Selected Facts

Nemanja Antic[*] Archishman Chakraborty[†]

This version: October, 2024.

Abstract

We study constrained information design where a sender must provide facts to persuade a receiver to accept or reject a proposal. We show that sender-optimal strategies correspond to maximal-weight matchings on a bipartite graph that incorporates a novel fact-selection constraint, alongside the usual ones. We characterize exactly when the sender can induce his ideal decisions. Receiver payoffs are independent of the sender's cardinal preferences, although these preferences determine the sender-optimal strategy. When the receiver can first specify the set of admissible facts, we identify conditions under which the receiver would (not) like to eliminate the sender's freedom to select facts.

JEL Classification: C72, D71, D72, D82.

Keywords: persuasion, matching, disclosure, evidence, cheap talk.

[*]Northwestern University, Evanston, Illinois; nemanja.antic@kellogg.northwestern.edu

[†]Yeshiva University, New York, New York; archishman@yu.edu.

1 Introduction

Experts often have to provide factual evidence in support of their arguments. But they also have freedom to select the facts they will emphasize. Politicians soliciting votes highlight the facts that are convenient for their case.[1] Television networks and newspaper editors give prominence to certain news stories while burying others that do not support their ideological loyalties. Managers reporting to the board of directors, or the investing public, have to submit evidence together with their projections and recommendations. But because of their expertise about internal operations, they may be able to cherry-pick this evidence. For the same reason, organizations may be able to manage the information they disclose to regulators, and experts testifying in court may be able to choose the facts they deem to be relevant.

In this paper, we consider the problem of an expert (the sender, he) who selects facts in order to persuade an uninformed observer (the receiver, she) to approve of a decision taken (or recommended) by the sender. The decision is to either accept or reject a proposal. The sender is privately informed about the state of the world. In each state he has access to some (indisputable) facts that are relevant for the decision. Relative to the receiver, the sender is biased in favor of accepting the proposal. The sender's report must (truthfully) reveal some facts that will subsequently be scrutinized by the receiver. His design problem is to select the facts that will be revealed, and those that will be concealed, as a function of his private information, while making sure the receiver always approves of his decision and never wants to overturn it and take a different decision.

To persuade the receiver, the sender must select her facts in a manner that pools states where she wants to accept the proposal but the receiver does not, with states where they both want to accept it. We show that this problem is formally equivalent to a matching problem on a suitably chosen bipartite graph that captures constraints on the facts that are available to the sender, while also accounting for the receiver's priors and preferences. One side of the graph corresponds to (agreement) states where the sender and receiver agree that the proposal should be accepted

[1] To justify the Iraq War in 2003, George W. Bush's administration claimed aluminum tubes bought by Iraq were for use in uranium-enriching centrifuges, suggesting Iraq had weapons of mass destruction. The suggestion turned out to be untrue (https://www.factcheck.org/2008/01/us-intelligence-on-wmds-in-iraq/). In the lead-up to the 2016 Brexit referendum, Boris Johnson and other pro-Brexit campaigners highlighted the UK's payments to the European Union, claiming that the UK sent £350 million a week to the EU. This fact was criticized to be misleading, in part because it excluded payments received by the UK from the EU. (https://www.independent.co.uk/news/uk/politics/vote-leave-brexit-lies-eu-pay-money-remain-poll-boris-johnson-a8603646.html).

whereas the other side corresponds to (conflict) states where they have a conflict of interest. Each edge of the graph corresponds to arguments the sender can make (i.e., facts she can reveal) that will persuade the receiver to follow the sender's recommended decision. We show that any matching on this graph can be used to construct a reporting strategy for the sender that satisfies Bayes plausibility and obedience constraints.

We apply Hall's marriage theorem (1935) to identify a sufficient condition under which the sender has a strategy that persuades the receiver to approve the sender's ideal decision in every state. This corresponds to a perfect matching on the bipartite graph described above. In effect, the sender *subverts* the receiver's attempt to monitor him and implements his own unconstrained optimal decision rule. The sufficient condition for subversion states that there must be enough diversity across states in the different ways the sender can select supporting facts to pool agreement states with conflict states. We provide examples of environments where the sufficient condition is met and the sender's optimal strategy takes a natural and simple form.

Our sufficient condition is also necessary. When it is not met, the sender must make a compromise and give up on her own ideal decision in some states. In general, the sender's optimal reporting strategy is the solution to a linear program. We show that it can be described as a maximal-weight matching on an edge-weighted bipartite graph, where the weights are derived from the sender's intensity of preferences (i.e., his cardinal utilities). Furthermore, the optimal reporting strategy also corresponds to a maximal-cardinality matching that does not depend on these weights.

Using this characterization, we identify the expected payoffs to both the sender and receiver from an optimal policy. We show that the receiver's expected payoff depends on the sender's ordinal preferences, but not on his cardinal preferences, even though the latter determine the sender's optimal reporting strategy. Thus, when the sender prefers to accept the proposal in every state, the receiver's expected payoff depends only on her own preferences and the priors.

We ask next if the receiver can set the rules of argumentation for the sender. Suppose that at an ex ante stage the receiver can specify a subset of the set of possible facts that are *admissible*. Thus, if the sender is a prosecutor and the receiver is a judge, the latter may choose to admit forensic evidence but not witness testimonies. The sender can only select facts that are deemed admissible by the receiver and he must have at least one such fact available in every state. What is the receiver-optimal set of admissible facts? Our graph-theoretic approach answers this question.

Consider first the special case of a *Cartesian* problem where the set of possible states is an n-dimensional product set that captures different aspects of the decision facing the two players. The available facts in each state are the realized values of each aspect and the sender can only provide $k < n$ of these values as supporting facts (choosing which ones to reveal). The receiver's ex ante design problem is to decide which values of each aspect to recognize as admissible facts.

In Cartesian problems, it is optimal for the receiver to pre-specify k admissible aspects. Any vector of values of these k admissible aspects counts as a fact, and nothing else does. This solution to the receiver's design problem minimizes the sender's freedom to select her facts in every state. The sender must reveal the realized values of the pre-specified admissible aspects. He has no choice in the matter. We can think of the admissible aspects as a *topic* of discussion that the sender is restricted to by the receiver. The receiver does not need to predict the arguments the sender will subsequently make (that are determined by the sender's cardinal preferences), in order to choose her optimal topic.

These results for Cartesian problems generalize to non-Cartesian problems (e.g., ones where the state space is not a product set), except that the receiver may not always want to eliminate the sender's freedom to select facts in *every* state. In general the receiver will want to minimize the diversity of arguments the sender can make *across* different states. We make this insight precise by showing that the receiver's problem reduces to maximizing a metric that we call the *Hall deficit*, derived from the necessary and sufficient condition for a perfect matching according to Hall's theorem. As in the Cartesian special case, the Hall deficit does not depend on the sender's cardinal preferences that determine the arguments the sender will make. It depends only on the receiver's own preferences and the priors.

The rest of the paper is organized as follows. Section 2 sets up the model. Section 3 casts the sender's problem as a matching problem on a suitable bipartite graph and provides necessary and sufficient conditions for the sender to subvert the receiver. Section 4 characterizes the optimal strategy for the sender more generally, while Section 5 describes the receiver-optimal specification of admissible facts. Section 6 discusses interpretations and variations of our model. Section 7 discusses the related literature and Section 8 contains the concluding remarks. The Appendix contains proofs of results not contained in the main text.

2 Model

A sender ("he") is privately informed about the state of the world $x \in X$, where X is a finite set of possible states. He prepares a report m recommending either that a proposal should be accepted ($d = a$) or it should be rejected ($d = r$). The report is scrutinized subsequently by an uninformed receiver ("she"). The receiver has a conflict of interest with the sender and may not agree with the sender's recommendation. The receiver attaches a prior probability $p(x) > 0$ to $x \in X$.

Both the receiver's and sender's payoff from rejecting the proposal is normalized to zero. If the proposal is accepted in state $x \in X$, the receiver's payoff is $u(x)$, while the sender's payoff is $v(x)$. Assume that if the receiver wants the proposal to be accepted then so does the sender: if $u(x) > 0$ then $v(x) > 0$. Let $A = \{x \in X : u(x), v(x) > 0\}$ be the set of states where the receiver and sender prefer to accept the proposal and let $R = \{x \in X : u(x), v(x) < 0\}$ denote the set of states where both parties prefer to reject the proposal. In the remaining states, $C = X \setminus (A \cup R)$, there is a conflict of interest between the sender and the receiver. The sender wants to accept the proposal when $x \in C$ while the receiver prefers to reject it.[2]

For every state x, there is a non-empty and finite set of associated facts $\mathcal{F}(x)$ that describes the available hard information (or evidence) at state x. The sender must choose one fact $\phi \in \mathcal{F}(x)$ from this set as part of her report. When $\phi \in \mathcal{F}(x)$ is revealed by the sender, the sender proves (and the receiver understands) that states outside the set $\{x|\phi \in \mathcal{F}(x)\}$ are impossible.[3] In each state x, the sender also recommends a decision $d \in \{a, r\}$ to either accept or reject the proposal. In addition, we allow her to include a (costless) message from a finite set $\mathbf{M}$ (that does not depend on x) in her report. Thus, in each state x, the sender's report $m \in \mathcal{M}(x) = \mathcal{F}(x) \times \{a, r\} \times \mathbf{M}$.[4]

The five-tuple $\mathcal{P} = \langle X, \{\mathcal{F}(\cdot)\}, p, u, v\rangle$ defines the sender's problem. Let $\mathcal{F} = \cup_{x \in X} \mathcal{F}(x)$ be the set of all possible facts, $\phi(m) = proj_{\mathcal{F}}(m)$ be the fact associated with report m and $d(m) = proj_{\{a,r\}}(m)$ be the decision recommended by report m. A reporting strategy $\sigma = \{\sigma(x)\}$ for the sender in a problem $\mathcal{P}$ is a collection of probability distributions over reports, $\sigma(x) \in \Delta(\mathcal{M}(x))$, one for each $x \in X$.

We say that the sender's problem $\mathcal{P}$ is *Cartesian* if it has two properties. First, $X = \times_{i=1,\ldots,n} X_i$, i.e., the set of possible states is a product set with each state $x = (x_1, ..., x_n) \in X$ defined by the realization of $n \geq 1$ decision-relevant *aspects*. Second, the sender must reveal the realized values of $k \geq 0$ of the n aspects, i.e., $\mathcal{F}(x) = \{\phi \subset \{x_1, ..., x_n\} : |\phi| = k\}$. The parameter k can be interpreted as a constraint on the sender's ability to produce evidence or it may reflect a constraint on the receiver's ability to process it. The case $k = 0$ corresponds to a standard unconstrained persuasion environment where the sender does not have to perfectly reveal any facts. In comparison, when $k > 0$ the sender is constrained by the requirement that he must reveal k components of the state.

Our model and results also cover environments beyond Cartesian ones. For instance, we allow

[2] We rule out the possibility of indifference (i.e., that either $u(x)$ or $v(x)$ equals zero), implying there is a unique full information optimal decision rule for both the sender and the receiver. Indifference by the receiver can be resolved in favor of the sender (and vice-versa), so this would not have a significant bearing on our results.

[3] The fact set $\mathcal{F}(x)$ can be redefined to accomodate the possibility of revealing multiple pieces of hard information, as we show below.

[4] Allowing the costless (cheap talk) message is a convenient way to handle randomizations in our proofs. But it has no bearing on our results. As will become clear, we can integrate over these messages and describe reporting strategies simply as a probability distribution over (recommended) decisions and revealed facts.

