of who has viewed it.

To make the example simpler, I will assume that A only has three effort levels to choose from: 2.222, which is optimal when A is informed that $v_B = 5$, 2.295, which is optimal when A is not informed of v_B, and 2.4, which is optimal when A is informed that $v_B = 15$. Similarly, I will assume that B only has four effort levels to choose from: 1.093, which is optimal when $v_B = 5$ but A is not informed of v_B, 1.111, which is optimal when A is informed that $v_B = 5$, 3.573, which is optimal when $v_B = 15$ but A is not informed of v_B, and 3.6, which is optimal when A is informed that $v_B = 15$.[4]

Payoffs are described in the following matrices, where agent i's payoff is given by $v_i(\frac{x_i}{x_A+x_B}) - x_i$. Decimals are rounded to three places except where further digits are needed to compare values.

When $v_B = 5$:

	1.093	1.111	3.573	3.6
2.222	4.481, .5555671	4.444, .5556	1.612, -.490	1.595, -.508
2.295	4.479, .5200	4.443, .51994	1.616, -.529	1.598, -.547
2.4	4.471, .472	4.436, .471	1.618, -.582	1.6, -.6

When $v_B = 15$:

	1.093	1.111	3.573	3.6
2.222	4.481, 3.853	4.444, 3.889	1.612, 5.6755	1.595, 5.6752
2.295	4.479, 3.746	4.443, 3.782	1.616, 5.5604	1.598, 5.5603
2.4	4.471, 3.601	4.436, 3.636	1.618, 5.39987795	1.6, 5.4

These payoff matrices define the basic game G, and I will endow it with the information structure $S = (T_A \times T_B, \pi)$, where $T_A = \{t_A\}$, $T_B = \{t_{B5}, t_{B15}\}$, and $\pi(t_A, t_{b5}|v_B = 5) = 1$, $\pi(t_A, t_{B15}|v_B = 15) = 1$. This information structure conveys no information about the state to A, but fully reveals the state to B. Next, define information structure $S^* = (T_A^* \times T_B^*, \pi^*)$, where $T_A^* = \{t_{B5}, t_{A15}\}, T_B^* = T_B$, and $\pi^*(t_{A5}, t_{B5}|v_B = 5) = 1$, $\pi^*(t_{A15}, t_{B15}|v_B = 15) = 1$. This information structure fully reveals the state to both A and B. I claim that S^* fails the hear-no-evil condition with respect to (G, S) for player A. To prove this, I will show that A strictly prefers the unique equilibrium of (G, S) to the unique equilibrium of (G, S^*).

It is easily verifiable that in game (G, S^*), when both A and B are informed of v_B, when $v_B = 5$, (2.222,1.111) is the unique pure-strategy equilibrium, and when $v_B = 15$, (2.4,3.6) is the unique pure-strategy equilibrium. In the appendix I show that (G, S^*) has no mixed-strategy equilibria. The expected payoff to A in this case is .5(4.444 + 1.6)=3.022.

Since each player's action here is fixed by the state of the world, we can describe this outcome with the decision rule $\sigma^*(\cdot|\theta)$, where $\sigma^*(2.222, 1.111|v_B = 5) = 1$, and $\sigma^*(2.4, 3.6|v_B = 15) = 1$. Since S^* is an expansion of S, and this outcome is a Bayes Nash equilibrium of S^*, we see that σ^* is a Bayes correlated equilibrium of our original game (G, S). Also note that σ^* can only be induced by an information structure equivalent to S^*: any information structure which does not fully inform A of v_B cannot result in a decision rule where A's action is fixed by v_B.

To determine whether σ^* is a hear-no-evil Bayes correlated equilibrium of (G, S), we must examine what happens when A commits to ignoring the additional information conveyed by S^*. When A rejects the additional information contained in S^*, the resulting information structure $\hat{S}^A$ is identical to S, since S^* conveyed no additional information to B. Thus in order to be a hear-no-evil Bayes correlated equilibrium, A must weakly prefer playing according σ^* to

[4]The optimal effort levels when A is uninformed were calculated using Proposition 1 in Zhang and Zhou (2014), and the optimal effort levels when A is informed were calculated using the formulas in the appendix of Denter, Morgan, and Sisak (2014).

playing some Bayesian Nash equilibrium of (G, S).

