Communication with Evidence in the Lab*

Jeanne Hagenbach[†] Eduardo Perez-Richet[‡]

May 5, 2018

Abstract

We study a class of sender-receiver disclosure games in the lab. Our experiment relies on a graphical representation of sender's incentives in these games, and permits partial disclosure. We use local and global properties of the incentive graph to explain behavior and performance of players across different games. Sender types whose interests are aligned with those of the receiver fully disclose, while other types use vague messages. Receivers take the evidence disclosed by senders into account, and perform better in games with an acyclic graph. Senders perform better in games with a cyclic graph. The data is largely consistent with a non-equilibrium model of strategic thinking based on the iterated elimination of obviously dominated strategies.

Keywords: Sender-receiver game; hard evidence; information disclosure; masquerade relation; skepticism; obvious dominance.

JEL classification: C72; C91; D82.

*We thank Larbi Alaoui, Leonie Baumann, Sourav Bhattacharya, Raicho Bojilov, Olivier Gossner, Guillaume Hollard, Asen Ivanov, Nicolas Jacquemet, Philippe Jehiel, Navin Kartik, Frédéric Koessler, Laurent Lamy, Aniol Llorente-Saguer, Daniel Martin, Sri Srikandan, Olivier Tercieux, Tristan Tomala and Arne Uhlendorff for interesting discussions and useful suggestions. We also thank seminar audiences at Ecole Polytechnique, Paris School of Economics, University Pompeu Fabra, University Autonoma de Barcelona, University of Essex, London School of Economics, HEC Lausanne and participants to the Cowles Foundation Economic Theory Conference and the 30th annual meeting of the European Economic Association. We are grateful to Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), to the French National Research Agency (ANR, project DYNAMITE) and to the iCODE (Institute for Control and Decision, research project of the Idex Paris-Saclay), for financial support.

†CNRS – Sciences Po and CEPR, e-mail: `jeanne.hagenbach@sciencespo.fr`

‡Sciences Po and CEPR, e-mail: `eduardo.perez@sciencespo.fr`

1 Introduction

In many economic interactions, agents provide hard evidence to influence the choice of others. Firms disclose product information to consumers. Policyholders provide medical check-ups to their insurers. At technology exhibitions, start-ups display their newest products to users and competitors. Job candidates make claims on CVs that employers can often cross-check on the Internet. In all these cases, the revealed information is verifiable and, yet, some of it can be voluntarily omitted, presented more or less advantageously or made imprecise. The objective of the informed party affects her incentive to shroud information, and the reading disclosed information by the uninformed party. In this paper, we report the results of an experiment that investigates how agents disclose and interpret evidence depending on the strategic context. To do so, we use a collection of simple sender-receiver disclosure games that is rich enough to vary the strategic context, while permitting a tractable common analysis. In general disclosure games, the predictive power of perfect Bayesian equilibrium is plagued by multiplicity, and fully revealing equilibria, while focal when they exist, can also be sustained by multiple strategies, and sometimes fail empirically. We show that a non-equilibrium approach based on the iterated elimination of obviously dominated strategies (Li, 2017) generates meaningful restrictions on behavior that are well matched by the data, and can provide guidance for equilibrium selection.

The empirical and experimental disclosure literatures focus on situations where the sender has monotonic incentives: She prefers to appear as having a "higher" informational type (better quality, greater ability, etc). By contrast, the games in our experiment span various incentives for the sender. This is important because non-monotonic incentives are omnipresent in economic situations. Consider, for example, the following situation. Company A and company B compete in the cell phone market. Company B is also present on the tablet market, where A is a potential entrant. A has just held its board meeting and it is known that they have reached a decision about entering the tablet market. This decision is private information but generates evidence, such as a new product, that can be disclosed or hidden. If A has decided to enter, it would prefer B to believe that it has not, so B does not engage into preventive R&D on a new tablet. If A has decided not to enter, it would prefer B to believe it has, so B devotes resources to the tablet market rather than to the cell phone market. In this case, the incentives of company A are cyclic: when committed to enter, A wants to appear as not entering and vice versa. More generally, cyclic incentives are natural in strategic situations with a Colonel Blotto flavor.[1] When having committed resources to one front, a player prefers the opponent to think she will attack on the other fronts. Non-monotonic incentives also arise naturally in situations where the receiver does not know the direction of the sender's bias,[2] or where the sender is addressing

[1] In Colonel Blotto games, two players simultaneously allocate their resources across battlefields. These games have been used to model research and development portfolio selection, political campaign resource allocation and auctions with simultaneous bidding.

[2] Consider a researcher advising a dean deciding between hiring an experimentalist or a theorist. The re-

a heterogeneous audience. Even in situations where the sender's bias is fixed and known, incentives to masquerade are not necessarily monotonic: The sender may want to appear as a slightly higher type but not as any higher type as in the Crawford and Sobel (1982) model. In the games we study, payoffs generate cyclic as well as acyclic but non-monotonic, and monotonic incentives.

Our experiment incorporates multiple games into a unified framework. These games are taken from a simple class of sender-receiver disclosure games in which the receiver essentially wants to identify the type of the sender, and the sender may have various incentives. We use a reduced representation of sender's incentives, taken from Hagenbach et al. (2014), as a graph whose nodes represent the different informational states (types) of the sender, and whose oriented arcs represent masquerading incentives. Thus, there is an arc from type t to type s if t wants to masquerade as s, that is, if the sender is better-off convincing the receiver that her information is s when it really is t. A sender type is categorized as *envious* if she has an incentive to masquerade as some other type, and *satisfied* otherwise. A satisfied sender's interests are aligned with the receiver's, whereas an envious sender's are not. We also distinguish between games with masquerading cycles, and those without cycles. Acyclic games have a fully revealing equilibrium (FRE), whereas cyclic games may or may not have one. In our experiment, the information of the sender is materialized by two colored cards that she observes, and can reveal as she sees fit. This simple design provides an original and natural way to allow for full disclosure (FD), partial disclosure (PD) and no disclosure (silence).

Our empirical analysis has two parts. In the first part, we document how players' performance and behavior depend on sender categories (envious or satisfied), the nature of received messages (FD, PD or silence) for receivers, and game categories (cyclic or acyclic). The main findings are as follows. First, receivers overwhelmingly take evidence into account both in beliefs and actions. That is, they report beliefs that do not put any weight on informational types that are ruled out by the evidence, and take actions that reflect consistent beliefs. Second, satisfied senders overwhelmingly fully disclose, whereas envious senders use vague messages (silence or PD). While intuitive, this result is not a necessary feature of perfect Bayesian equilibrium. Finally, while satisfied senders perform equally well in all games, envious senders tend to perform better, and receivers worse, in cyclic games (with or without FRE).

In the second part, we confront several theoretical models of strategic behavior in our class of games with the data. The models we consider are perfect Bayesian equilibrium (PBE), fully revealing equilibria (FRE), and a procedure of iterated elimination of obviously dominated strategies (IEODS) based on the notion of obvious dominance recently introduced in Li (2017).