	WMD	*No WMD*
Saddam	$\mathcal{F}(x) = \{Saddam,\ WMD\}$	$\mathcal{F}(x) = \{Saddam\}$
Sam	$\mathcal{F}(x) = \{Sam,\ WMD\}$	$\mathcal{F}(x) = \{Sam\}$

Figure 1: Non-Cartesian example where the state space has a product structure

the set of possible states not to be a product set and a fact to contain the realized values of at least k of the aspects (and not exactly k). We also allow some values of some aspects not to count as hard evidence. In the example below, the state space is a product set described by the values of two aspects, the sender's name and whether or not he possesses weapons of mass destruction (WMDs).

As shown by the fact sets, the sender can always prove what his name is. He can also provide evidence he possesses WMDs. But he cannot prove that he does not possess WMDs. Thus, we allow the available hard evidence $\mathcal{F}(x)$ to not include all that the sender knows in state x. Our model also allows for a *grain of truth* specification of fact sets, $\mathcal{F}(x) = \{\phi \in 2^X : x \in \phi\}$, whereby an available fact in state x is any subset of the state space that contains x (see, e.g., Milgrom, 1981); or *Dye structures*, $\mathcal{F}(x) = \{x, \varnothing\}$, where any fact set consists of the actual state and another fact $\varnothing$ that is available in all states (see, e.g., Dye, 1985). We discuss the implications of our results for these instances of our model in detail below. For now we note that fact sets are primitives and we impose no restrictions on them except for requiring that each $\mathcal{F}(x)$ be non-empty.

The timing of moves for the constrained persuasion problem described above is as follows. First, the sender commits to a reporting policy σ. Then the state of the world x is realized and a report $m \in \mathcal{M}(x)$ is drawn according to $\sigma(x)$. After observing the report m, the receiver considers whether or not to obey the recommended decision, given her priors p and knowledge of σ. We focus on the reporting strategy σ that maximizes the sender's ex ante expected payoff.

Given a report m from the sender with recommendation $d(m) \in \{a, r\}$, the receiver is willing to obey the sender's recommendation if and only if

$$\mathbb{E}[u|m : d(m) = a] \geq 0 \geq \mathbb{E}[u|m : d(m) = r]. \tag{1}$$

This obedience constraint (Kamenica and Gentzkow, 2011) ensures the receiver has the incentive to follow the sender's recommendation. We can assume without loss that the sender must make prepare her report m in a way that ensures (1) holds. As an alternative specification that is formally equivalent, we can allow the sender to take the decision himself while making sure the receiver has no incentive to overturn his decision. Under this interpretation, (1) captures a notion of deniability

(Antic, Chakraborty and Harbaugh, 2024). When the sender has decision rights but must comply with regulations, or is subject to ex post public scrutiny, meeting this constraint allows him to maintain deniability that he has served the public interest. In what follows, we will refer to (1) as the deniability constraint.

3 Subversive strategies

We start our analysis by showing how the problem described above can be mapped into a matching problem on an appropriately defined graph. Since for any state $x \in R$, the sender and the receiver agree the proposal should be rejected, and the sender is able to commit to sending a report that recommends rejection and reveals any fact in $\mathcal{F}(x)$, the deniability constraint (1) will always holds in this case. So the sender gets his ideal outcome for all $x \in R$. The crux of the sender's problem is to get the receiver to agree to an accept decision as much as possible in the remaining states. This involves pooling states in the set A, where the sender and receiver agree that acceptance is best, with states in C where there is a conflict of interest.

Fix $\mathcal{P} = \langle X, \mathcal{F}, p, u, v \rangle$ and consider a bipartite graph G with vertices $V(G) = A \cup C$ and edges $E(G)$. The bipartition of the vertices $V(G)$ are the sets A and C. An edge $\{x, x'\} \in E(G)$ connects $x \in C$ and $x' \in A$ if and only if $\mathcal{F}(x) \cap \mathcal{F}(x') \neq \emptyset$. Thus, two vertices are connected by an edge as long as they have a fact in common. We will sometimes write $G = (A, C)$ to denote this bipartite graph.

Call $M \subseteq E$ a *matching* if $\{x, x'\} \in M$ implies $\{x, x''\} \notin M$ and $\{x', x''\} \notin M$ for all $x'' \in V(G)$. That is, a matching is a subset of edges such that each vertex is in at most one edge of the matching. Vertex $x \in V(G)$ is matched by M if $\{x, x'\} \in M$ for some $x' \in V(G)$; otherwise x is unmatched by M. A matching M on $G = (A, C)$ is *C-perfect* if every $x \in C$ is matched by M. A matching M on G is *perfect* if every $x \in V(G)$ is matched by M. For a subset of vertices $S \subseteq V(G)$, let $N(S) = \{x' \in V(G) \mid \{x, x'\} \in E,\ x \in S\}$ denote the *neighbors* of S.

Panel (a) of Figure 2 provides an example of a Cartesian problem $\mathcal{P}$, with $n = 2$, $k = 1$, $X_1 = \{0, 1, 2\}$, $X_2 = \{0, 1\}$ and uniform priors. In all our figures, the sets A, C and R are colored blue, yellow and red, respectively. Suppose that the receiver's utility is $u(x) = +1$ for $x \in A$, with $u(x) = -1$ otherwise. With such preferences, the receiver only wishes to avoid mistakes (i.e., choosing $d = r$ for $x \in A$ and $d = a$ for $x \notin A$) and she cares equally about both kinds of mistakes in every state. The deniability constraint (1) then becomes

$$\Pr[A \mid m : d(m) = a] \geq \frac{1}{2} \geq \Pr[A \mid m : d(m) = r]. \tag{2}$$

While we allow for arbitrary receiver utility functions u, this balance of probabilities special case of

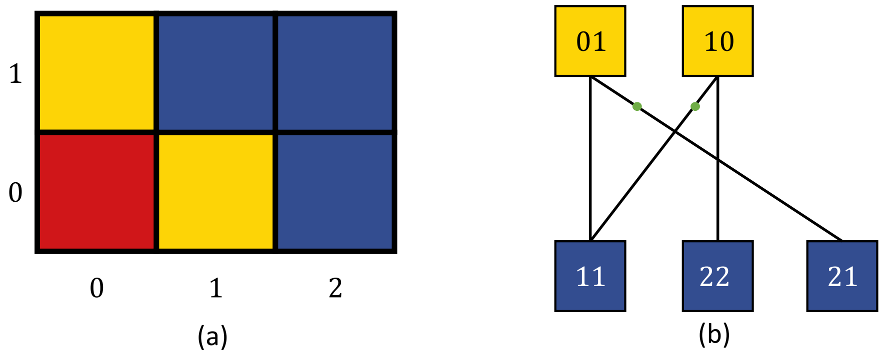

Figure 2: A state space and its graph

the deniability constraint, (2), will be important in what follows.[5] We allow for any sender's utility function v, as long as $v(x) > 0$ if $x \in A \cup C$ and $v(x) < 0$ otherwise.

Figure 2(b) shows the graph G associated with the problem in panel (a). Each edge of G corresponds to $k = 1$ facts that are shared by the two vertices defining the edge. This graph has a C-perfect matching, with the edges included in the matching highlighted (by green circles) in panel (b). We construct a reporting strategy for the sender from this matching as follows.

The sender chooses $d = a$ and reveals $x_2 = 1$ when the state is either $(0, 1)$ or $(2, 1)$, and also chooses $d = a$ but reveals $x_1 = 1$ when the state is either $(1, 0)$ or $(1, 1)$. In state $(2, 0)$ he again chooses $d = a$ while revealing $x_1 = 2$, while in state $(0, 0)$ he chooses $d = r$ and reveals $x_1 = 0$. Since priors are uniform, the relevant deniability constraint (2) is met in all cases. For instance, when the receiver observes $x_1 = 1$ and $d = a$, she concludes the state of the world is either $(1, 0) \in C$ or $(1, 1) \in A$. Since these two possibilities are equally likely, $\Pr[A \mid m : x_1 = 1, d(m) = a] = 1/2$ so that (2) is met.

Notice that the sender obtains his ideal decisions using such a reporting strategy since $d = a$ whenever $x \in A \cup C$ and $d = r$ whenever $x \in R$. Whenever there is a reporting strategy that makes it possible to implement the sender's ideal decision rule, while satisfying the deniability constraint, the sender avoids any burden of scrutiny by the receiver. He subverts the receiver's agenda and implements his own unconstrained optimal decisions. We call such a reporting strategy a *subversive (reporting) strategy*. In the rest of this section we identify necessary and sufficient conditions for the existence of subversive strategies, going beyond the special assumptions of the example in Figure 2.

The key tool for our analysis is Hall's matching theorem (1935) that takes as a primitive a

[5]It is similar to the "balance of probabilities" standard of proof used for U.S. civil law cases. When (2) is met, the balance of probabilities favors acquitting the sender.

bipartite graph $G = (A, C)$. It states that a C -perfect matching exists on G if and only if

$$|N(S)| \geq |S| \text{ for all } S \subseteq C. \tag{HC}$$

We may think of elements of C as agents and elements of A as objects. Each agent $x \in C$ only finds objects in $N(\{x\})$ acceptable to her. A C-perfect matching assigns one acceptable object in A to every agent in C, and the same object is not assigned to two different agents (i.e., the matching is one-to-one). Condition (HC) is a statement about the diversity of preferences among the agents for the objects. For instance, if two agents find one and the same object as the only acceptable one, (HC) is not met and it is impossible to assign a different acceptable object to each agent. Thus, (HC) is necessary for the existence of a C-perfect matching. Hall's theorem shows it is also sufficient.

If we assume uniform priors and the balance of probabilities version of the deniability constraint, then our version of the matching problem involves pooling, or matching, elements of the conflict set C with elements of the agreement set A. This is the key feature of the example in Figure 2. We assume that each conflict point in C can be pooled with an agreement point in A only via a "valid argument" (i.e., one that has a supporting fact). So (HC) is a statement about the diversity of possible arguments in our setting. Hall's theorem states that a subversive reporting strategy exists if and only if the requirement of providing supporting facts does not sufficiently restrict the possible arguments available to the sender. The sender can select his facts in a manner that allows his own ideal decisions to be implemented.[6]

In the rest of this section, we extend (HC) to a broader domain, allowing for general priors and receiver utility functions u, and going beyond Cartesian problems. We find necessary and sufficient conditions for the existence of subversive reporting strategies. In the subsequent section, we characterize the sender's optimal reporting strategies, covering cases where subversion is impossible. In all cases, the core intuition of our approach is contained in Hall's condition (HC).

To go beyond the simplifying assumptions underlying the example of Figure 2, to each $x \in A \cup C$ assign a weight $w(x) = \|p(x)u(x)\| > 0$. Unless mentioned otherwise, we will assume the weights $w(x)$ are rational for each $x \in A \cup C$. Let h be the largest number such that $w(x)/h$ is an integer for all $x \in A \cup C$ (i.e., $h = 1/L$, where L denotes the least common multiple of the denominators of $\{w(x)\}_{x \in A \cup C}$). For each $x \in A \cup C$ create $w(x)/h$ "clones" of x, denoted by ${}^{i}_{c}x$, $i = 1, ..., w(x)/h$.

[6] For the unconstrained Cartesian problem with $k = 0$, the sender does not need to provide any facts and any $x \in C$ can be matched with any $x' \in A$, i.e., G is a complete bipartite graph. In this case (HC) reduces to requiring $|C| \leq |A|$. At the other extreme where $k = n$, G is completely disconnected and $N(S)$ is empty for each x, so that (HC) fails and subversion is impossible.

Let ${}_cA$ be the set of all clones of all states $x \in A$ and ${}_cC$ be the set of all clones of all $x \in C$. Define the cloned bipartite graph ${}_cG = ({}_cA, {}_cC)$ derived from G as follows: $\{{}^i_c x, {}^j_c x'\} \in E({}_cG)$ if and only if $\{x, x'\} \in E$, for each $i = 1, ..., w(x)/h$, $j \in 1, ..., w(x')/h$. Notice that when $u(x) \in \{-1, +1\}$ and all $x \in A \cup C$ are equiprobable, ${}_cG$ coincides with G. We are ready for our first result.