In game (G, S), when A is not informed of v_B, but B is, there exists a pure-strategy equilibrium where $x_A = 2.295$, $x_B(5) = 1.093$, and $x_B(15) = 3.573$. To see this, first observe that B is best responding in both cases. Given this, choosing effort 2.222 yields A payoff .5(4.481+1.612)=3.0465, choosing effort 2.295 yields A expected payoff .5(4.479 + 1.616)= 3.0475, and choosing effort 2.4 yields A expected payoff .5(4.471+1.618)=3.0445, hence 2.295 is a best response for A. In the appendix I show I that there are no other equilibria of (G, S). Note that this equilibrium gives A an expected payoff of 3.0475, which is higher than the 3.022 A would receive under complete information.

Thus, A strictly prefers the unique equilibrium of (G, S) to the unique equilibrium of (G, S^*). Since the decision rule σ^* can only be induced by an information structure equivalent to S^*, and S^* fails the hear-no-evil condition with respect to (G, S) for player A, we can conclude that σ^* is not a hear-no-evil Bayes correlated equilibrium of (G, S).

4 Agent Welfare

The following example shows that a BCE which fails the hear-no-evil condition for each player may nonetheless Pareto-dominate a HNEBCE.

There are two players. The state of the world ω is either 0 or 1, and it is common knowledge that $Pr(\omega = 0) = .1$. The row player must choose either 0 or 1, and the column player must choose either L or R. Payoff matrices for each state are given below. I will refer to the row player's action as a_1 and the column player's action as a_2.

0:

	L	R
0	-19, 1000	1, -12
1	-20, -1	0, 0

1:

	L	R
0	-20, 1000	0, -12
1	-19, -1	1, 0

The row player prefers to choose action 1 whenever $Pr(\omega = 1|\cdot) > .5$. The column player prefers to choose action R when $a_1 = 1$, and action L when $a_1 = 0$.

It is a BCE for the row player to always match his action to the state of the world, and the column player to choose action L in state 0, and action R in state 1. This gives the row player expected payoff .1*-19 + .9*1 = -1, and the column player expected payoff .1*1000 + .9*0 =100

However it is not a HNEBCE, since if the row player could publicly reject his information, the unique equilibrium would be to always play (1,R), which earns him expected payoff .9.

Now consider the following twist. There are two copies of the game played simultaneously, one where player 1 plays the role of row player and player 2 plays the role of column player, and one where player 1 plays the role of column player and player 2 plays the role of row player. Each player's strategy must specify what they do in each game, so each player has four possible strategies: 0L, 0R, 1L, 1R.

Again, it is a BCE for the row player player to always match his action to the state and the column player player to choose L in state 0 and R in state 1. Since each player gets to play each role once, this earns each player expected payoff -1 + 100 = 99.

However, this is again not a HNEBCE. If one player unilaterally rejected their information, then in the game where they are the row player, the unique equilibrium will be (1,R), earning utility .9, and in the game where they are the column player, they can simply choose action L regardless of state (since they will no longer be informed of the state), earning payoff .1*1000 + .9*-1=99.1. This gives them expected payoff 100 between the two games, so they are better off than if they had not rejected the mediator's information.

Importantly, there is no HNEBCE of this combined game that can generate expected payoffs of (99,99) or higher. The only way for a player to earn a payoff of 99 is for the outcome (0,L) to occur with probability at least .09 in the game where they are the column player (since no outcome can generate more than utility 1 when they are the row player). But if the row player knows the column player will play L with probability .09, his utility from that game can be no higher than .91*1 + .09 * -19 = -.8. Rejecting his information will ensure outcome (1,R) is always played when he is the row player, earning payoff .9 for that game. This is an increase of 1.7. However, rejecting his information will also lower his payoff in the game where he is the column player. When he has no information as the column player, he is best off always choosing action L (instead of only choosing L when his opponent would choose 0). But this lowers his utility by at most 1. Thus rejecting his information gives him more utility (1.7) in the game where he is a row player than it loses him (1) in the game where he is the column player. Hence rejecting his information is strictly preferred. Thus any BCE where outcome (0,L) arises with probability at least .09 fails the HNE condition, and thus no HNEBCE can give either player a payoff of 99 or higher.

This has an interesting implication. If a mediator seeks to maximize any increasing function of agent welfare, his preferred outcome fails the hear-no-evil condition. If some regulator were to restrict the mediator to only inducing HNEBCE outcomes, both players would be made worse off. This suggests that even if a piece of information may be harmful to the agent possessing it, censoring that information may not be socially beneficial.