---

searcher knows which candidate is best but also has a hidden bias. His type set is $\{tT, tE, eT, eE\}$ where tE corresponds to a bias towards the theorist but information that the experimentalist is best. Then type eT would like to be perceived as eE or tE, whereas type is tE would like to be perceived as type tT or eT. This induces a cycle between types tE and eT.

These concepts are defined, and their implications for our class of disclosure games derived in the theory section of this paper (Section 3). PBE is defined as usual in disclosure games. Its predictive power in our class of games is low as it induces few behavioral and payoff restrictions. Consequently, it is difficult to exclude as a model of players behavior even though we do find some discrepancies between beliefs reported by the subjects and empirical frequencies, and subjects do sometimes fail to optimize with respect to reported beliefs. FRE can only be applied to games for which it exists, but it induces significant restrictions on strategies, allowing us to develop multiple measures of FRE. We show that FRE matches more than 80 to 90% of the data according to these different metrics in the case of simple acyclic games. In the case of more complicated acyclic games, the fit is good, but we document significant departures from FRE. For cyclic games that have a FRE, however, FRE only matches a mall share of the data.

Finally, we consider IEODS. As in Li (2017), a strategy obviously dominates another if, under all possible scenarios following the first informational node at which they differ, the dominant one leads to a better payoff than the dominated one. IEODS is an appealing model of behavior for boundedly rational players as comparisons between strategies do not require contingent thinking. In the theory section, we show that, in the class of disclosure games we consider, the first three rounds of elimination can be decomposed in specific steps: In the first round, the receiver eliminates inconsistent strategies; in the second round, satisfied sender types eliminate vague messages and envious ones eliminate fully disclosing messages, and those whose consistent interpretations cannot lead to any benefit; in the third round, the receiver eliminates strategies that attribute a vague message to satisfied senders, or to envious senders who cannot gain from this message.

We suggest three ways of approaching the data based on the IEODS procedure. First, we consider IEODS as a solution concept. We show that more than 96% of our receiver observations, and 85% of our sender observations correspond to strategies that survive IEODS, showing that IEODS is a good predictor of behavior in our disclosure games. Of course, IEODS generally does not pin down a unique behavior in our games, but neither does any other solution concept we consider, and IEODS is in fact relatively effective at reducing the players' sets of strategies. Second, we consider IEODS as a model of strategic thinking in our disclosure games. Rounds of eliminations can be thought of as the levels of players in the spirit of the level k literature (for a survey, see Crawford et al., 2013). While it is true that players largely perform all rounds of elimination, the failure rates can become quite high for some steps of the elimination procedure, up to 33% for the fourth round which is the highest possible one in the games of the experiment. Furthermore, we use our decomposition of IEODS in steps within rounds to show that failure rates are neither homogeneous within rounds nor increasing in round (or level). Finally, we suggest using IEODS as a refinement of PBE by looking at PBE in strategies that survive IEODS (denoted IEODS+PBE). This is in line with the argument that, when there are multiple equilibria, selection is often the result of learning dynamics so that initial responses influence selected outcomes (Crawford, 1995; Camerer, 2003). We show that, compared to PBE, IEODS+PBE narrows down predictions around empirical averages in the payoff space.

2 Related Literature

While there is an extensive experimental literature[3] on cheap talk models (Crawford and Sobel, 1982), communication with evidence (Grossman, 1981; Milgrom, 1981) has given rise to a much thiner experimental literature. Early papers are mostly concerned with disclosure by a seller in a market environment. Forsythe et al. (1989) and King and Wallin (1991a) consider an informed seller, as in the models of Grossman (1981) and Milgrom (1981), and show that the unraveling principle works well, leading to a fully revealing equilibrium after some learning. King and Wallin (1991a) allow for partial disclosure through disclosure of given subsets of possible values.[4] They observe that departures from full disclosure arise as disclosure options to the sender increase and buyers have no knowledge of these options. In this case, receivers are insufficiently skeptical. Jin et al. (2015) run an experiment about quality disclosure in which they conclude that unraveling fails as lower types do not disclose, and receivers are too optimistic about sellers who choose not to disclose. However, they show that with feedback about the interactions players learn and reach equilibrium. Benndorf et al. (2015) study an environment with multiple senders, interpreted as workers, with the option to disclose their productivity at a cost. There is no strategic receiver, but payoffs are designed to see how far disclosure of higher types pushes disclosure of lower types. Their subjects under-reveal, especially when of lower productivity. In a recent experiment which we discuss further below, Schipper and Li (2018) study strategic thinking in games with monotonic incentives.

All these experiments consider environments with monotonic incentives for senders. We are interested in understanding the behavior of senders and receivers in situations with a variety of (monotonic or non-monotonic) incentives. Our paper is rooted in a body of theoretical works that extends Grossman (1981) and Milgrom (1981) to richer setups. Seidmann and Winter (1997) consider disclosure with preferences as in Crawford and Sobel (1982) that lead to non-monotonic incentives, and introduce the notion of worst-case types, which gives an operational definition of skepticism in non-monotonic environments.[5] Okuno-Fujiwara et al. (1990) gener-

[3]See for example Dickhaut et al. (1995); Blume et al. (1998, 2002); Cai and Wang (2006); Wang et al. (2010) for sender-receiver games, Battaglini and Makarov (2014) for a setup with multiple receivers, Vespa and Wilson (2016) for multiple senders, and Lai et al. (2015) for multidimensional cheap talk.

[4]King and Wallin (1991b) and Dickhaut et al. (2003) also allow for partial disclosure in a model which adds uncertainty about the seller's information as in Dye (1985).

[5]Mathis (2008) looks at conditions on the set of evidence that make it possible to extend the result of Seidmann and Winter (1997) to partial provability.

alize Milgrom (1981) to environments in which players exchange verifiable information before playing an incomplete information game, provide conditions for fully revealing equilibria to exist and examples in which full revelation is not an equilibrium. Van Zandt and Vives (2007) perform a similar exercise in supermodular games. Giovannoni and Seidmann (2007) show that there is a unique pooling equilibrium in one-dimensional action games if the the bias of the sender is strongly centripetal in both directions. Hagenbach et al. (2014) provide results that encompass all previous full revelation results in this literature, and introduce a new approach based on properties of the masquerade relation, i.e. the incentive graphs of informed players. In a recent theoretical contribution, Rappoport (2017) studies games with monotonic incentives when the evidence structure does not allow for fully revealing equilibria, and relates the degree of equilibrium skepticism to the amount of evidence implicit in the type distribution.

This paper is related to the literature on strategic thinking, and in particular to the level-k (Nagel, 1995; Stahl and Wilson, 1994, 1995; Costa-Gomes et al., 2001; Costa-Gomes and Crawford, 2006) and cognitive hierarchy (Camerer et al., 2004) literatures. Crawford (2003) is the first to apply level-k models to communication in a context in which intentions rather than private information are signalled. However, our strategic thinking model is based on the notion of obvious dominance introduced by Li (2017) to study obviously strategy-proof mechanism design.[6] IEODS is naturally related to the rationalizability literature (Bernheim, 1984; Pearce, 1984).