Theorem 1 *A subversive reporting strategy exists iff*

$$\mathbb{E}[u|S \cup N(S)] \geq 0 \text{ for all non-empty } S \subseteq C. \tag{3}$$

Proof.

Step 1 ('if'**).** Suppose (HC) obtains on ${}_cG$ so that by Hall's theorem a C-perfect matching exists on ${}_cG$. Let ${}_cM$ be this matching. We first construct a subversive reporting strategy from ${}_cM$. Next we show that when (3) obtains on G, (HC) obtains on ${}_cG$.

Consider any $x \in C$, $x' \in A$ such that $\{x, x'\} \in E$ and $\{{}^i_c x, {}^j_c x'\} \in {}_cM$ for some $i = 1, ..., w(x)/h$, $j \in 1, ..., w(x')/h$. Let $\mu(\{{}^i_c x, {}^j_c x'\}) \in \mathbf{M}$ be a unique message that identifies the particular match $\{{}^i_c x, {}^j_c x'\} \in {}_cM$ (i.e., μ is a one-to-one map from ${}_cM$ to $\mathbf{M}$). The sender's report m consists of revealing a fact in $\mathcal{F}(x) \cap \mathcal{F}(x')$ that defines the edge $\{x, x'\}$, a decision $d(m) = a$, as well as the message $\mu(\{{}^i_c x, {}^j_c x'\})$. She sends report m with probability $h/w(x)$ when the realized state is x, and with probability $h/w(x')$ when the realized state is x'. Note that

$$\begin{aligned} \Pr[A \mid m : d(m) = a] &= \frac{\frac{h}{w(x')}p(x')}{\frac{h}{w(x')}p(x') + \frac{h}{w(x)}p(x)} = -\frac{u(x)}{u(x') - u(x)} \\ &= 1 - \Pr[C \mid m : d(m) = a] \end{aligned}$$

implying

$$\mathbb{E}[u|m : d(m) = a] = \Pr[A \mid m : d(m) = a]\, u(x') + \Pr[C \mid m : d(m) = a]\, u(x) = 0$$

so that the deniability constraint (1) is met.

Since ${}_cM$ is a C-perfect matching, for each $x \in C$ the the probabilities of such randomized reports add up to one. For $x' \in A$, these probabilities may add up to some number $q_{x'} < 1$ less than one, since some ${}^j_c x'$ may not be matched under ${}_cM$. In these cases, x' is perfectly revealed by the report with the remaining probability $1 - q_{x'}$, and the sender chooses $d(m) = a$. The sender also perfectly reveals all $x \in R$ and chooses $d(m) = r$. In both cases, (1) is met. Since $d(m) = a$ if and only if $x \in A \cup C$, this reporting strategy is subversive.

It remains to show (3) on G implies (HC) on ${}_cG$. Pick ${}_cS \subseteq {}_cC$. The claim is trivial if ${}_cS$ is empty so suppose ${}_cS \neq \emptyset$. Construct $S \subseteq C$ as follows: $x \in S$ if ${}^i_c x \in {}_cS$ for some $i = 1, ..., w(x)/h$.

Thus, S contains only those elements of C for which at least one clone is contained in ${}_cS$. Let ${}_cS' = \cup_{x\in S} \cup_i \left\{{}^i_c x\right\} \subseteq {}_cC$ be the set of all clones of all elements of S. Note that for any $x \in C$, $N(\{{}^i_c x\}) = N(\{{}^j_c x\})$ for every pair of clones ${}^i_c x$ and ${}^j_c x$ of x. Thus,

$$\begin{aligned} |N({}_cS)| &= \left|N({}_cS')\right| \\ &= \sum_{x'\in N(S)} \frac{w(x')}{h} \geq \sum_{x\in S} \frac{w(x)}{h} \\ &= \left|{}_cS'\right| \geq |{}_cS|, \end{aligned}$$

where the inequality in the second line follows from (3) applied to ${}_cS'$ and the inequality in the last line follows from noting ${}_cS \subseteq {}_cS'$.

Step 2 ('only if'**).**

Suppose there exists a subversive reporting strategy σ. Let $supp\ \sigma(x)$ be the support of $\sigma(x)$ and $supp_a\ \sigma(x) = \{m \in supp\ \sigma(x) : d(m) = a\}$. To satisfy (1), any $m \in supp_a\ \sigma(x)$ must also belong to $supp_a\ \sigma(x')$ for some $x' \in N(\{x\})$. Pick $S \subseteq C$ with $x \in S$. Define $\pi(x) \subseteq N(\{x\}) \subseteq A$ as follows: $x' \in \pi(x) \Leftrightarrow supp_a\ \sigma(x) \cap supp_a\ \sigma(x') \neq \emptyset$. Let $\pi(S) = \cup_{x\in S}\pi(x)$ and $supp_a\ \sigma(S) = \cup_{x\in S} supp_a\ \sigma(x)$.

Note that $\pi(S) \subseteq N(S)$ and, using the law of iterated expectations, further that

$$\begin{aligned} \mathbb{E}[u|S \cup N(S)] = \quad & \Pr[S \cup \pi(S) \mid S \cup N(S)]\mathbb{E}[u \mid S \cup \pi(S)] \\ & + \Pr[N(S) \smallsetminus \pi(S) \mid S \cup N(S)]\mathbb{E}[u|N(S) \smallsetminus \pi(S)] \end{aligned}$$

Since $u(x) > 0$ for all $x \in N(S) \smallsetminus \pi(S) \subseteq A$, this yields

$$\mathbb{E}[u|S \cup N(S)] \geq \mathbb{E}[u \mid S \cup \pi(S)] \Pr[S \cup \pi(S) \mid S \cup N(S)]$$

But, using the law of iterated expectations again,

$$\mathbb{E}[u|S \cup \pi(S)] = \sum_{m\in supp_a\ \sigma(S\cup\pi(S))} \Pr[m \mid S \cup \pi(S)]\mathbb{E}[u \mid m] \geq 0,$$

where $supp_a\ \sigma(S \cup \pi(S)) = \cup_{x\in S\cup\pi(S)} supp_a\ \sigma(x)$; and the inequality follows from the fact that σ is a subversive reporting strategy that meets (1) for all $m \in supp_a\ \sigma(S \cup \pi(S))$. This establishes (3) on G. ■

Each vertex of the cloned graph ${}_cG$, when viewed as the primitive, can be thought to be equiprobable, with probability equal to h. Furthermore, even when u takes a general form, because of how the cloned graph was constructed, it is as if the receiver uses the balance of probability standard (2) when evaluating a pooled message corresponding to an edge $\{{}^i_c x, {}^j_c x'\} \in {}_cM \subseteq E\left({}_cG\right)$. So Theorem 1 links Hall's condition (HC) when applied to ${}_cG$ to an equivalent condition (3) when

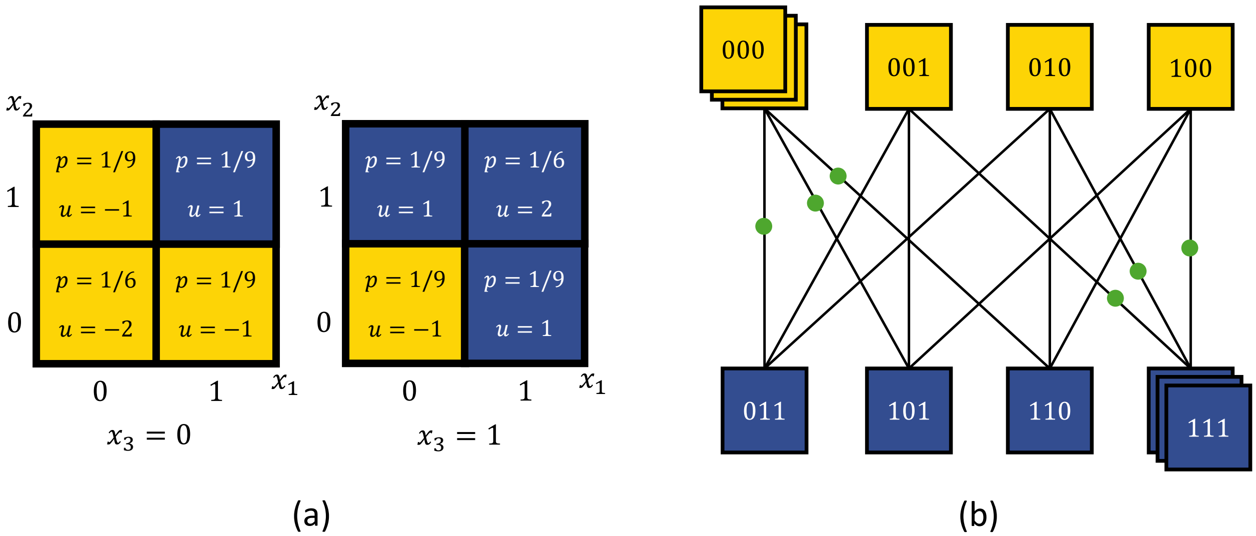

Figure 3: A subversive strategy as a matching

applied to the original graph G. It makes precise the sense in which assuming uniform priors and a balance of probabilities standard are without loss of generality, even if actual receiver preferences and priors are general.

Condition (3) (or (HC)) is easy to check in some instances of our model. For instance, with grain of truth fact sets $\mathcal{F}(x) = \{\phi \in 2^X, x \in \phi\}$, the set X of all possible states is always an available fact and so the graph G (and hence ${}_cG$) is complete. As a result, one needs to check (3) only for the case of the "grand coalition" $S = C$ with $N(C) = A$. Indeed, this is true in any setting where G is complete, e.g., a Cartesian problem with $k = 0$, or Dye fact sets $\mathcal{F}(x) = \{x, \varnothing\}$. These settings are like the usual unconstrained information design environments. In contrast, Figure 3 illustrates Theorem 1 in a constrained environment of a three-dimensional Cartesian problem with $k = 1$.

In Figure 3, the sender wishes to persuade the receiver to accept a proposal (e.g., pass a bill) in every state (i.e., R is empty). The sender knows, and can prove, whether the bill would help ($x_i = 1$) or hurt ($x_i = 0$) each of three citizens (or constituencies), $i \in \{1, 2, 3\}$. But he can provide evidence only for one citizen. The receiver prefers to pass the bill only if a majority of citizens benefit and her payoffs are shown in the panel (a) of the figure, together with priors p that display correlation across aspects. Notice that $w(x) = 1/3$ for $x \in \{(0,0,0), (1,1,1)\}$ and $w(x) = 1/9$ otherwise. Using this, in panel (b) we depict the cloned graph ${}_cG$ corresponding to this problem, as well as a perfect matching ${}_cM$ on ${}_cG$ (see marked edges).[7]

[7] To avoid clutter, we depict ${}_cG$ simply by replicating the vertices that have multiple clones but not depicting all

Using ${}_cM$ we can recover the sender's subversive reporting strategy, as follows. Since the sender always prefers to accept the proposal he will always recommend $d = a$. We only need to determine the fact he chooses to reveal in each state. When only one citizen benefits from the bill, $x \in \{(1,0,0),(0,1,0),(0,0,1)\}$, the sender reveals the identity of the citizen who benefits. Each such report is pooled with a report that is sent when $x = (1,1,1)$. In state $(1,1,1)$, the sender randomizes uniformly, revealing $x_i = 1$ with probability $1/3$ for each $i \in \{1,2,3\}$. Thus, when the sender reveals $x_1 = 1$, the receiver attaches a posterior probability of $1/3$ to the state $(1,1,1)$ and the remaining probability to $(1,0,0)$. The deniability constraint is met and the receiver is willing to pass the bill in this case; and similarly in cases where the sender reveals $x_2 = 1$ and $x_3 = 1$.