5 Coordination games

Consider the following parameterization of a coordination game, as examined in section IV of Taneva (2019). The state of the world is either 0 or 1, and each is equally likely. When the state is ω the payoffs are as given in G_ω, where $c \geqslant 0$ and $d \geqslant 0$.

G_0 :

	0	1
0	c, c	d, 0
1	0, d	0, 0

G_1 :

	0	1
0	0, 0	0, d
1	d, 0	c, c

Players care about coordinating their actions with the state of the world and with each other. Each player prefers to correctly match their action to the state of the world, regardless of the action of the other player. This parameterization captures a range of coordination games. When $c > d$, agents prefer to be correct together, as in an investment game. When $d > c$, agents prefer to be correct alone; for example, if agents must decide whether to plant a dry crop or a wet crop, where being correct alone guarantees an agent monopoly profits, while being correct together guarantees them duopoly profits. When $c = d$, agents care only about matching the state, and not about the other agent's action. For a more in-depth analysis of the coordination game and its Bayes correlated equilibria, see Taneva (2019).

Taneva (2019) assumes the information designer treats the two agents as indistinguishable, and restricts attention to such "symmetric" BCE. Such symmetric equilibria can be described by two parameters: the probability r of both agents matching their actions to the state, and the probability q of a given agent matching his action to the state. For example, the outcome in which both agents always match their actions to the state is described by $(q, r) = (1, 1)$, and the outcome where exactly one agent always matches his action to the state is described by $(q, r) = (.5, 0)$.

Notably, Taneva (2019) only considers BCE of the basic game, where neither player has any information beyond the mediator's signal. Maintaining (i) the restriction to symmetric equilibria, and (ii) the assumption that neither player has any baseline information, I will show that every remaining BCE is a HNEBCE.

Below I identify one equilibrium that will always exist when one player rejects their information.

Lemma 5.1 *If one player rejects his information, then there is a BNE of the game that arises where: (i) he randomize between his actions with equal probability, and (ii) his opponent chooses the action she believes is more likely to match the state.*

Proof If one player does not know anything about the state, and chooses to randomize uniformly between his two actions, his opponent can do no better than to choose the action she believes is more likely to match the state. If her strategy is to try to match the state, the first player can do no better than randomizing uniformly between his two actions, since he is indifferent between them. ∎

Recall that for an information structure to satisfy the hear-no-evil condition, we require only that each player prefers his favorite equilibrium that might arise after hearing the mediator's information to his least-favorite equilibrium that might arise after unilaterally rejecting his information. That is, information structure S^* will satisfy this condition if each player i prefers any BNE of

(G, S^*) to any BNE of $(G, \hat{S}^i)$, $i = 1, 2$, where again $\hat{S}^i$ is the information structure where player i chooses not to hear the mediator's recommendation, and the other player does choose to hear it. To prove that a given BCE is a HNEBCE, the general approach I will take here is to show that there is some BNE of (G, S^*) in which player i's opponent plays the same strategy as in some BNE of $(G, \hat{S}^i)$. Since his opponent's strategy does not change between these two BNE, the BNE of (G, S^*) cannot be worse for player i than the corresponding BNE of $(G, \hat{S}^i)$ (since in the BNE of (G, S^*), player i has the option to deviate to the strategy he plays in $(G, \hat{S}^i)$).

Lemma 5.2 *All symmetric BCE with $q > .5$ are HNEBCE.*

Proof Suppose $q > .5$. Then when P2 follows the mediator's recommendation, she is taking the action she believes is more likely. Suppose P1 rejects the mediator's signal (and so does not hear the mediator's recommendation). In the BNE described by the previous lemma, P2 will be taking the action she believes is more likely. Thus P1's choice to reject his information does not change his opponent's action, and so rejecting his information makes him worse off. $\blacksquare$

When $c > d$, any signal from the mediator that induces a BCE with $q < .5$ can also induce a BCE with $q > .5$. This is because, for whatever strategies are associated with the BCE with $q < .5$, each agent can also play the strategy which takes the opposite action, and if both agents take this new strategy, neither will want to deviate. I formalize this intuition in the following lemma.