The most closely related experimental paper is Schipper and Li (2018). They study strategic thinking in the form of cautious rationalizability in disclosure games with monotonic incentives and complete unraveling, so that all PBE are fully revealing. Even though we used the first steps of an iterated elimination procedure in the earlier version of this paper, Schipper and Li (2018) are the first to systematically apply such an approach in an experiment on disclosure games. The disclosure game they study, however, does not belong to the class of games we use in our experiment. Their cautious rationalizability solution concept is equivalent to iterated elimination of weakly dominated strategies on the normal form. This solution concept and IEODS make exactly the same predictions in all games of both experiments[7], and appear to explain the data quite well in both cases. So we view our results as complimentary in highlighting the usefulness of the procedure in explaining how subjects play disclosure games in the lab.

Finally, the empirical literature on disclosure has also focused on environments with monotonic masquerading incentives. Dranove and Jin (2010) give a nice overview of the literature on quality disclosure. Brown et al. (2012) look at cold openings in the movie industry, and observe that consumers fail to correctly infer the average quality of movies that are withheld

[6] See also the burgeoning literature on this topic (Ashlagi and Gonczarowski, 2016; Bade and Gonczarowski, 2017; Pycia and Troyan, 2018; Troyan, 2016; Zhang and Levin, 2015).

[7] However, we can provide examples of disclosure games in which they make different predictions.

from critics. Luca and Smith (2015) study disclosure of international rankings by MBAs and show how business schools manage to strategically present information in favorable ways.

3 Theoretical Background

In this section, we provide the necessary theoretical background for our analysis. We first introduce a class of simple sender-receiver disclosure games that comprises all games in our experiment. Then, we define the solution concepts that we will discuss as models of behavior for our experiment, and describe how they operate on our class of games. All proofs for this section are in Appendix A.

A class of disclosure games. We consider sender-receiver disclosure games that can be described in the following framework. The sender's type is drawn from a finite set T according to a distribution $p(\cdot)$ with full support. The set of messages, or pieces of evidence, $M(t)$ to which she has access is contingent on her type. Hence a message m in $M = \cup_{t\in T} M(t)$ certifies that the type of the sender lies in the subset $\mathcal{E}(m) = \{t \in T : m \in M(t)\} \subseteq T$. We call this correspondence $\mathcal{E}$ the *evidence structure* of the game. A subset of types $T' \subseteq T$ is *certifiable* if there exists a message m such that $\mathcal{E}(m) = T'$. We refer to any message that certifies T as *silence* since it provides no information. If the sender certifies a singleton $\{t\}$, we say that her message is *fully disclosing*. We refer to any message that is not fully disclosing as being *vague*. If a message is vague but is not silence, we say that it is *partially disclosing*.

Assumption 1. *Each sender type can at least remain silent or fully disclose.*

The sender first chooses a message m to send, and the receiver then chooses an action a from a finite set A after observing m. The payoffs of the sender and the receiver are given by $u_S(a,t)$ and $u_R(a,t)$. We focus on games such that, for every t, the utility function of the receiver is uniquely maximized by a single action $a^*(t)$. Our second assumption, summarized below, is that the receiver's payoff is positive if the receiver plays her optimal action and 0 otherwise. In the remainder of the paper, we refer to $a^*(t)$ as the *optimal action* (of the receiver). We denote $a^*(T')$ the set of actions that are optimal for some type t in T'.

Assumption 2. *For every t, the receiver optimal action $a^*(t)$ is the unique action such that $u_R\big(a^*(t),t\big) > 0$ and, for every $a \neq a^*(t)$, $u_R(a,t) = 0$.*

To state our last assumption, we first need to introduce some additional definitions.

Masquerade Relation. In disclosure games, it is useful to introduce the *masquerade relation* to describe sender incentives as in Hagenbach et al. (2014). We say that type t wants to masquerade as type $t' \neq t$, and write $t \to t'$, if

$$u_S\big(a^*(t'), t\big) > u_S(a^*(t), t).$$

This relation is best pictured as a graph whose vertices are the elements of T, and where there is an oriented arc between t and t' if $t \to t'$. Examples from the experiment are provided in Table 2.

We say that type t is *satisfied* if there is no other type t' that t wants to masquerade as, and denote the set of satisfied types by Sat. We refer to non satisfied types as *envious*, and denote the set of envious types by Env. We say that a sequence of types $t^1, \cdots, t^k$ forms a *masquerading cycle* if, for every $1 \leq \ell \leq k$, $t^\ell \to t^{\ell+1}$, where by convention $t^{k+1} = t^1$. A masquerade relation is cyclic if it admits a masquerading cycle, and acyclic otherwise. By extension, we talk about cyclic and acyclic games.

Next, we introduce our last assumption. It says, first, that envious types are indifferent between any type they do not want to masquerade as and themselves, and, second, that satisfied types strictly prefer being identified to leading the receiver to any other action.

Assumption 3. *If t is envious, and $t \not\to t'$, then $u_S(a^*(t'), t) = u_S(a^*(t), t)$. If t is satisfied, then, for any $t' \neq t$, $u_S(a^*(t), t) > u_S(a^*(t'), t)$.*

In the remainder of this section, we focus on the class of disclosure games that satisfy our three assumptions. This class includes all games in the experiment.

Examples from the Experiment. In the games of the experiment, a sender observes two cards, each of them either yellow or blue, and can decide which cards she shows the receiver. The type set of the sender is therefore $T = \{YY, YB, BB\}$, the set of messages is $M = \{\emptyset, Y, B, YY, YB, BB\}$. The evidence structure is summarized by Figure 1 which represents all certifiable subsets. Among all subset of type, only $\{YY, BB\}$ is not certifiable. The action set of the receiver is $A = \{a, b, c\}$. The payoff matrix of the receiver is given by Table 1 in all our experiments, so optimal actions always are $a^*(YY) = a$, $a^*(YB) = b$ and $a^*(BB) = c$. In contrast, the payoff of the sender varies across sessions, giving rise to different masquerade relations. Table 2 shows three examples of sender matrices and associated masquerade relations from the experiment. In all our games, payoffs only take value 0 or 3.