When $x \in \{(0,1,1),(1,0,1),(1,1,0)\}$ and only one citizen is hurt by the bill, the sender reveals the identity of the citizen who is hurt, i.e., $x_i = 0$ for some i. Each such report is pooled with a report that is sent when $x = (0,0,0)$ where the sender again randomizes, revealing each of the three components of x with probability $1/3$. Once again it is easy to check that deniability will be met for each possible selected fact. Thus, in this example, the sender persuades the receiver to always pass the bill, either via evidence that supports a claim that a majority of citizens will benefit from it, or via evidence that suggests only a minority will be hurt by it. By Theorem 1, such a subversive strategy exists if and only if (3) holds.[8]

To obtain Theorem 1 we assumed each $w(x)$ is rational. This restriction is not important for the 'only if' part, as an inspection of the proof reveals. Since rationals are dense in the reals, it is also not a substantive restriction for the 'if' part. Any problem with an irrational weights can be approximated by a problem with rational weights, as Lemma 3 in the Appendix shows.

4 Optimal strategies

When a subversive strategy exists, it is a sender-optimal strategy. But when (3) fails, a C-perfect matching does not exist on ${}_cG$ and subversion is impossible. In this section we characterize the sender's optimal reporting strategies for all situations, whether or not subversion is possible. To do so, it will be helpful to clone states in the same way as was done in Section 3.

We proceed as follows. First, we set up the sender's optimization problem as a linear program on the cloned space ${}_cA \cup {}_cC$. Then we show that this problem has a solution in pure strategies in the cloned space.[9] As a consequence, a sender-optimal strategy can be described as a maximum-

of the additional (identical) edges these clones create.

[8]In Figure 3, subversion is impossible when $k > 1$ because (3) fails for $S = \{(0,0,0)\}$ when $k = 2$, and for all $S \subseteq C$ when $k = 3$, because $N(S)$ is empty.

[9]This is consistent with randomizations by the sender in the original space, as shown in the previous section.

weight matching on an edge-weighted graph ${}_{\eta}G$ that is constructed from the cloned graph ${}_{c}G$, using the sender's cardinal utility function v. We show further that such a maximum-weight matching must also be a maximum cardinality matching on ${}_{c}G$. This allows us to derive expressions for the sender's and receiver's expected payoffs from any sender-optimal reporting strategy.

As before, the sender is able to guarantee her first-best outcome on all states $x \in R$, since a reject decision always satisfies the deniability constraint (1). For each $x \in A \cup C$, let $\widehat{v}(x) = h\frac{v(x)}{\|u(x)\|}$, where h was defined in Section 3. For a clone ${}_{c}^{i}x \in {}_{c}A \cup {}_{c}C$ of $x \in A \cup C$, $\widehat{v}(x)$ gives the payoff to the sender at ${}_{c}^{i}x$ from accepting the proposal, but scaled up by the constant h. In order to avoid notational clutter, we erase the distinction between a state x and its clone ${}_{c}^{i}x$ in what follows and let $\widehat{v}(x)$ be the sender's payoff from $d = a$ at $x \in {}_{c}A \cup {}_{c}C$.

Using the convention $\mathcal{F}({}_{c}^{i}x) = \mathcal{F}(x)$, we assume each $x \in {}_{c}A \cup {}_{c}C$ sends a report $m \in \mathcal{M}(x) = \mathcal{F}(x) \times \{a, r\} \times \mathbf{M}$, and let$\mathcal{M}_a(x) = \{m \in \mathcal{M}(x) : d(m) = a\}$ with $\mathcal{M}_a = \cup_x \mathcal{M}_a(x)$. A reporting strategy σ picks $\sigma(x) \in \Delta(\mathcal{M}(x))$ for each $x \in {}_{c}A \cup {}_{c}C$. A solution to the following program is a sender-optimal reporting strategy.

$$\begin{aligned}
&\max_{\sigma} \sum_{x \in {}_{c}C} \sum_{m \in \mathcal{M}_a} \widehat{v}(x)\, \sigma(x)[m] \qquad (4)\\
\text{s.t.}\quad &\sum_{x \in {}_{c}C} \sigma(x)[m] - \sum_{x \in {}_{c}A} \sigma(x)[m] \le 0, \text{ for all } m \in \mathcal{M}_a\\
&\sum_{m \in \mathcal{M}_a(x)} \sigma(x)[m] \le 1, \text{ for all } x \in {}_{c}A \cup {}_{c}C\\
&\sigma(x)[m] \ge 0, \text{ for all } x,\ m.
\end{aligned}$$

To see why this describes the sender's problem, note first that it is never optimal for the sender to send a report in $\mathcal{M}_a$ (i.e., recommend $d = a$) with positive probability when the state belongs to R. If instead he chooses $d = r$ he obtains higher payoffs from doing so, and the deniability constraint is now relaxed for any $m \in \mathcal{M}_a$ that he may have been sending with positive probability. Symmetrically, the sender must always send a report in $\mathcal{M}_a$ for $x \in {}_{c}A$ because he can ensure $d = a$ by revealing such a state. For these reasons, only states in ${}_{c}C$ appear in the objective function, where we recall that all states ${}_{c}A \cup {}_{c}C$ are equally weighted. Since the sender obtains zero payoffs from $d = r$, it also suffices to restrict attention to messages in $\mathcal{M}_a$. The first constraint of the program (4) is the deniability constraint (1), adapted to the cloned space. The inequality in the second constraint allows $m \notin \mathcal{M}_a$ to be sent with strictly positive probability, while the last constraint is a non-negativity constraint on probabilities.

Lemma 1 *The linear program (4) has a solution in pure strategies:* $\sigma(x)[m] \in \{0, 1\}$ *for all* $x \in {}_{.c}A \cup {}_{c}C$, $m \in \mathcal{M}_a$.

The proof (see the Appendix) uses some classic results in linear programming. It consists of showing (via a result attributed to D. Gale) that the coefficient matrix of the constraint set is totally unimodular, i.e., every square submatrix of the coefficient matrix has determinant equal to 0 or ± 1. Total unimodularity implies every extreme point of the constraint set consists of integers. Since the objective function is also linear, the result follows. Lemma 1 is useful because it allows us to restrict attention to matchings on the cloned graph ${}_cG$, as we look for sender optimal reporting strategies.

Lemma 2 *Any pure strategy σ with $\sigma(x)[m] \in \{0,1\}$ for $x \in {}_{.c}A \cup {}_cC$, $m \in \mathcal{M}_a$ that satisfies the constraints of (4) can be described as a matching on ${}_cG$, and conversely.*

Let $\eta : E({}_cG) \to \mathbb{R}$ be an edge-weight function for the graph ${}_cG$. For any edge $\{x, x'\} \in E({}_cG)$, with $x \in {}_cC$, $x' \in {}_cA$, we will use the weight $\eta(\{x, x'\}) = \widehat{v}(x)$, and call the resulting edge-weighted graph ${}_\eta G = \{{}_cG, \eta\}$. The sum of the edge-weights of a matching M on ${}_\eta G$ is the weight of the matching M. Let $\nu^*({}_\eta G)$ be the weight of the *maximal-weight matching* on ${}_\eta G$ and call it the *matching number* of ${}_\eta G$. In general, a maximum-weight matching may not maximize the number of matches (a *maximal-cardinality matching*). Nevertheless, because of the way the edge weights η are constructed in our problem, the two will coincide as our next result shows.

Theorem 2 *The sender's optimal reporting strategy can be described as a maximal-weight matching on ${}_\eta G$. It is also a maximal-cardinality matching on ${}_\eta G$ (equivalently, on ${}_cG$).*

Proof. Lemmas 1 and 2 show that a solution to (4) can be described as a maximum-weight matching on ${}_\eta G$. Given the definition of the edge-weights, it follows that the value of the program (4) must equal the matching number $\nu^*({}_\eta G)$,

To complete the proof, we need to show that any maximum-weight matching on ${}_\eta G$ must also be a maximum-cardinality matching. We prove the contrapositive. Let M be a matching on ${}_\eta G$ that is not of maximum-cardinality. By Berge's theorem (1957) there is an M-augmenting path in ${}_\eta G$ (Lovasz and Plummer, 2009, Theorem 1.2.1), denoted P. Such a path starts from an unmatched vertex, alternates between edges that are in the matching M and edges that are not, and ends on another unmatched vertex.

Figure 4 shows an example for the matching in panel (a) that is represented by the green dot. The augmeting path for this matching starts at the unmatched vertex x and finishes on the unmatched vertex $\hat{x}'$ and it is given by the set of edges $\{(x, \hat{x}), (\hat{x}, x'), (x', \hat{x}')\}$. It creates a new matching $\{(x, \hat{x}), (x', \hat{x}')\}$ that is shown in panel (b). In general, the augmenting path creates a new matching $M' = P \triangle M$, where $\triangle$ denotes the symmetric set difference.[10]

[10] The symmetric set difference $P \triangle M = (P \cup M) \setminus (P \cap M)$.

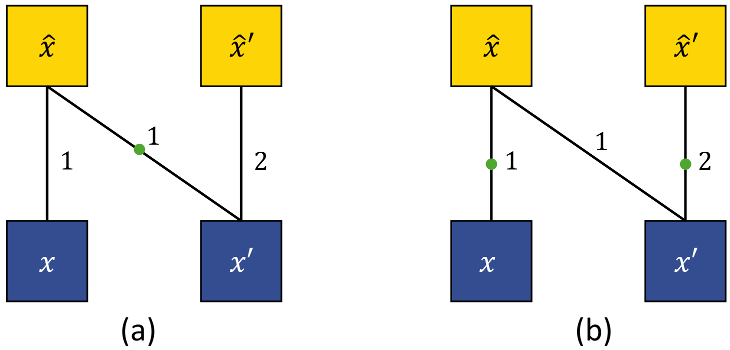

Figure 4: Augmenting paths

This new matching M' matches each vertex that was matched under M (now to a different neighbor), and also matches the two additional vertices at the start and end of P (one in ${}_cA$ and one in ${}_cC$). Since the edge weights in our problem depend only on the vertices $x \in {}_cC$, and $v(x) > 0$ for all such x , we see that the edge-weight of the new matching M' must be greater than the edge-weight of M. Therefore, M cannot be a maximum-weight matching.[11] ■

Since the sender will choose $d = a$ at each $x \in {}_cA$, whether or not it is matched, Theorem 2 implies the sender's expected payoff equals $\sum_{x \in_c A} \widehat{v}(x) + \nu^*({}_\eta G)$. The sender gets what she wants in all states $x \in A$. The weight of the maximum weight matching $\nu^*({}_\eta G)$ corresponds to her expected payoff from the states $x \in C$ on which the receiver agrees to accept the proposal. The second part of the theorem, which states that any maximum weight matching is also a maximum cardinality matching, allows us to pin down the the receiver's expected utility.

Proposition 1 *The receiver's expected payoff is*

$$\sum_{x \in A \cup C} p(x) u(x) + \max_{S \subseteq C} \left[- \sum_{x \in S \cup N(S)} p(x) u(x) \right].$$

It is independent of the sender's optimal strategy σ (i.e., the sender-optimal matching).

Proposition 1 is a corollary of Theorem 2 and its proof follows from the König-Ore formula which characterizes the number of unmatched vertices in any maximum cardinality matching.[12] The rest is simply accounting for the mapping between our model and the graph ${}_cG$. The second

[11] In general matching problems, the weight of an edge $(x, \hat{x})$ will depend on both vertices x and $\hat{x}$. Thus, a maximum-weight matching may not also be a maximum-cardinality matching in models with "lying costs" (see, e.g., Kartik, 2009) where the sender's payoff directly depends on the report in addition to the state.