Lemma 5.3 *If $c > d$, all symmetric BCE are HNEBCE.*

Proof If $q > .5$, we are done, so suppose $q < .5$ for some symmetric BCE (q, r). Then by definition there is some information structure S such that action distribution (q, r) is induced a BNE of (G, S). Let s_1 and s_2 denote the strategies played by player 1 and player 2 respectively in this BNE of (G, S). Consider the strategies s'_1 and s'_2, defined such that player i takes the opposite action under strategy s'_i as they do under strategy s_i. Then s'_1, s'_2 is also a BNE of (G, S): if either player unilaterally deviates, they will be less likely to match their action to the state, and less likely to match their action to the other player's. Since $c > d$, such a unilateral deviation is never profitable. The associated symmetric BCE can be written as $(q', r') = (1 - q, 2q - r)$. Notably, $q' > .5$, and so by the previous lemma, information structure S satisfies the hear-no-evil condition with respect to no information for the game G.

Note that the proof of this lemma hinges on the weakness of the HNE condition. Currently it requires only that p1 prefers the p1-best equilibrium that could arise after hearing the mediator's information to the p1-worst equilibrium that could arise after rejecting the mediator's information. However, it may also be reasonable to compare the p1-worst equilibria in either case.

As an example, $c = 100, d = 1$, and $q = r = .1$ (this is a BCE because neither player wants to unilaterally deviate from following the mediator's recommendation). If player 1 expects that the other player will follow the mediator's recommendation, he could unilaterally reject his information and be strictly better off regardless of which subsequent BNE is played.

In this case the public rejection of information can also serve as a way to avoid bad equilibria, which is not captured by the current "weak" hear-no-evil condition. Under the current definition, players are maximally optimistic about what will happen after seeing the mediator's signal, and maximally pessimistic about what will happen without it.

Next I move to the case where $c \leqslant d$.

Lemma 5.4 *If $c \leqslant d$, all symmetric BCE are HNEBCE.*

Proof If $q > .5$ we are done, so assume that $q < .5$ for some symmetric BCE (q, r). Then for some information structure S^* and strategies s_1, s_2, the strategy pair (s_1, s_2) supports the BNE (q, r) of the game (G, S^*). Consider the strategy pair (s'_1, s'_2), where s'_i is defined as the strategy that takes the opposite action of strategy s_i for each realization of S^*, for $i = 1, 2$. I claim that neither player has an incentive to deviate from s'_1, s'_2, since neither player had an incentive to deviate from s_1, s_2. Deviating from s'_1, s'_2 would make an agent strictly less likely to match their action to the state, and no more likely to match the state alone, than if they deviated from s_1, s_2. Thus deviating from s_1, s_2 is more appealing to each agent than deviating from s'_1, s'_2. Since s_1, s_2 was a BNE, then s'_1, s'_2 must also be a BNE, which I will also refer to as the symmetric BCE (q', r'), where $q' = 1 - q$ and $r' = 1 - (2q - r)$. Note that player 1 has the option of responding to strategy s'_2 by randomizing uniformly over his actions, but chooses s'_1 as a best response instead. Thus player 1 prefers this equilibrium to the equilibrium when player 1 rejects his information where he randomizes uniformly over his actions while player 2 plays strategy s'_2. ∎

The following proposition summarizes the results of these lemmas.

Proposition 5.5 *Any symmetric BCE of the base game is an HNEBCE, for any $c \geqslant 0, d \geqslant 0$.*

6 Conclusion

In this paper I have laid out a refinement of Bayes correlated equilibrium which requires that players be willing to hear the mediator's recommendation, in addition to being willing to follow the recommendation once they hear it. While the original concept of Bayes correlated equilibrium is useful for making predictions when information acquisition is *covert*, so that no agent knows the information structure acquired by another agent nor its realization, my "hear-no-evil Bayes correlated equilibrium" serves the same purpose when information acquisition is *overt*, and agents are aware of the information structures acquired by other agents but not their realizations. This is relevant in situations where it can be hard to hide what one is researching - for example, if a firm must build a new facility to conduct a certain type of R&D, or if a researcher must apply for a grant before beginning work.

In contrast to the set of Bayes correlated equilibra, we have seen that the set of hear-no-evil equilibria need not be decreasing in the players' baseline level of information. Instead, for any game where that is true for the set of hear-no-evil equilibria, the set of hear-no-evil equilibria must be equal to the set of Bayes correlated equilibria. I have illustrated the equilibrium concept with an example of a contest where a player strictly prefers not to learn his opponent's value of winning. While it may seem intuitive that a given HNEBCE should be better for agent welfare than a BCE which fails the hear-no-evil condition, I have also provided a further example to show this is not the case, where a Pareto-optimal outcome is a BCE but not a HNEBCE. Finally, we have seen that for a broad class of 2x2x2 coordination games, every symmetric BCE satisfies the hear-no-evil condition.