Worst-case types, skepticism and acyclicity. The notion of *worst-case type*, first introduced by Seidmann and Winter (1997), captures the idea of skepticism. A worst-case type of a subset $T' \subseteq T$ is a type $t' \in T'$ that no other type $t \in T'$ wants to masquerade as. Graphically, it is a vertex in T' with no incoming arc from other types in T'. The set of worst-case types of T' is denoted by $\text{wct}(T') = \{t' \in T' : t \to t' \Rightarrow t \notin T'\}$. By extension, we write $\text{wct}(m)$

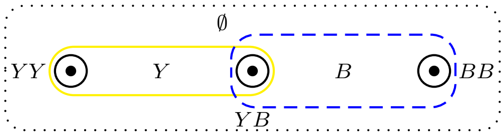
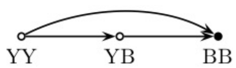


Figure 1: *Evidence structure in the experiment.*

	a	b	c
YY	3	0	0
YB	0	3	0
BB	0	0	3

Table 1: *Receiver payoff matrix in the experiment.*

Game 1: monotonic

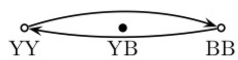


	a	b	c
YY	0	3	3
YB	0	0	3
BB	0	0	3

Game 2: double cycle

	a	b	c
YY	0	3	0
YB	3	0	3
BB	0	3	0

Game 3: long cycle



	a	b	c
YY	0	0	3
YB	0	3	0
BB	3	0	0

Table 2: *Sender payoff matrices and masquerade relations for three games from the experiment. Empty nodes are for envious types, filled nodes for satisfied types.*

for $\mathrm{wct}(\mathcal{E}(m))$, and speak about the worst-case type of a message. For example, in Game 1 of Table 2, type YB is a worst-case type of message Y but not of the the silent message; in Game 2, the only messages with worst-case type are the fully disclosing ones; in Game 3, type YB is the only worst-case type of the silent message.

Worst-case types are key to full revelation as they provide an operational definition of skepticism. Attributing vague messages to worst-case types allows the receiver not to reward obfuscation attempts while remaining consistent with the evidence. However, worst-case types need not exist. In Game 2 of Figure 2, for example, message Y has no worst-case type because types YY and YB form a masquerading cycle. The existence of worst-case types, and hence the possibility of skepticism, is crucially related to the absence of masquerading cycles, as shown by the following lemma.

Lemma 1. *Every subset of types $T' \subseteq T$ admits a worst-case type if and only if the masquerade relation is acyclic.*

However, skepticism remains possible in the presence of cycles in the graph if some subsets of types cannot be certified. This is the case in Game 3 of Table 2. If the sender could certify the subset $\{YY, BB\}$, then the masquerading cycle between these two types would make it impossible to find a worst-case type but, in the absence of such a message, skepticism is always possible.

Strategies and Equilibrium. A strategy of the sender is a mapping $\mu(\cdot)$ from types T to messages M such that $\mu(t) \in M(t)$. Let $\mathcal{M}^0$ be the set of sender strategies. A strategy of the receiver is a mapping $\alpha(\cdot)$ from messages M to actions A. Let $\mathcal{A}^0$ be the set of receiver strategies. Denote by $\beta_m \in \Delta(T)$ the belief of the receiver following message m, and let $\beta = \big(\beta_m\big)_{m\in M}$ be her belief system. Our equilibrium concept is perfect Bayesian equilibrium (PBE). A triple (μ, α, β) forms a PBE if, for every $m \in M$ and every $t \in T$, it satisfies

$$\alpha(m) \in \arg\max_a \sum_t u_R(a,t)\beta_m(t) \qquad \text{(receiver optimality)}$$

$$\mu(t) \in \arg\max_{m\in M(t)} u_S\big(\alpha(m), t\big) \qquad \text{(sender optimality)}$$

$$\mu(t) = m \Rightarrow \beta_m(t) = \frac{p(t)}{\sum_{t':\mu(t')=m} p(t')} \qquad \text{(Bayesian rationality)}$$

$$\beta_m\big(\mathcal{E}(m)\big) = 1 \qquad \text{(evidence consistency)}$$

For the remainder of the paper, it is useful to extend the idea of consistency to receiver strategies. We say that a strategy α of the receiver is a *consistent strategy* if for every message $m \in M$, we have $\alpha(m) \in a^*\big(\mathcal{E}(m)\big)$. By Assumption 2, this is equivalent to $\alpha(m)$ being optimal for some consistent belief β_m. Otherwise, we say that α is inconsistent. Similarly, we say that action a is *consistent* with message m if $a = a^*(t)$ for some type $t \in \mathcal{E}(m)$.

A belief system is *skeptical* if every belief is concentrated on the set of worst-case-types of the corresponding message. That is if, for every $m \in M$ such that $\text{wct}(m) \neq \emptyset$, $\beta_m\big(\text{wct}(m)\big) = 1$. It is also useful to extend the notion of skepticism from beliefs to strategies. We say that a receiver *acts skeptically* if the action she picks following a message is the optimal action associated with a worst-case type for this message, that is if, for every $m \in M$ such that $\text{wct}(m) \neq \emptyset$, $\alpha(m) \in a^*\big(\text{wct}(m)\big)$.

Fully revealing equilibria. We say that a PBE is a fully revealing equilibrium (FRE) if the strategy of the sender is separating, that is, $\mu(t) \neq \mu(t')$ when $t \neq t'$. Skepticism is necessary and sufficient for the existence of FRE. Indeed, we can extend results from Hagenbach et al. (2014) to characterize FRE as follows in the class of games we consider.[8]

[8] In Appendix A, we extend this result to mixed strategy equilibria.

Proposition 1. *In our class of games, the following statements are equivalent:*

(i) There exists a FRE (μ, α, β).

(ii) The receiver strategy $\alpha(\cdot)$ is such that, for every m, $\alpha(m) \in a^*\big(\mathrm{wct}(m)\big)$, and the sender strategy $\mu(\cdot)$ is separating and such that, for every t, $t \in \mathrm{wct}\big(\mu(t)\big)$.

That is, in FRE, the receiver acts skeptically and the sender only sends messages for which she is a worst-case type. In particular, for FRE to exist, it must be that every sender type has access to a distinct message for which she is a worst-case type. This requirement is satisfied in our class of games where full disclosure is always possible. Furthermore, for FRE to exist, every certifiable subset of types must have a skeptical interpretation (a worst-case type). It follows from Lemma 1 that all acyclic games have a FRE. Cyclic games can also have a FRE, as it is the case for Game 3 in Table 2 (but not for Game 2).

Iterated Elimination of Obviously Dominated Strategies (IEODS). We propose a heuristic for strategic reasoning in disclosure games based on the notion of obvious dominance of Li (2017). Obvious dominance compares two strategies according to the worst and best case scenarios they generate starting from the first information set at which they diverge. We next state the definition of obvious dominance for sender and receiver strategies. Because the idea is to iteratively eliminate obviously dominated strategies, we include in the definition a set of available strategies for the other party.

As a preliminary, consider a set of possible strategies $\mathcal{M}$ for the sender. Then $\mathcal{M}(T) = \{\mu(t) : t \in T, \mu \in \mathcal{M}\}$ is the set of messages that can be received given $\mathcal{M}$, and $\mathcal{I}(m|\mathcal{M}) = \{t \in T : \exists \mu \in \mathcal{M},\, \mu(t) = m\}$ is the set of possible interpretations of m under $\mathcal{M}$. For example, the set of possible interpretations of m under the full set of sender strategies is exactly the set of consistent interpretations: $\mathcal{I}(m|\mathcal{M}^0) = \mathcal{E}(m)$.