[12] See Theorem 1.3.1 in Lovász and Plummer (2009).

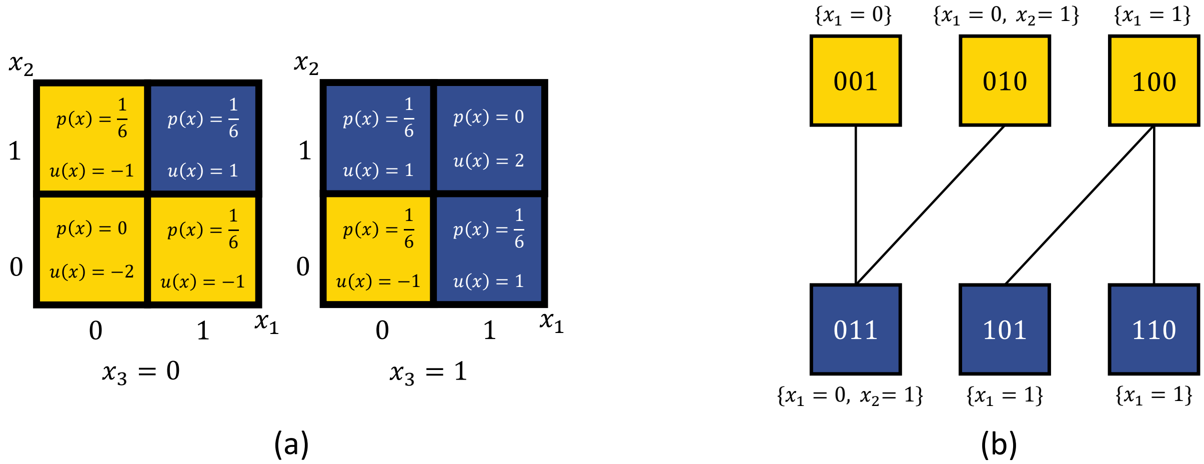

Figure 5: Optimal strategies

term in the expression for the receiver's expected payoff has an interesting interpretation. Using the relationship between the original graph G and the cloned graph ${}_cG$ we may write

$$\frac{1}{h}\max_{S\subseteq\mathcal{C}}\left[-\sum_{x\in S\cup N(S)} p(x)\,u(x)\right]=\max_{S\subseteq_c\mathcal{C}}\left[|S|-|N(S)|\right]\equiv\delta_H \tag{HD}$$

We call the number δ_H the "Hall deficit" on ${}_cG$. When Hall's condition (HC) is satisfied on ${}_cG$, and subversion is possible, $\delta_H=0$. In this case, the empty set solves the maximization problem in (HD). Otherwise, δ_H is positive. Proposition 1 shows that the receiver's total payoff equals what she would get if the sender gets his ideal decisions plus a term proportional to the Hall deficit δ_H.

When subversion is not possible, the sender has to make choices about the conflict states in C that he can match with states in A. These choices, and the sender's optimal strategy σ, will depend on the intensity of his preferences, i.e., on the weights $\widehat{v}$ that are derived from his cardinal payoff function v. But since δ_H does not depend on σ, the receiver's expected payoff does not depend on the actual choices the sender makes. This is illustrated by the example in Figure 5.

As with the example of Figure 3, we suppose in Figure 5 that the sender knows if a bill would help or hurt each of three citizens and the sender always wants the bill to be passed. Priors and receiver preferences are different from the earlier example, as shown in Figure 5. As for the fact sets, we suppose that the sender can always prove whether the bill will help or hurt citizen 1. In addition, he can also prove the bill will help citizen 2 but only when it hurts citizen 1. That is, $\mathcal{F}(x)=\{x_1,x_2\}$ for $x\in\{(0,1,0),(0,1,1)\}$ with $\mathcal{F}(x)=\{x_1\}$ otherwise. The cloned graph ${}_cG$, which is identical to G in this non-Cartesian example, is shown in the figure.

As can be seen from the graph, when $x = (1, 0, 0)$ and only citizen 1 benefits from the bill, the sender can persuade the receiver to pass the bill by revealing $x_1 = 1$ and matching this state either with $(1, 1, 0)$ or with $(1, 0, 1)$ and meet the deniability constraint. However, the sender can get the bill passed and meet deniability in just one of the two other conflict states $(0, 1, 0)$ and $(0, 0, 1)$. This is because (HC) fails for $S = \{(0, 1, 0), (0, 0, 1)\}$ which has $N(S) = \{(0, 1, 1)\}$, implying subversion is not possible.

A sender with cardinal preferences that reflect a strong preference to pass the bill when citizen 2 benefits from it will prefer to match $(0, 1, 0)$ with $(0, 1, 1)$. In contrast, a sender with a strong desire to pass the bill when citizen 3 benefits may match $(0, 0, 1)$ with $(0, 1, 1)$. In general, one may expect that different kinds of senders, with different cardinal preferences over the conflict states, may be more or less constrained by the set of available facts that they can exploit in each state. For instance, in Figure 5 the sender who has a strong preference to pass the bill when citizen 3 benefits may seem more constrained than a sender who prefers to pass it when citizen 2 does, because the latter can sometimes prove his favored citizen benefits while the former cannot.

This suggests, in turn, that the receiver may prefer facing a sender who is more constrained by the available fact sets, compared to another sender who is less so. Proposition 1, a consequence of Theorem 2, shows this intuition is incorrect. While the sender's optimal strategy depends on his cardinal preferences, the receiver is equally well off from facing all possible kinds of senders. This is because each possible sender-optimal strategy (a maximum-weight matching on ${}_{\eta}G$) must also be a maximum cardinality matching on ${}_{c}G$, and ${}_{c}G$ does not require specifying the cardinal preferences of the sender. This fact is the key input for our results in the next section.

5 Receiver-optimal fact sets

So far we have taken the fact sets available to the sender as primitives and restricted the receiver to a passive role. In this section, we ask how the receiver would like to modify the fact sets at the very beginning, before the sender commits to a reporting strategy. We model the receiver's design problem as follows.

Take as given the non-empty fact sets $\mathcal{F}(x)$, $x \in X$, with $\mathcal{F} = \cup_{x \in X} \mathcal{F}(x)$. We suppose the receiver can specify a smaller set $\mathcal{F}' \subseteq \mathcal{F}$ that the sender is restricted to, subject to the constraint $\mathcal{F}'(x) \equiv \mathcal{F}' \cap \mathcal{F}(x)$ be non-empty for each $x \in X$. Designing the set $\mathcal{F}'$ of *admissible facts* is the only instrument available to the receiver. We can think of $\mathcal{F}'$ as describing the rules of argumentation and evidence production that the receiver imposes on the sender at the outset. Notice the receiver can specify $\mathcal{F}'$ given only her knowledge of each $\mathcal{F}(x)$ but without knowing the realized state x.

The only constraint on the receiver is that she must allow the sender to present some fact in every state, i.e., each $\mathcal{F}'(x)$ must be non-empty. Each possible $\mathcal{F}'$ creates a subgraph $G(\mathcal{F}')$ of G that (possibly) removes some edges of G. Let $\delta_H(\mathcal{F}')$ be the Hall deficit on $G(\mathcal{F}')$.

Theorem 3 *Let $\mathcal{F}^*(\mathcal{P}) \subseteq \mathcal{F}$ be the receiver-optimal set of admissible facts for a problem $\mathcal{P}$. It solves $\max_{\mathcal{F}' \subseteq \mathcal{F}} \delta_H(\mathcal{F}')$ subject to the constraints $\mathcal{F}'(x)$ be non-empty for each $x \in X$. If problem $\mathcal{P}'$ differs from $\mathcal{P}$ only in the specification of the sender's cardinal payoff function v, then $\mathcal{F}^*(\mathcal{P})$ is receiver-optimal also for $\mathcal{P}'$.*

Proof. The proof follows from Proposition 1. Notice first that we can work interchangeably with the graph G and its cloned graph ${}_cG$. Each specification of a set $\mathcal{F}'$ of admissible facts gives rise to a subgraph $G(\mathcal{F}')$ and its corresponding clone ${}_cG(\mathcal{F}')$. On this subgraph, the receiver's expected payoff is given by Proposition 1; see, in particular, expression (HD). It is independent of v and the matching used by the sender on ${}_cG(\mathcal{F}')$. It follows that the set $\mathcal{F}'$ that maximizes the receiver's expected payoff must have the maximum Hall deficit $\delta_H(\mathcal{F}')$ among all feasible alternatives; and that this solution will be optimal also for a problem $\mathcal{P}'$ that differs from $\mathcal{P}$ only in the specification of v. ■

We first apply Theorem 3 to the special case of a Cartesian problem with $k = 1$. Let $\mathcal{F}^i$ be an admissible fact set that restricts the sender to present facts only on a particular aspect i, i.e., $\mathcal{F}^i(x) = \{x_i\}$ for each x. Any value of any other aspect is not an admissible fact. Let δ_H^i be the Hall deficit of $G(\mathcal{F}^i)$.

Proposition 2 *Consider a Cartesian problem with $k = 1$. The set of admissible facts $\mathcal{F}^{i^*}$ that restricts the sender to present facts only on a particular aspect $i^* = \arg\max_i \delta_H^i$, is receiver-optimal. In particular, i^* does not depend v.*

Proof. Consider the constraint on the receiver-optimal fact set $\mathcal{F}^*$ that $\mathcal{F}^*(x)$ be non-empty for each x. A necessary condition for this to obtain is that all possible values of some aspect i must belong to $\mathcal{F}^*$. For if there is no aspect all of whose possible values are acceptable facts, there must exist a state $x \in X$ every component of which is not an acceptable fact, because the set of possible states X is product set. Then $\mathcal{F}^*(x)$ is empty for such an x, a contradiction.

Note next that this necessary condition is also sufficient for the non-emptiness of each fact set, i.e., a fact set of the form $\mathcal{F}^i$ is feasible for the receiver's problem. Since allowing any additional facts, in addition to all possible values of aspect i, only (weakly) reduces the Hall deficit of the resulting subgraph, an admissible fact set of the form $\mathcal{F}^i$ is (weakly) better for the receiver than a larger fact set $\mathcal{F}' \supseteq \mathcal{F}^i$. Since this is true for each i, an admissible fact set $\mathcal{F}^{i^*}$, that restricts

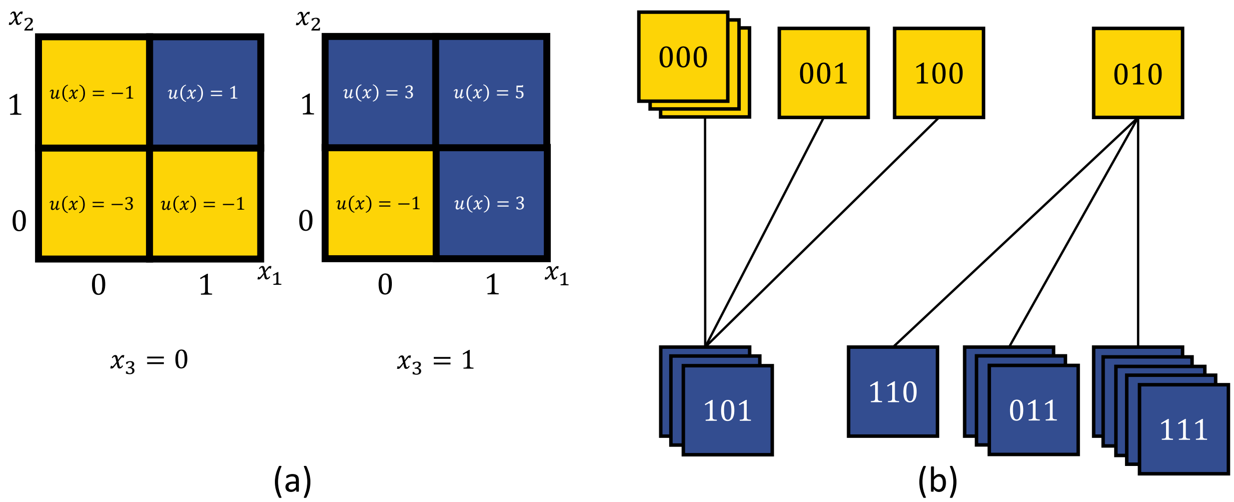

Figure 6: Admissible fact sets and Hall deficit I

the sender to provide facts only on some aspect i^* of the problem, must be receiver-optimal. Using Theorem 3, we see $i^* = \arg\max_i \delta_H^i$ and it does not depend on v. ■

Figure 6 provides a Cartesian example with $k = 1$ that illustrates Proposition 2. As with the earlier examples, the sender wants the receiver to pass a bill and he knows whether or not the bill would help or hurt each of three citizens. Priors are uniform iid. The agreement and conflict sets as well as receiver payoffs u from accepting the proposal are shown in the figure.