7 Appendix

To see that there are no other pure-strategy equilibria of (G, S), note that if A plays 2.222 with probability 1 then when $v_B = 5$ B will play 1.111 and when $v_B = 15$ B will play 3.573. But in this case A would prefer to deviate to 2.295. And if A plays 2.4 with probability 1, then when $v_B = 5$ B will play 1.093 and when $v_B = 15$ B will play 3.6. But then A would again prefer to deviate to 2.295. Thus there are no other pure-strategy equilibria when both parties have complete information.

It remains to rule out mixed-strategy equilibria. Let $\sigma_i(x_i)$ denote the probability of agent i choosing effort level x_i. First consider the case of full information.

When $v_B = 5$, note that 3.573 and 3.6 are dominated strategies for B, and will not be played in any equilibrium. Given that B will not be playing these strategies, A will choose not to play 2.4 or 2.295, since both will be dominated by 2.222. Given this, B will strictly prefer to play 1.111, and there are no mixed equilibria.

When $v_B = 15$, B will not play 1.093 or 1.111, as both are dominated strategies. Given this, A will not play 2.222 or 2.295, as both are dominated by 2.4. The strict best response for B in this case is 3.6, so there are no mixed equilibria.

Hence there are no other equilibria of (G, S^*) besides the pure-strategy equilibrium outlined above.

Finally, consider the game (G, S) when B knows v_B but A is uninformed.

When $v_B = 5$, B will play either 1.093 or 1.111. In order for B to be willing to randomize between them, we must have $.55555671\sigma_A(2.222) + .52\sigma_A(2.295) + .472\sigma_A(2.4) = .5556\sigma_A(2.222) + .51994\sigma_A(2.295) + .471\sigma_A(2.4)$.

When $v_B = 15$, B will play either 3.573 or 3.6. In order for B to be willing to randomize between them, we must have $5.6755\sigma_A(2.222) + 5.5604\sigma_A(2.295) + 5.39987795\sigma_A(2.4) = 5.56752\sigma_A(2.222) + 5.5603\sigma_A(2.295) + 5.4\sigma_A(2.4)$.

In order for A to be willing to randomize between 2.222 and 2.295, we must have $4.481\sigma_B(1.093) + 4.444\sigma_B(1.111) + 1.612\sigma_B(3.573) + 1.595\sigma_B(3.6) = 4.479\sigma_B(1.093) + 4.443\sigma_B(1.111) + 1.616\sigma_B(3.573) + 1.598\sigma_B(3.6)$.

In order for A to be willing to randomize between 2.295 and 2.4, we must have $4.479\sigma_B(1.093)+4.443\sigma_B(1.111)+1.616\sigma_B(3.573)+1.598\sigma_B(3.6) = 4.471\sigma_B(1.093)+4.436\sigma_B(1.111) + 1.618\sigma_B(3.573) + 1.6\sigma_B(3.6)$.

In order for A to be willing to randomize between 2.222 and 2.4, we must have $4.481\sigma_B(1.093)+4.444\sigma_B(1.111)+1.612\sigma_B(3.573)+1.595\sigma_B(3.6) = 4.471\sigma_B(1.093)+4.436\sigma_B(1.111) + 1.618\sigma_B(3.573) + 1.6\sigma_B(3.6)$.

Putting these three equations in matrix form, obtain:

$$\begin{bmatrix} 4.481 & 4.444 & 1.612 & 1.595 \\ 4.479 & 4.443 & 1.616 & 1.598 \\ 4.481 & 4.444 & 1.612 & 1.595 \end{bmatrix} \begin{bmatrix} \sigma_B(1.093) \\ \sigma_B(1.111) \\ \sigma_B(3.573) \\ \sigma_B(3.6) \end{bmatrix} = \begin{bmatrix} 4.479 & 4.443 & 1.616 & 1.598 \\ 4.471 & 4.436 & 1.618 & 1.6 \\ 4.471 & 4.436 & 1.618 & 1.6 \end{bmatrix} \begin{bmatrix} \sigma_B(1.093) \\ \sigma_B(1.111) \\ \sigma_B(3.573) \\ \sigma_B(3.6) \end{bmatrix}$$

which simplifies to

$$\begin{bmatrix} .002 & .001 & -.004 & -.003 \\ .008 & .007 & -.002 & -.002 \\ .01 & .008 & -.006 & -.005 \end{bmatrix} \begin{bmatrix} \sigma_B(1.093) \\ \sigma_B(1.111) \\ \sigma_B(3.573) \\ \sigma_B(3.6) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Row reducing this yields

$$\begin{bmatrix} 1 & 0 & -4.333 & -3.167 \\ 0 & 1 & 4.667 & 3.333 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_B(1.093) \\ \sigma_B(1.111) \\ \sigma_B(3.573) \\ \sigma_B(3.6) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which has no solution with $\sigma_B(\cdot) \in [0, 1]$. Thus there is no equilibrium in which A randomizes over all three effort levels.