Definition 1 (Obvious Dominance – Receiver)**.** *A receiver strategy α obviously dominates α' given a subset $\mathcal{M}$ of sender strategies, denoted by $\alpha \rhd_{\mathcal{M}} \alpha'$, if, for every $m \in \mathcal{M}(T)$ such that $\alpha(m) \neq \alpha'(m)$,*

$$\min_{t \in \mathcal{I}(m|\mathcal{M})} \left\{u_R\big(\alpha(m), t\big)\right\} \geq \max_{t \in \mathcal{I}(m|\mathcal{M})} \left\{u_R\big(\alpha'(m), t\big)\right\},$$

and, in case of equality, either $\left\{u_R\big(\alpha(m), t\big) : t \in \mathcal{I}(m|\mathcal{M})\right\}$ or $\left\{u_R\big(\alpha'(m), t\big) : t \in \mathcal{I}(m|\mathcal{M})\right\}$ is not a singleton.

Elimination of obviously dominated strategies for a receiver in our games requires that a receiver facing message m considers all types of senders that might have sent her message m given $\mathcal{M}$, and selects actions according to the best and worst payoff they can generate in all these possible scenarios. Consider Game 1 of Table 2, for example, and let $\mathcal{M} = \mathcal{M}^0$ so that all sender strategies are available. Suppose that the message is Y. This message can only have been sent by types YY and YB. Hence choosing a delivers two possible payoffs: 3 or 0. Similarly, b delivers either 3 or 0. However, c can only deliver 0 because type BB is ruled out. Therefore playing c after Y is obviously dominated by both a and b.

Definition 2 (Obvious Dominance –Sender)**.** *A sender strategy μ obviously dominates μ' given a subset $\mathcal{A}$ of receiver strategies, denoted by $\mu \rhd_{\mathcal{A}} \mu'$, if, for every $t \in T$ such that $\mu(t) \neq \mu'(t)$,*

$$\min_{\alpha \in \mathcal{A}} \left\{ u_S\big(\alpha(\mu(t)), t\big) \right\} \geq \max_{\alpha \in \mathcal{A}} \left\{ u_S\big(\alpha(\mu'(t)), t\big) \right\},$$

and, in case of equality, either $\left\{ u_S\big(\alpha(\mu(t)), t\big) : \alpha \in \mathcal{A} \right\}$ or $\left\{ u_S\big(\alpha(\mu'(t)), t\big) : \alpha \in \mathcal{A} \right\}$ is not a singleton.

Elimination of obviously dominated strategies for a sender in our games therefore requires that a sender of type t considers all receiver actions that each of her messages can generate. Consider Game 1 in Table 2, and suppose that receiver strategies are consistent. Then consider a sender of type YY. If she sends the silent message $\emptyset$, consistent actions are a,b and c so possible payoffs are 0 and 3. If she sends Y, consistent actions are a and b, and possible payoffs are again 0 and 3. By contrast, if she fully discloses by choosing message YY, the only consistent action is a, and the only possible payoff 0. Therefore full disclosure is obviously dominated by both Y and $\emptyset$.

The iterated deletion procedure can be defined as follows. For every $k \geq 0$,

$$\mathcal{A}^{k+1} = \left\{ \alpha \in \mathcal{A}^k : \forall \alpha' \in \mathcal{A}^k, \alpha' \not\rhd_{\mathcal{A}^k} \alpha \right\},$$

and

$$\mathcal{M}^{k+1} = \left\{ \mu \in \mathcal{M}^k : \forall \mu' \in \mathcal{M}^k, \mu' \not\rhd_{\mathcal{M}^k} \mu \right\}.$$

It is easy to see that this procedure terminates in a finite number of steps and that nonempty sets of sender and receiver strategies survive this process. We can think of k as indexing rounds of elimination or the level of a player who performs exactly the first k rounds of elimination.

To illustrate this procedure, Table 3 shows how IEODS operates on our Game 1 of Table 2 with the monotonic masquerade[9]. Next, we follow this procedure step by step. At the first round of elimination, only the receiver is active. Following each message of the sender, any consistent choice of action may hit it right, hence it generates a payoff of 3 in the best-case scenario, and there may also be a worst-case scenario in which it gets it wrong and generates a null payoff. An inconsistent choice, by contrast, leads to a null payoff for sure. Therefore all inconsistent choices are eliminated, and in $\mathcal{A}^1$ remain all consistent strategies. In the second round, the sender draws the conclusions from consistence (and the receiver is inactive). If

[9]We adopt this nice compact representation of rounds of elimination from Schipper and Li (2018).

Round	Player	Type	$\emptyset$	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
		YY	✓	✓	•	× (2.2)	•	•
2	S	YB	✓	× (2.3)	✓	•	× (2.2)	•
		$\boxed{BB}$	× (2.1)	•	× (2.1)	•	•	✓
3	R		a, b (3.1)	a (3.2)	b (3.1)	a	b	c
		YY	✓	× (4)	•	×	•	•
4	S	YB	✓	×	✓	•	×	•
		$\boxed{BB}$	×	•	×	•	•	✓

Table 3: *Iterated Elimination procedure for the monotonic masquerade. For receiver, we write surviving actions following each message (column) at the corresponding round. For each type of sender, • denotes a message that is not available, × an eliminated message, and ✓ a surviving message. Type BB is framed to denote that it is satisfied. Numbers in parenthesis indicate the step of reasoning that leads to keep/eliminate the corresponding action(s)/message.*

she is of the satisfied type, BB, she can obtain 3 for sure by fully disclosing, whereas any other message may lead to at least one other consistent actions, and therefore has a worst-case outcome of 0. Therefore, full disclosure is the only undominated message. If she is of an envious type, YY or YB, then full disclosure leads to 0 for sure, whereas silence can lead to a consistent receiver action that pays off 3. Therefore full disclosure is dominated. Among vague messages, envious types must also eliminate messages that cannot lead to a positive payoff action. Here, it is dominated for type YB to pick Y as all receiver actions that are consistent with Y give a null payoff to YB. In the third round, only the receiver is active, and she updates her possible interpretations of each message based on the former round of elimination by the sender. Silence can only be used by YY and YB since BB is satisfied and should fully disclose, so action c is eliminated. Message Y can only be sent by YY since only YB could possibly send it but prefers the dominating strategy of being silent. So b is eliminated following Y. B can only be sent by YB because BB is satisfied, so only b survives. Finally, in the last round, type YY sender concludes that she should not use Y as the receiver would unmask her (0 in every possible scenario), whereas she can get 0 or 3 by playing silence.

The reasonings we have illustrated for Game 1 are common to all games in our class, and we can in fact decompose the first three rounds of the iterated deletion procedure in several independent steps within rounds as follows. Some of these steps are very natural. For example, receiver is consistent, satisfied types fully disclose, envious types do not.