By Proposition 2, the receiver wants to allow the sender to present facts only about a particular citizen, the one that results in the maximum Hall deficit in the resulting subgraph. This Hall deficit is easy to calculate in Cartesian problems. In particular, when i is the only admissible aspect, $\mathcal{F}'(x) = \{x_i\}$ for each x, the Hall deficit of the resulting graph is given by

$$\delta_H^i = \frac{1}{h} \sum_{x_i} \max \left[-\sum_{x_{-i}} p(x)u(x), 0 \right]. \tag{5}$$

To see why this is the case, consider the graph in Figure 6 where the receiver is only willing to admit facts about citizen 2 as evidence, i.e., $\mathcal{F}'(x) = \{x_2\}$. Notice first that this graph consists of two subgraphs, one for the case $x_2 = 0$ and another for the case $x_2 = 1$, that are disconnected from each other. Notice also that each of these two subgraphs must be a complete bipartite graph because all its vertices share a fact corresponding to a particular value of the admissible aspect x_2. This implies the overall Hall deficit δ_H^2 can be calculated simply by first focusing on the "grand coalition" of all conflict states in each subgraph, and then summing over the subgraphs. Formula (5) follows from this observation. Using it, we see that $\delta_H^2 = 2 + 0 = 2$ and, similarly, $\delta_H^1 = 2$ while

$\delta_H^3 = 4$. Thus, in Figure 6, it is optimal for the receiver to ask the sender to only provide facts about citizen 3. This is at least as good as any other set of admissible facts the receiver can specify.

It is straightforward to extend Proposition 2 to Cartesian problems with $k > 1$. For instance, if $k = 2$, identical arguments to those in the proof of Proposition 2 establish that the receiver will admit all possible values of two aspects i and j, and no values of any other aspects, i.e., fact sets of the form $\mathcal{F}^*(x) = \{(x_i, x_j)\}$ are optimal. The choice of the two unrestricted aspects will be determined by the Hall deficits of the subgraphs created by each pair of unrestricted aspects. Given any such pair $\{i, j\}$, this Hall deficit can be calculated by focusing on the grand coalition on a subgraph corresponding to a particular value of $\{x_i, x_j\}$, and then summing across these disconnected subgraphs to obtain a formula similar to (5). We focus on Cartesian problems with $k = 1$ in Proposition 2 purely for the sake of expositional convenience.

We can think of the receiver optimal admissible aspect(s) as a topic or subject matter imposed on the sender before he makes his arguments. This optimal topic does not depend on v. Even though the sender's cardinal preferences v determine the arguments the sender will actually make, the receiver does not need to predict these arguments to determine her choice of topic. Her optimal topic describes the facts she would like to know, given only that the sender wants to accept the proposal (i.e., $x \notin R$). In particular, when R is empty and the sender always wants to accept the proposal, the receiver's optimal topic depends only on priors and her own preferences.

In Cartesian problems, with her choice of a topic, the receiver eliminates the sender's freedom to select facts. For each x, the sender is forced to reveal the values of the k aspects the receiver accepts. He cannot choose between facts. This intuitive implication of Proposition 2 does not extend to non-Cartesian problems. Figure 7 provides an example where each possible state is equally likely and described by the values of two aspects. It is not a Cartesian problem because $\{0, 1\} \times \{0, 1\}$ has zero probability, i.e., the set of possible states X is not a product set. The sender must reveal the value of one of the two aspects and the initial (unrestricted) fact sets are of the form $\mathcal{F}(x) = \{x_1, x_2\}$. Receiver preferences take the form $u \in \{-1, +1\}$, R is empty and the sets A and C are as depicted in panel (a). Panel (b) shows the corresponding (cloned) graph and it allows subversion.

Proposition 2 does not cover this example and we have to use Theorem 3 to determine the receiver-optimal admissible fact set. Consider the admissible fact set $\mathcal{F}'$ depicted in Figure 7(a) that requires the sender to either reveal the value of x_1 or of x_2, provided the revealed value is in the set $\{2, 3\}$. In effect, the receiver asks the sender to reveal which of the two aspects has a value higher than its expected value. Such a fact set is feasible because $\mathcal{F}'(x)$ is non-empty for each $x \in X$. The graph corresponding to this fact set is obtained by deleting the dotted edges from the

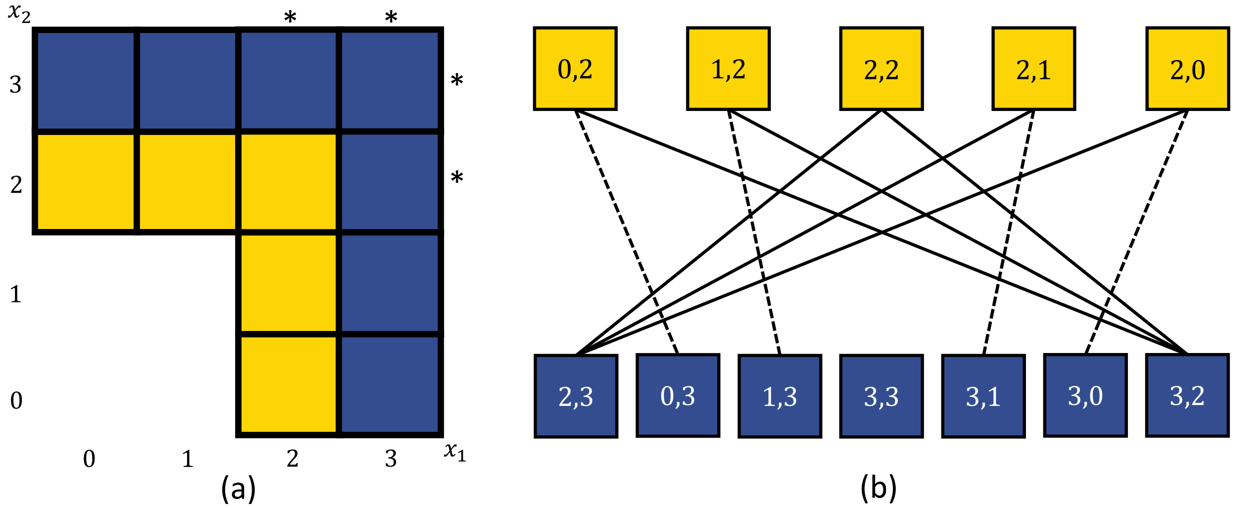

Figure 7: Admissible fact sets and Hall deficit II

graph in panel (b). It has $\delta_H = 3$ and it allows the sender to make two matches.

It is not difficult to directly verify that this admissible fact set has the largest possible Hall deficit among all possibilities.[13] By Theorem 3, it is optimal for the receiver even though it gives the sender flexibility to select facts in some states. Senders with different cardinal preferences v may make use of this flexibility in different ways and make different matches. But all of them will only make two matches in total.

In particular, this admissible fact set is better for the receiver than restricting the sender to provide facts only on a particular aspect because such fact sets allow the sender to make three matches, as can be easily verified from the figure. While eliminating the sender's freedom to select facts is always optimal in Cartesian problems, this is not generally true. What matters is minimizing the diversity of possible arguments available to the sender *across* different conflict states, i.e., finding the subgraph that has the largest Hall deficit.

Finally, even in Cartesian problems, if the receiver faces additional constraints on the admissible fact set, she may have to leave the sender with some freedom to select facts. For instance, certain facts $\mathcal{U} \subseteq \mathcal{F}$ may be "undeniable", in the sense that the receiver-optimal admissible fact set $\mathcal{F}^* \subseteq \mathcal{F}$ has to satisfy the additional constraint $\mathcal{U} \subseteq \mathcal{F}^*$. Proposition 2 (in particular, expression 5) does not cover such cases but Theorem 3 does.

[13] The non-emptiness constraint on $\mathcal{F}'(x)$ applied to states of the form $(x_1, 2)$ and $(2, x_2)$ implies the sender must be able to make at least two matches, and so the Hall deficit can at most be three, for any feasible fact set. The fact set depicted in Figure 7 achieves this upper bound on δ_H.

6 Additional remarks

In this section, we offer a brief discussion of some key features of our model, along with some comparative statics results and alternative assumptions.

On receiver-optimal fact sets. So far we have assumed that in designing admissible fact sets the receiver requires the sender to provide a supporting fact regardless of the decision the sender recommends (or takes). We now ask what will happen if the receiver can specify admissible fact sets $\mathcal{F}(x)$ that are empty when $x \in R$. Since in these states the sender will recommend $d = r$, this amounts to the receiver waiving the requirement for the sender to provide a fact when he recommends rejection.[14]

Relaxing some of the (non-emptiness) constraints on the receiver's design problem in this way cannot make the receiver worse off. To see that she may be strictly better off, consider an altered version of the example of Figure 7. Suppose that instead of having zero probability, $\{0,1\} \times \{0,1\} = R$ with $p(R) > 0$. This changes the example to a Cartesian problem. If every admissible fact set $\mathcal{F}'(x)$ (including for $x \in R$) must be non-empty, by Proposition 2, it is optimal for the receiver to admit all values of one aspect. This results in a Hall deficit of two and allows the sender to make three matches.

If instead the receiver can let admissible fact sets $\mathcal{F}'(x)$ be empty for $x \in R$, her design problem is the same as that in Figure 7, where $\{0,1\} \times \{0,1\}$ had zero probability. As shown earlier, the receiver-optimal fact sets in Figure 7 admit $x_i \in \{2,3\}$ as a fact, for $i = 1, 2$. This results in a Hall deficit of three and allows the sender to make only two matches. So the receiver is strictly better off from dropping the requirement of a supporting fact when the sender proposes rejection. More generally, once we allow admissible fact sets to be empty for $x \in R$, Proposition 2 applies only to Cartesian problems where R is empty.[15]

Comparative statics on fact sets. Fix the set of possible states X together with preferences and priors that define a problem. Consider two senders, α and β , who differ only in the facts that they have available in different states. Call α more *adept* than β if $\mathcal{F}_\alpha(x) \supseteq \mathcal{F}_\beta(x)$ for all $x \in X$, with strict inclusion for some $x \in X$. It is easy to see that, ceteris paribus, α must earn at least as high an expected payoff as β because he can provide all the facts (and so make all the matches) β can, and perhaps some more. We discuss below some examples where α and β do equally well and identify conditions for α to do strictly better than β.

[14] We continue to require the sender to provide a fact when he recommends $d = a$, implying admissible fact sets must be non-empty for each $x \notin R$. Dropping this requirement would transform our constrained information design problem to the standard unconstrained one.

[15] Theorem 3 continues to hold with the modification that $\mathcal{F}'(x)$ must be non-empty only for $x \notin R$.