Next we must check if there is an equilibrium in which A randomizes between two effort levels. If A randomizes between 2.222 and 2.295, then effort level 3.6 will be strictly dominated for B. When $v_B = 15$ B will always choose effort level 3.573, and when $v_B = 5$ B will choose either 1.093 or 1.111. However, if $\sigma_B(3.573) = .5$, then regardless of the values of $\sigma_B(1.093)$ and $\sigma(1.111)$, A will prefer to only choose effort level 2.295 and not randomize.

If A randomizes between 2.295 and 2.4, then when $v_B = 5$ B will always choose effort level 1.093, and when $v_B = 15$, B will choose either effort level 3.357 or 3.6. However, if $\sigma_B(1.093) = .5$, then regardless of $\sigma_B(3.357)$ or $\sigma_B(3.6)$, A will prefer to choose effort level 2.295 and not randomize.

If A randomizes between 2.222 and 2.4, then when $v_B = 5$ B will choose either 1.093 or 1.111, and when $v_B = 15$ B will choose either 3.357 or 3.6. In order for A to be willing to randomize, it must be true that $4.481\sigma_B(1.093) + 4.444\sigma_B(1.111)+1.612\sigma_B(3.573)+1.595\sigma_B(3.6) = 4.471\sigma_B(1.093)+4.436\sigma_B(1.111)+ 1.618\sigma_B(3.573) + 1.6\sigma_B(3.6)$, or equivalently

$$.01\sigma_B(1.093) + .008\sigma_B(1.111) - .006\sigma_B(3.573) + -.005\sigma_B(3.6) = 0.$$

Since $\sigma_B(1.111) = .5 - \sigma_B(1.093)$ and $\sigma_B(3.6) = .5 - \sigma_B(3.573)$, we can rewrite this as

$$.01\sigma_B(1.093)+.004-.008\sigma_B(1.093)-.006\sigma_B(3.573)+-.0025+.005\sigma_B(3.573) = 0$$

and again as

$$.002\sigma_B(1.093) - .001\sigma_B(3.573) + .0015 = 0.$$

This simplifies to $2\sigma_B(1.093) + 1.5 = \sigma_B(3.573)$. Since this has no solution for $\sigma_B(\cdot) \in [0, 1]$, there is no equilibrium where A mixes between 2.222 and 2.295.

Hence there are no other equilibria of (G, S) besides the unique pure-strategy equilibrium outlined.

References

[1] George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970.

[2] Bruno Bassan, Olivier Gossner, Marco Scarsini, and Shmuel Zamir. Positive value of information in games. *International Journal of Game Theory*, 32(1):17–31, 2003.

[3] Bruno Bassan, Marco Scarsini, and Shmuel Zamir. "I don't want to know!": Can it be rational? *Discussion paper*, 1997.

[4] David Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability" University of California Press Berkeley, Calif*, pages 93–102, 1951.

[5] David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

[6] Philipp Denter, John Morgan, and Dana Sisak. 'where ignorance is bliss,'tis folly to be wise': Transparency in contests. *Available at SSRN 1836905*, 2014.

[7] Jack Hirshleifer. The private and social value of information and the reward to inventive activity. *The American Economic Review*, 61(4):561–574, 1971.

[8] Abraham Neyman. The positive value of information. *Games and Economic Behavior*, 3(3):350–355, 1991.

[9] Ludovic Renou. Commitment games. *Games and Economic Behavior*, 66(1):488–505, 2009.

[10] Anne-Katrin Roesler and Balázs Szentes. Buyer-optimal learning and monopoly pricing. *American Economic Review*, 107(7):2072–80, 2017.

[11] Ina Taneva. Information design. *American Economic Journal: Microeconomics*, 11(4):151–85, 2019.

[12] Jun Zhang and Junjie Zhou. Information disclosure in contests: A bayesian persuasion approach. *The Economic Journal*, 126(597):2197–2217, 2016.