Round 1. *$\mathcal{M}^1 = \mathcal{M}^0$ and $\mathcal{A}^1$ is the set of consistent receiver strategies.*

To describe the next round, we introduce the notion of *advantageous* obfuscation. A vague message m is advantageous for an envious type t if there exists a consistent interpretation of this message $t' \in \mathcal{E}(m)$ such that $t \to t'$. For instance, in Game 1, it is advantageous for the sender of type YB to send message B but not to send message Y. Note that remaining silent is always advantageous. Intuitively, an envious sender who understands that the receiver is consistent should only send advantageous messages.

Round 2. *$\mathcal{A}^2 = \mathcal{A}^1$, and $\mathcal{M}^2$ is the set of sender strategies in $\mathcal{M}^1$ that satisfy all of the following:*

(2.1) *If t is satisfied, then $\mu(t)$ is fully disclosing.*

(2.2) *If t is envious, then $\mu(t)$ is vague.*

(2.3) *If t is envious, then $\mu(t)$ is advantageous for t.*

Round 3. *$\mathcal{M}^3 = \mathcal{M}^2$ and $\mathcal{A}^3$ is the set of receiver strategies in $\mathcal{A}^2$ that satisfy all of the following:*

(3.1) *If m is a vague message in $\mathcal{M}^2(T)$, then $\alpha(m) \notin a^*(Sat)$.*

(3.2) *If m is a vague message in $\mathcal{M}^2(T)$, and m is not advantageous for t, then $\alpha(m) \neq a^*(t)$.*

4 Experimental Design

Our experiment consists of a collection of sender-receiver games. In each of these games, the sender observes two cards, each of which is either yellow, Y, or blue, B. She chooses which cards to reveal, if any. The receiver observes the cards revealed by the sender, and then takes a decision a, b or c.

Subjects were first given written instructions and had to fill a comprehension test.[10] Then they played 20 or 30 rounds of a fixed sender-receiver game preceded by two rounds of practice. At the beginning of a session, each subject was randomly assigned a role, labeled as *player 1* (sender) or *player 2* (receiver), for the whole session. At the end of each round, both players in every pair were given full feedback (cards of the sender, action of the receiver and payoffs) about their interaction. Senders and receivers were then randomly rematched for the next round.[11]

The 21 experimental sessions were conducted at the Experimental Laboratory in Ecole Polytechnique, France, with a total of 354 subjects who were mostly science students from Ecole Polytechnique and ENSTA, two French engineering schools, and some administrators at Ecole Polytechnique. The final payment of each subject was the sum of a 5 euros show-up fee and a variable payment based on her total score in the games she played. The average variable

[10] The english translation of the instructions and comprehension test, corrected orally by the experimentalist, are given in Appendix G. All decisions were taken on computers and screenshots are provided in Appendix F.

[11] We did not use a perfect stranger matching process, so rematch was in principle allowed. On average, over all sessions, a given pair of subjects was formed 3.9 times. Each combination of two subjects and two cards was formed 2.3 times on average.

payment per subject was 15 euros. In some sessions, subjects had to go through an additional belief-elicitation task and could earn up to 4 more euros. For each session, Table 5 details the number of rounds played, the number of subjects and the value of a point in euros (Exchange Rate). The last column of the Table states whether beliefs were elicited or not.

Information and messages. In our design, the type set is given by the feasible combinations of cards $T = \{YY, YB, BB\}$. The three combinations were drawn with equal probabilities by the computer and independently in each round. The sender was allowed to reveal both cards, none or any one of her choice. The corresponding message space is $M = \{\emptyset, Y, B, YY, YB, BB\}$, with obvious interpretations. The evidence structure is depicted in Figure 1 in Section 3. A subject *fully discloses*, abbreviated as FD, if she reveals both cards, *partially discloses*, abbreviated as PD, if she reveals one card, and remains *silent* if she reveals no card. We also refer to silence and PD as *vague messages*.

Payoffs and masquerading graphs. In each session, subjects could at all time see two matrices showing sender and receiver payoffs in the game they were playing. The payoff matrix of the receiver was the same across all sessions, whereas the payoff matrix of the sender was the unique part of the design that varied across sessions. Examples of the matrices we used are given by Table 1 and Table 2 in Section 3. To each game corresponds a particular masquerade relation, and Figure 2 shows all the masquerading graphs that we used.[12] Table 5 shows the masquerading graphs and sender payoff matrices associated to each of the 21 experimental sessions. Note that the same sender payoff matrix was sometimes used in several sessions.

Over all these sessions, we collected data for 3690 one-shot sender-receiver games. Table 4 shows the distribution of observations across acyclic and cyclic games, envious and satisfied senders.

	Envious	Satisfied	Total
Acyclic	1039	821	1860
Cyclic	1372	458	1830
Total	2411	1279	3690

Table 4: *Distribution of the data across game and sender categories.*

Beyond cyclic and acyclic, it is useful to classify our games into four categories: simple acyclic games with a single envious type (Acyclic 1: 580 observations), acyclic games with

[12] A systematic exploration of all possible graphs on three types would have required many more sessions. In total, there are $2^6 = 64$ possible graphs, and only 36 if we take symmetries into account (that is, make YY and BB equivalent, and keep YB distinct as required by the structure of the set of evidence). When selecting the graphs for our experiments, we aimed at having a similar number of acyclic and cyclic graphs. We also wanted graphs which exhibited various number of envious types and arrows.

two envious types (Acyclic 2: 1280 observations), cyclic games with a FRE (Cyclic FRE: 650 observations), and other cyclic games (Cyclic other: 1180 observations). For each session, Table 5 reports the category of the game used, and whether this game has a FRE.

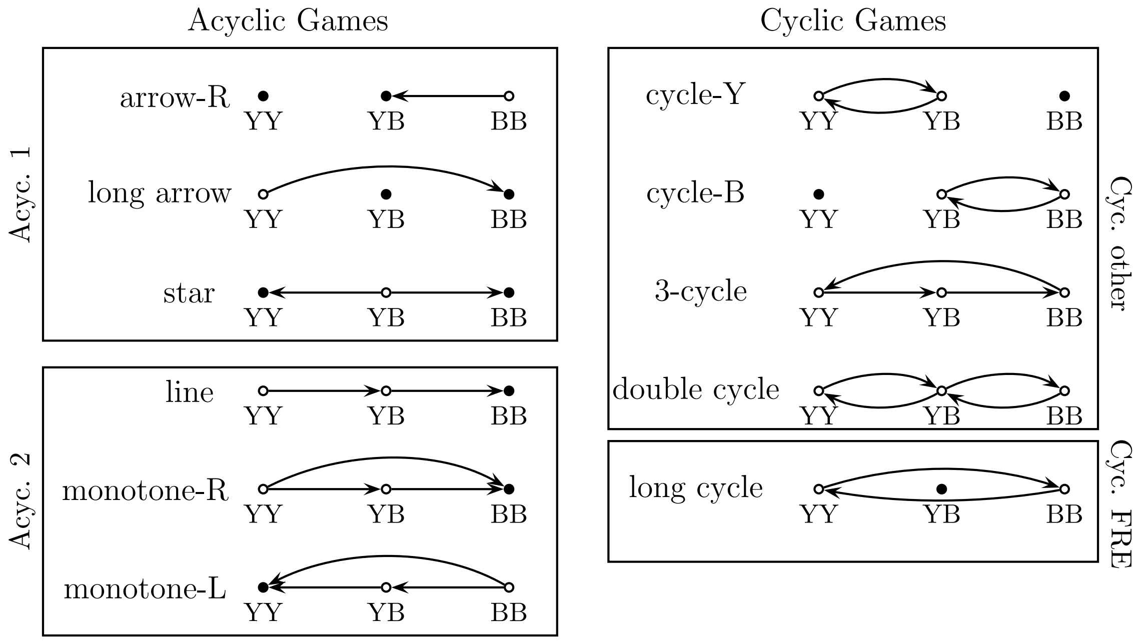


Figure 2: *Masquerading graphs and game categories in the experiment. Empty nodes are for envious types, filled nodes for satisfied types.*