For the first example, suppose X is a n-dimensional product set and assume that $\mathcal{F}_\beta(x) = \{\phi \subset \{x_1, ..., x_n\} : |\phi| = k\}$ for some $k < n$. Let $\mathcal{F}_\alpha(x) = \{\phi \subset \{x_1, ..., x_n\} : |\phi| \geq k\}$, so that α is more adept than β. Compare the graph G_α that describes α's problem to G_β that describes β's problem. Note that whenever $\mathcal{F}_\alpha(x) \cap \mathcal{F}_\alpha(x')$ is non-empty for some $x \in C$ and $x' \in A$, so is $\mathcal{F}_\beta(x) \cap \mathcal{F}_\beta(x')$. Thus, G_β must have the same set of edges as G_α, i.e., the two graphs are identical. This implies α and β must employ the same optimal strategy (specifically, matching) and so they must be equally well off, and the receiver must also be equally well off from facing either of them.

For the second example, suppose β is the sender in the example of Figure 3, and assume that any logical implication of a fact for β is a fact for the more adept sender α.[16] For instance, β can prove the fact $\phi =$"citizen 1 benefits from the bill" in state $x = (1, 0, 0)$. This implies the fact $\psi =$"at least one citizen benefits from the bill" is also true at x. We suppose ψ is a fact for α at x. Notice that by using ψ, α can match $x = (1, 0, 0)$ with $x' = (0, 1, 1)$, which is not possible for β given his fact sets. So, the graph G_α has more edges than G_β. Nonetheless, α will do exactly as well as β in this example since β has a subversive strategy. For a more adept sender α to do strictly better than a less adept sender β (and the receiver to be worse off when facing α), it is necessary and sufficient that G_α has a smaller Hall deficit than G_β, by Theorem 2 (and Proposition 1).

This comparative static result does not extend to cases where the receiver first designs an optimal set of admissible facts tailored to the sender she will face. The receiver may be able to constrain a more adept sender more than a less adept one and so she may be prefer facing the former. For instance, consider the example of Figure 6 and suppose β's fact sets are as in that example. But α also has all logical implications of β's facts available as facts. As shown earlier, it is optimal for the receiver to constrain β to reveal facts about citizen 3, giving a Hall deficit of 2. But the receiver is better off against α because she can ask him simply to reveal whether or not a majority benefits from the bill to obtain her ideal payoffs (i.e., a Hall deficit of six).[17]

Limited commitment. We assume the sender can commit to a reporting strategy. In general, this commitment is valuable but there are some special cases where the sender has no incentive to deviate from his strategy after learning the state. We conclude this section with a brief description of some simple situations where the value of commitment is zero.

The easiest case to consider where commitment has no value is when the conditions of Theorem 1 obtain and the sender has a subversive reporting strategy. Since the sender achieves his ideal payoffs

[16] If $\phi \in \mathcal{F}_\beta(x)$ and $\phi \Rightarrow \psi$ (logically), then $\psi \in \mathcal{F}_\alpha(x)$.

[17] For an example where the receiver is worse off against the more adept sender, even after optimizing her admissible fact sets, suppose $X = \{x, x'\}$, with $x \in C$ and $x' \in A$ (and uniform priors with $u \in \{-1, +1\}$). Suppose $\mathcal{F}_\alpha(x) = \{\phi, \psi\}$, $\mathcal{F}_\alpha(x') = \{\phi\}$, $\mathcal{F}_\beta(x) = \{\psi\}$, $\mathcal{F}_\beta(x') = \{\phi\}$. Then $\mathcal{F}^*_\alpha = \{\phi\}$ and $\mathcal{F}^*_\beta = \{\phi, \psi\}$ are receiver-optimal fact sets for the two senders and the receiver is better off facing the less adept sender β.

with such a strategy there is no incentive to deviate from it. Another easy case where the value of commitment could be zero is one where the sender can deviate from the decision recommendation but he is committed to the facts his strategy (or experiment) is supposed to disclose. When X is a product set and the sender can reveal at least k components of the state, the sender can obtain his commitment outcome by fully revealing the state $x \in C$ whenever his optimal commitment strategy recommends rejecting the proposal in that state.

The value of commitment is also zero in situations the receiver can detect that a deviation from the experiment may have taken place but cannot deduce from the observed report what the actual deviation is (e.g., an illegal after hours login or other suspicious activity). One can sustain the commitment outcome in this case by assuming the receiver attaches degenerate (skeptical) beliefs to some $x \in C$ (and prefers rejection), after detecting a deviation has occurred, provided such a state exists that is consistent with the facts revealed in the report; otherwise, the receiver follows the sender's recommendation. Since the sender always gets his desired decision for $x \in A$ under his optimal commitment strategy, these beliefs eliminate his incentive to deviate from the experiment.

A final case where the value of commitment is easy to pin down is the grain of truth specification of our model where $\mathcal{F}(x) = \{\phi \in 2^X : x \in \phi\}$. In this case, the sender cannot profitably deviate from his optimal commitment strategy after learning the state if and only if for each $x \in C$, either the proposal is accepted for sure or it is rejected for sure. To see this, consider the only if direction first and suppose the sender-optimal strategy under commitment induces both decisions with positive probability for some $x \in C$. Then the sender has an incentive to deviate and reveal a fact that is available to him, and which induces acceptance instead of rejection, after learning the state is x.

For the if direction, note that the sender-optimal commitment strategy can always be implemented by using the fact $\{x, x'\} \in \mathcal{F}(x) \cap \mathcal{F}(x')$ for each matched pair $\{x, x'\}$ with $x \in C$, $x' \in A$. Such a report cannot be mimicked by any $x'' \in C$, $x'' \neq x$, and so the sender has no profitable deviation available after he learns the state. Outside of these easy cases, the problem of identifying sender-optimal incentive compatible reporting strategies is non-trivial and it is the subject of our ongoing research.

7 Related literature

This paper belongs to the information design literature (Rayo and Segal, 2010; Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019; Kamenica, 2019). In contrast to the standard belief-based approach, we use graph-theoretic techniques to capture constraints on the kinds of experiments that the sender can design. Bayes plausibility is built into the structure of our cloned graphs in a way that ensures the obedience (or, deniability) constraint is satisfied for every matching. The fact selection constraints are captured by the presence (or absence) of edges in our graph. They imply that the sender can induce any posterior for which all states in the support share some common fact. In contrast, Kolotilin, Corrao and Wolitzky (2024) and Kolotilin and Wolitzky (2024) describe the Bayes plausible posteriors that can arise in a persuasion setting in the language of (assortative) matching but they have no additional constraints on experiments.[18]

Our paper is also closely related to the work of Glazer and Rubinstein (2004, 2006). They study a mechanism design problem in which receiver commits to a persuasion rule that selects a fact after the sender sends a message. They characterize receiver-optimal persuasion rules, when the sender has state-independent preferences (i.e., seeks to maximize the likelihood of the proposal being accepted). We study the mirror-image problem of sender-optimal rules of persuasion, under general preferences for both the sender and the receiver, when the sender has commitment power and selects the facts to be revealed.[19] We also solve the receiver's problem of designing a restricted set of admissible facts that imposes additional constraints on the arguments available to the sender.

The early literature on games of hard information disclosure (Grossman and Hart, 1980; Grossman, 1981; Milgrom, 1981; Dye, 1985; Matthews and Postlewaite, 1985; Milgrom and Roberts, 1986; Fishman and Hagerty, 1990; Seidmann and Winter, 1997) considers models without commitment and focuses on receiver-preferred equilibria supported by skeptical off-path beliefs. According to these beliefs, any undisclosed information is taken to be detrimental to the sender.[20] In the context of this literature, Glazer and Rubinstein (2004, 2006) provide a foundation for equilibrium selection in favor of the receiver via conditions under which their receiver-optimal mechanism does not need the receiver to commit. This equivalence between equilibria without commitment and optimal mechanisms with commitment was also studied by Sher (2011), Hart, Kremer and Perry (2017), and Rappaport (2024) in the context of disclosure models with hard evidence. We differ from this literature in assuming the sender can commit to his strategy so that skeptical off-path beliefs have no role to play. While this commitment power is valuable in general, in some special cases it may be unnecessary. In such cases, we effectively select the sender's preferred equilibrium

[18]Patil and Salant (2024) and Alon, Auster, Gayer and Minardi (2024) study non-Bayesian persuasion environments where the sender's experiments are subject to specific constraints.

[19]When the receiver's verification strategy is deterministic in Glazer and Rubinstein (2004, 2006), the sender can predict and voluntarily reveal the fact the receiver will choose to verify, so the key difference between these papers and ours is our focus on sender-optimal constrained information design.

[20]Dziuda (2011) considers a disclosure game of where the sender can be an honest type, so that no reports are off-path.

in the induced game without commitment.

Dasgupta, Krasikov and Lamba (2022) study a monopoly pricing model where a buyer learns his value by designing an experiment that either provides hard information about his type, or produces a null signal that can be mimicked by all types. The buyer is worse off when he chooses whether or not to disclose such hard information to the seller, compared to a benchmark where the buyer can misreport his signal in any way. We consider more general constraints on information design in a binary action sender-receiver setting.[21]

8 Conclusion

We consider the problem of a sender trying to persuade a receiver to accept or reject a proposal. The sender must disclose some evidence but has a choice of what evidence to reveal. We show that this constrained persuasion problem can be solved in generality by casting it as a matching problem on a suitably chosen bipartite graph. The sender's optimal strategy can be described both as a maximum-weight matching and a maximum-cardinality matching on this graph. Hall's theorem provides the conditions under which the sender can persuade the receiver to implement the sender's ideal decisions. It also allows us to concisely describe the receiver's expected payoffs and the set of admissible facts that she would like to constrain the sender to.

9 Appendix

Lemma 3 *If $w(x)$ is irrational for some $x \in X$, a subversive reporting strategy exists if $\mathbb{E}[u|S \cup N(S)] > 0$ for all non-empty $S \subseteq C$.*

Proof of Lemma 3. For each $x \in X$, approximate $u(x)$ by $\widehat{u}(x) \in \mathbb{Q}$, so that $\widehat{u}(x) \leq u(x)$ and $\mathbb{E}[\widehat{u} \mid S \cup N(S)] > 0$ for all $S \subseteq C$, where the expectation is taken with respect to the prior p. Such an approximation exists by the density of rationals in the reals. For each $x \in X$ approximate $p(x)$ by $\widehat{p}(x) \in \mathbb{Q}$ such that $\widehat{p}(x) > p(x)$, and for each $x \notin C$ approximate $p(x)$ by $\widehat{p}(x) \in \mathbb{Q}$ such that $\widehat{p}(x) < p(x)$, so that $\sum_{x \in X} \widehat{p}(x) = 1$ and $\mathbb{E}[\widehat{u} \mid S \cup N(S)] \geq 0$ where the last expectation is taken with respect to priors $\widehat{p}$. Applying Proposition 1 to the problem where u and p are replaced by $\widehat{u}$ and $\widehat{p}$ yields a C-perfect matching, and an associated subversive reporting strategy. Observe that this matching in the approximated problem is still a valid matching in the original problem since the approximations where chosen to ensure (1) is satisfied. ■

[21] Roesler and Szentes (2017), Ali, Lewis and Vasserman (2023) and Madarasz and Pycia (2024) also consider models of information design in monopoly pricing environments.

Proof of Lemma 1.

The linear program (4) can be expressed in matrix form, i.e., finding a vector σ which solves

$$\max_{\sigma} \{\widehat{v} \cdot \sigma : B\sigma \leq b\} . \tag{6}$$

To see this, we let the rows of the vectors and matrices represent the list of $\cup_{x \in {}_cC \cup {}_cA} \{x\} \times \mathcal{M}$, which means that row (x, m) of the vector σ gives the value of $\sigma(x)[m]$. Let $\widehat{v}$ be the vector of $\widehat{v}(x)$ for each (x, m) (repeating the same value for each $m \in \mathcal{M}_a$). The matrix B captures all of the relevant constraints and can be expressed as

$$B = \left(\begin{array}{c} B_1 \\ \hline B_2 \\ \hline B_3 \end{array} \right),$$

that is, B is the concatenation of three matrices B_1, B_2 and B_3. Let B_1, which represents the first $|\mathcal{M}_a|$ rows of B be defined as follows. For each $m \in \mathcal{M}_a$, we have a row where the element multiplying $\sigma(x)[m]$ takes on the value 1 if $x \in {}_cC$ and $m \in \mathcal{M}_a(x)$, the value -1 if $x \in {}_cA$ and $m \in \mathcal{M}_a(x)$ and the value 0 otherwise. For B_2, which represents the next $|{}_cC \cup {}_cA|$ rows we have a row for each x where the element multiplying (x, m) takes on the value of 1 if $m \in \mathcal{M}_a(x)$ and 0 otherwise. For B_3 which is a square matrix representing the next $|\cup_{x \in {}_cC \cup {}_cA} \{x\} \times \mathcal{M}|$ rows of B we just have $B_3 = -I$, the negative of the identity matrix. The vector b is defined as $|\mathcal{M}_a|$ entries of 0, followed by $|{}_cC \cup {}_cA|$ entries of 1, followed by $|\cup_{x \in {}_cC \cup {}_cA} \{x\} \times \mathcal{M}_a|$ zero elements.

Theorem 21.5 of Schrijver (1998) states that B^T is unimodular if and only if there exists an integer vector σ which solves the matrix form linear program (6). We will show something stronger: that B is totally unimodular (TUM), which implies that B^T is TUM and hence unimodular. Recall that an integral matrix B is TUM if every square submatrix of B has determinant equal to 0 or ± 1. It is immediate that B is TUM if and only if $B_{12} := \left(\begin{array}{c} B_1 \\ \hline B_2 \end{array} \right)$ is TUM, since the rows that are removed have just one -1 value (and the rest of the values are zero), which means that if a submatrix with that row included is not TUM then the submatrix formed by deleting this row and the column where it takes on the value -1 must also fail to be TUM.

That B_{12} is TUM is due to a result attributed to David Gale in Heller and Tomkins (1956, Appendix, pp. 253): A matrix consisting of entries 0, ± 1 with exactly two non-zero entries in each column is TUM if and only if the rows of the matrix can be partitioned into two sets, such that two nonzero entries in a column are in the same partition element if they have opposite signs and in different partition elements if they have the same sign (see Schrijver, 1998 example 3 on page 276). Each column of B_{12} has two non-zero entries, one in B_1 and one in B_2. The rows of B_{12} can be partitioned as follows: let one partition element be the rows of B_2 involving $x \in {}_cC$, and the other partition element be all of the remaining rows. ■

Proof of Lemma 2. For each vertex $x \in {}_cC$ and $x' \in {}_cA$ that is part of a matching on ${}_cG$, we can let $\sigma(x)[m] = \sigma(x')[m] = 1$ for some $m \in \mathcal{M}_a$, with $\sigma(x'')[m] = 0$ for all other $x'' \in {}_cA \cup {}_cC$. Such a message m exists because $\{x, x'\}$ is an edge of ${}_cG$ and so they share a common fact. It is easy to see such a strategy will will satisfy all constraints of (4). In the other direction, a feasible pure strategy of problem (4) chooses $m \in \mathcal{M}_a$ with probability one on each vertex in ${}_cA \cup {}_cC$. Since σ satisfies the first constraint of (4) (the deniability constraint), for each vertex in $x \in {}_cC$ there will be (at least one) $x' \in {}_cA$ that sends the same message m. Since every such pair $\{x, x'\}$ must have a fact in common, they correspond to an edge of ${}_cG$, and so we can select each of these edges as part of our matching. Observe that this may result in unmatched elements of ${}_cA$, but that all elements of ${}_cC$ sending message m are matched. Repeating the process for all messages m that are picked by σ gives us a matching on ${}_cG$. ■

Proof of Proposition 1. By the König-Ore formula (Lovász and Plummer, 2009, Theorem 1.3.1) any maximum-cardinality matching on ${}_cG$ (or ${}_\eta G$, since edge-weights are not relevant) results in exactly

$$def\,({}_cG) = \max_{S \subseteq {}_cC} \left[|S| - |N\,(S)|\right] \tag{7}$$

vertices in ${}_cC$ that are unmatched.[22] Since each clone of an $x \in A \cup C$ has the same set of neighbors as any other clone, a solution $S \subseteq {}_cC$ of (7) must contain all the clones of a set $S' \subseteq C$, implying $N\,(S) \subseteq {}_cA$ must be the set of all clones of $N(S') \subseteq A$. In other words, the maximization problem in (7) can also be solved over subsets of C instead of ${}_cC$. Since each unmatched vertex of ${}_cC$ is worth an extra h utility to the receiver, we have

$$\begin{aligned}
h\,(def\,({}_cG)) &= h \max_{S \subseteq {}_cC} \left[|S| - |N\,(S)|\right] \\
&= h \max_{S \subseteq C} \left|\left\{{}^i_c x : x \in S\right\}\right| - \left|\left\{{}^i_c x : x \in N\,(S)\right\}\right| \\
&= h \max_{S \subseteq C} \sum_{x \in S} \frac{w(x)}{h} - \sum_{x \in N(S)} \frac{w(x)}{h} \\
&= \max_{S \subseteq C} \sum_{x \in S} -p\,(x)\,u\,(x) - \sum_{x \in N(S)} p\,(x)\,u\,(x) \\
&= \max_{S \subseteq C} \left[- \sum_{x \in S \cup N(S)} p\,(x)\,u\,(x)\right],
\end{aligned}$$

[22] This number $def\,({}_cG)$ is known as the ${}_cC$-deficiency of ${}_cG$. Since S can be empty $def\,({}_cG) \geq 0$.

using the fact that each $x \in A \cup C$ has $w(x)/h$ clones. The expression for the receiver's expected utility now follows from noting that it equals

$$h[|{}_cA \cup {}_cC| + def\left({}_cG\right)] = \sum_{x \in A \cup C} p(x)\, u(x) + \max_{S \subseteq C} \left[- \sum_{x \in S \cup N(S)} p(x)\, u(x) \right] \tag{8}$$

This completes the proof of the Corollary. ■

References

[1] Ali, S. Nageeb, Greg Lewis, and Shoshana Vasserman. 2023. "Voluntary Disclosure and Personalized Pricing," *Review of Economic Studies,*90(2): 538-571.

[2] Antić, Nemanja, Archishman Chakraborty and Rick Harbaugh. 2024. "Subversive Conversations," *Journal of Political Economy,* forthcoming.

[3] Auster, Sarah, Shiri Alon, Gabi Gayer and Stefania Minardi. 2024. "Persuasion with Limited Data: A Case-Based Approach," *Working Paper.*

[4] Battaglini, Marco. 2002. "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica*, 70(4): 1379–1401.

[5] Bergemann, Dirk and Stephen Morris, 2019. "Information design: a unified perspective," *Journal of Economic Literature*, 57(1): 44-–95.

[6] Crawford, Vincent P. and Joel Sobel, 1982. "Strategic Information Transmission," *Econometrica*, 1431–1451.

[7] Duan, R., and Su, H. H. 2012. "A scaling algorithm for maximum weight matching in bipartite graphs." In Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms: 1413-1424.

[8] Dasgupta, Sulagna, Ilia Krasikov and Rohit Lamba. 2022. "Hard Information Design," *Working Paper.*

[9] Dziuda, Wioletta. 2011. "Strategic Argumentation," *Journal of Economic Theory*, 146(4): 1362-1397.

[10] Dye, Ronald. 1985. "Disclosure of nonproprietary information. " *Journal of Accounting Research*, 23(1): 123–145.

[11] Glazer, Jacob and Ariel Rubinstein. 2004. "On Optimal Rules of Persuasion," *Econometrica*, 72(6): 1715–36.

[12] Glazer, Jacob and Ariel Rubinstein. 2006. "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach." *Theoretical Economics* 1(4): 395-–410.

[13] Grossman, Sanford J. 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics* 24(3): 461—83.

[14] Grossman, Sanford, and Oliver Hart. 1980. "Disclosure Laws and Takeover Bids," *Journal of Finance* 35(2): 323—34.

[15] Hall, Phillip., 1935. "On Representatives of Subsets," *Journal of the London Mathematical Society*, 10: 26–30.

[16] Hart, Sergiu, Ilan Kremer, and Motty Perry. 2017. "Evidence games: Truth and commitment, " *American Economic Review* 107(3): 690–713.

[17] Heller, I. and Tompkins, C. (1956), "An Extension of a Theorem of Dantzig's", in Kuhn, H.W.; Tucker, A.W. (eds.), *Linear Inequalities and Related Systems*, Annals of Mathematics Studies, vol. 38, Princeton (NJ): Princeton University Press, pp. 247–254.

[18] Kamenica, Emir. 2019. "Bayesian persuasion and information design, " *Annual Review of Economics*, 11(1), 249–272.

[19] Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian Persuasion," *American Economic Review*, 101(6): 2590–2615.

[20] Kartik, Navin. 2009. "Strategic communication with lying costs." *The Review of Economic Studies* 76(4): 1359–1395.

[21] Kolotilin, Anton, Roberto Corrao and Alexander Wolitzky. 2024. "Persuasion and Matching: Optimal Productive Transport," *Journal of Political Economy,* forthcoming.

[22] Kolotilin, Anton and Alexander Wolitzky. 2024. "Distributions of Posterior Quantiles via Matching," *Theoretical Economics,* forthcoming.

[23] Lovász, László, and Michael D. Plummer. 2009. *Matching Theory.* American Mathematical Society: 367.

[24] Madarasz, Kristof and Marek Pycia. 2023. “Information Choice: Cost over Content,” *Working Paper.*

[25] Matthews, Steven and Andrew Postlewaite. 1985. “Quality testing and disclosure,” *The RAND Journal of Economics* 16(3): 328–340.

[26] Milgrom, Paul 1981. “Good News and Bad News: Representation Theorems and Applications,” *Bell Journal of Economics* 12(2): 380-–91.

[27] Milgrom, Paul and John Roberts. 1986. “Relying on Information of Interested Parties,” *The RAND Journal of Economics* 17(1), 18-–32.

[28] Patil, Sanket and Yuval Salant. 2024. “Optimal sample sizes and statistical decision rules,” *Theoretical Economics*, 19(2), 583-604.

[29] Rappaport, Daniel. 2024. “Evidence and Skepticism in Verifiable Disclosure Games,” *Working Paper.*

[30] Rayo, Luis and Ilya Segal. 2010. “Optimal information disclosure,” *Journal of Political Economy*, 118(5): 949–987.

[31] Roesler, Anne-Katrin and Balázs Szentes. 2017. “Buyer-Optimal Learning and Monopoly Pricing,” *American Economic Review*, 107(7): 2072-80.

[32] Romer, Thomas, and Howard Rosenthal. “Political resource allocation, controlled agendas, and the status quo,” *Public Choice* 33(4): 27–43.

[33] Schrijver, A. 1998. *Theory of Linear and Integer Programming.* Wiley.

[34] Sher, Itai. 2011. “Credibility and Determinism in a Game of Persuasion,” *Games and Economic Behavior* 71(2):409–419.

[35] Seidmann, Daniel and Eyal Winter. 1997. “Strategic Information Transmission with Verifiable Messages,” *Econometrica*, 65(1), 163-–169.