Belief elicitation. We elicited the beliefs of subjects in 8 of the 21 sessions (4 with acyclic games, 4 with cyclic games, 136 subjects in total). In each of these sessions, we elicited beliefs once, after the final round of play. Each sender was asked to indicate the frequency with which she thought that each action was played after each possible message. Each receiver was asked to indicate the frequency with which she thought that each card combination had been the true one given each possible message.[13] We then compared the reported frequencies to the empirical frequencies to remunerate subjects.[14] A subject whose beliefs exactly matched the empirical frequency could earn 4 euros. Appendix D shows average reported beliefs and true empirical frequencies.

[13] Appendix F reports the tables used for beliefs elicitation, that were filled in computers.

[14] We used the quadratic scoring rule $s(f, f^e) = 16\left(1 - \frac{1}{18}\sum_{k=1}^{18} \delta_k \left(\frac{f_k - f_k^e}{100}\right)^2\right)$, with f_k and f_k^e the believed and empirical frequencies in square k of the elicitation tables. If a message had never occurred in a session, the indicator δ_k was set to 0 for the squares corresponding to this message. Under this rule, subjects maximized their expected payoff by reporting their true subjective beliefs.

Session	Masquerade	Sender Payoff	Game Category	FRE	Rounds	Subjects	Exch. Rate	Beliefs
1	YY YB BB	$\begin{pmatrix} 3&0&0 \\ 0&3&0 \\ 0&3&0 \end{pmatrix}$	Acyclic 1	Yes	20	18	0.20	No
2	YY YB BB	$\begin{pmatrix} 0&0&3 \\ 0&3&0 \\ 0&0&3 \end{pmatrix}$	Acyclic 1	Yes	20	20	0.25	No
3	YY YB BB	$\begin{pmatrix} 3&0&0 \\ 3&0&3 \\ 0&0&3 \end{pmatrix}$	Acyclic 1	Yes	20	20	0.20	No
4	YY YB BB	$\begin{pmatrix} 0&3&0 \\ 0&0&3 \\ 0&0&3 \end{pmatrix}$	Acyclic 2	Yes	20	20	0.20	No
5	YY YB BB	$\begin{pmatrix} 0&3&0 \\ 0&0&3 \\ 0&0&3 \end{pmatrix}$	Acyclic 2	Yes	20	20	0.25	Yes
6	YY YB BB	$\begin{pmatrix} 0&3&0 \\ 0&0&3 \\ 0&0&3 \end{pmatrix}$	Acyclic 2	Yes	20	20	0.25	Yes
7	YY YB BB	$\begin{pmatrix} 0&3&3 \\ 0&0&3 \\ 0&0&3 \end{pmatrix}$	Acyclic 2	Yes	20	18	0.25	No
8	YY YB BB	$\begin{pmatrix} 0&3&3 \\ 0&0&3 \\ 0&0&3 \end{pmatrix}$	Acyclic 2	Yes	20	20	0.25	Yes
9	YY YB BB	$\begin{pmatrix} 0&3&3 \\ 0&0&3 \\ 0&0&3 \end{pmatrix}$	Acyclic 2	Yes	20	14	0.25	Yes
10	YY YB BB	$\begin{pmatrix} 3&0&0 \\ 3&0&0 \\ 3&3&0 \end{pmatrix}$	Acyclic 2	Yes	20	16	0.25	No
11	YY YB BB	$\begin{pmatrix} 0&3&0 \\ 3&0&0 \\ 0&0&3 \end{pmatrix}$	Cyclic other	No	20	20	0.66	No
12	YY YB BB	$\begin{pmatrix} 0&3&0 \\ 3&0&0 \\ 0&0&3 \end{pmatrix}$	Cyclic other	No	20	18	0.25	Yes
13	YY YB BB	$\begin{pmatrix} 3&0&0 \\ 0&0&3 \\ 0&3&0 \end{pmatrix}$	Cyclic other	No	30	10	0.20	No
14	YY YB BB	$\begin{pmatrix} 3&0&0 \\ 0&0&3 \\ 0&3&0 \end{pmatrix}$	Cyclic other	No	20	14	0.25	Yes
15	YY YB BB	$\begin{pmatrix} 0&3&0 \\ 3&0&3 \\ 0&3&0 \end{pmatrix}$	Cyclic other	No	20	20	0.25	No
16	YY YB BB	$\begin{pmatrix} 0&3&0 \\ 0&0&3 \\ 3&0&0 \end{pmatrix}$	Cyclic other	No	30	10	0.20	No
17	YY YB BB	$\begin{pmatrix} 0&3&0 \\ 0&0&3 \\ 3&0&0 \end{pmatrix}$	Cyclic other	No	20	16	0.25	No
18	YY YB BB	$\begin{pmatrix} 0&0&3 \\ 0&3&0 \\ 3&0&0 \end{pmatrix}$	Cyclic FRE	Yes	30	10	0.20	No
19	YY YB BB	$\begin{pmatrix} 0&0&3 \\ 0&3&0 \\ 3&0&0 \end{pmatrix}$	Cyclic FRE	Yes	20	20	0.25	No
20	YY YB BB	$\begin{pmatrix} 0&0&3 \\ 0&3&0 \\ 3&0&0 \end{pmatrix}$	Cyclic FRE	Yes	20	16	0.25	Yes
21	YY YB BB	$\begin{pmatrix} 0&0&3 \\ 0&3&0 \\ 3&0&0 \end{pmatrix}$	Cyclic FRE	Yes	20	14	0.25	Yes

Table 5: *Summary of sessions.*

5 Theoretical Predictions and Empirical Strategy

Our empirical analysis is divided in two parts. The first part (Section 6) is descriptive, and seeks to describe the behavior and performance of players according to game, sender type, and message categories. The second part (Section 7) compares the different models of strategic behavior introduced in Section 3. Before presenting our results, we discuss the predictions of these models for the games of the experiment, our empirical strategy and hypotheses for this second part.

PBE. All of our games have multiple PBE and, more generally, PBE makes few testable predictions that can be checked across all of our games. While this makes it difficult to assess the extent to which PBE explains the data, we can provide a comparison of empirical payoffs with the range of PBE theoretical payoffs for each of our games. Next, PBE assumes receiver consistency, which also corresponds to the first round of elimination in IEODS. In our games, this implies that receivers act consistently in the sense that they never pick the action associated with a type of the sender which is ruled out by the evidence they have received. We check to which extent the data satisfies this assumption. Finally, we also try to assess the extent to which players best-respond to their beliefs, and the extent to which they form accurate beliefs by looking at the beliefs we elicited in some sessions. Overall, the design of our experiment and the multiplicity of equilibria make it difficult to test for PBE, so these results are suggestive at best. The other solution concepts, by contrast, provide more verifiable implications.

FRE. It seems natural to focus on FRE as a prediction when one exists. However, this does not eliminate multiplicity, as an envious sender could send multiple vague messages, or fully disclose in a FRE. But FRE still make some testable predictions across games. The first prediction is that receivers must be fully accurate, that is, always play the optimal action associated to the true type of the sender. The other two predictions correspond to the characterization in Proposition 1.

Observation 1. *If the players play according to FRE, receivers must be fully accurate and always act skeptically, and senders must always send messages for which they are a worst-case type.*

This provides us with three natural indicators for FRE. We use them to check the extent to which participants play according to FRE in all sessions where a FRE exists. This includes all sessions in which an acyclic game is played, as well as sessions 18 to 21 in which the game is cyclic but has a FRE (it is Game 3 in Table 2).

IEODS. The iterative elimination described in Section 3 makes simple predictions on the behavior of players at each round based on the decomposition we obtained. These predictions can be described in general terms as follows. In all games of the experiment, IEODS finishes in four rounds or less.

Round 1 Receiver acts consistently, i.e. never picks an action that corresponds to a type that has been ruled out by evidence.

Round 2.1 Satisfied sender fully discloses.

Round 2.2 Envious sender sends a vague message.

Round 2.3 Envious sender obfuscates advantageously, i.e. sends a vague message whose consistent interpretations can possibly lead to a positive payoff.

Round 3.1 If possible, receiver never attributes a vague message to a satisfied sender.

Round 3.2 If possible, receiver never attributes a vague message to an envious sender for whom this message is not advantageous.

Round 4 If possible, envious sender only sends vague messages that can possibly lead to a positive payoffs given the interpretations of a round 3 receiver.

Furthermore, there is a natural filiation in the chain of reasonings that lead to the different rounds as summarized in Figure 3. We do not expect different steps of reasoning in the same round to all require the same level of sophistication. For example, elimination rounds 2.1 and 2.2 seem to require less sophistication than round 2.3. Similarly, it seems possible for a receiver to follow a chain of reasonings leading him to perform round 3.1, but not 3.2.

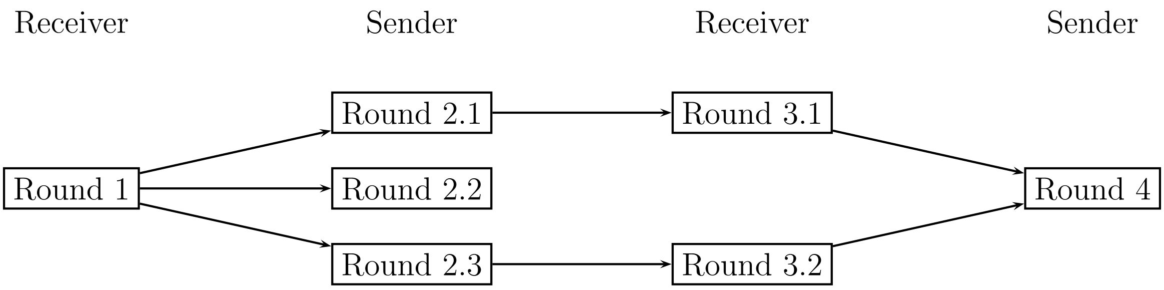


Figure 3: *Decomposition of IEODS reasoning.*

In Table 9 and Table 10 of Appendix B, we provide a game by game description of surviving actions and messages after each round of elimination. This leads to the following observation on the relationship between FRE and IEODS.

Observation 2. *In acyclic games, there exists a strategy profile supporting a FRE that survives IEODS. In simple acyclic games (Acyclic 1), all strategy profiles that survive IEODS support a FRE, but it is not the case in other acyclic games (Acyclic 2). In our cyclic game with FRE, all strategy profiles that support a FRE are eliminated.*

We explore three ways in which IEODS can help understand the data. First, we assess the extent to which IEODS predicts behavior in our games by looking at the fraction of sender and receiver observations that pass all rounds of elimination. Second, we use IEODS as a model of strategic thinking and explore the ability of players to perform the different rounds of elimination. To do this we look at failure rates for the different steps of reasoning, i.e. the rate at which players failed to pass a step, when failure was possible. We consider all steps in the decomposition of IEODS, thereby providing a finer exploration of depth of reasoning in these games. Finally, we explore the extent to which IEODS can provide a refinement to PBE by comparing the range of payoffs predicted by PBE in strategies that survive IEODS to empirical payoffs.

6 Overview

In this section, we provide an overview of the data through an examination of average payoffs, and a description of sender and receiver behavior.

6.1 Gains from Communication

To get a general sense of the data, we start by looking at the average scores of senders and receivers, as described on Figure 4. In every game, the receiver gets a payoff of 3 if she takes her unique optimal action, and 0 otherwise. Each of her three possible actions is equally likely to be optimal at the outset, so her average score in the absence of communication would be 1. Hence, we say that receivers gain from communication whenever their average payoff is above 1. The average payoff of the sender if the receiver chooses her action according to the uniform distribution is between 1 and 4/3, depending on the game. This range is depicted as the red hatched region in Figure 4. We say that senders gain from communication if their average payoff is above 4/3, and lose from communication if their average payoff is below 1.[15]

According to these definitions, both senders and receivers gain from communication. Satisfied senders gain but envious senders lose from communication. The loss of envious senders comes mainly from acyclic games. Receivers gain from communication regardless of the precision of the message they receive, and of cyclicality. But their scores are lower after vaguer

[15] The absence of communication is a possible benchmark. The reader interested in how subjects perform relative to the best and the worst equilibria in each game can go to Appendix C. It presents, for each session, the range of equilibrium and empirical payoffs.

messages, and in cyclic games. As suggested by theory that links acyclicity to existence of FRE, acyclic games favor receivers and harm envious senders. The poorer performance of receivers when facing vague messages, even in acyclic games, however, is not compatible with FRE. Indeed, while FRE does not preclude the use of vague message, it implies that the receivers should guess the type of the sender anyway.

The fact that receivers obtain an average payoff close to 3 when they face full disclosure confirms that the subjects understood that senders only could show cards they had truly observed. The fact that satisfied senders obtain an average payoff close to 3 both in cyclic and acyclic games shows that subjects managed to convey information effectively when their interests were aligned with those of the receivers.

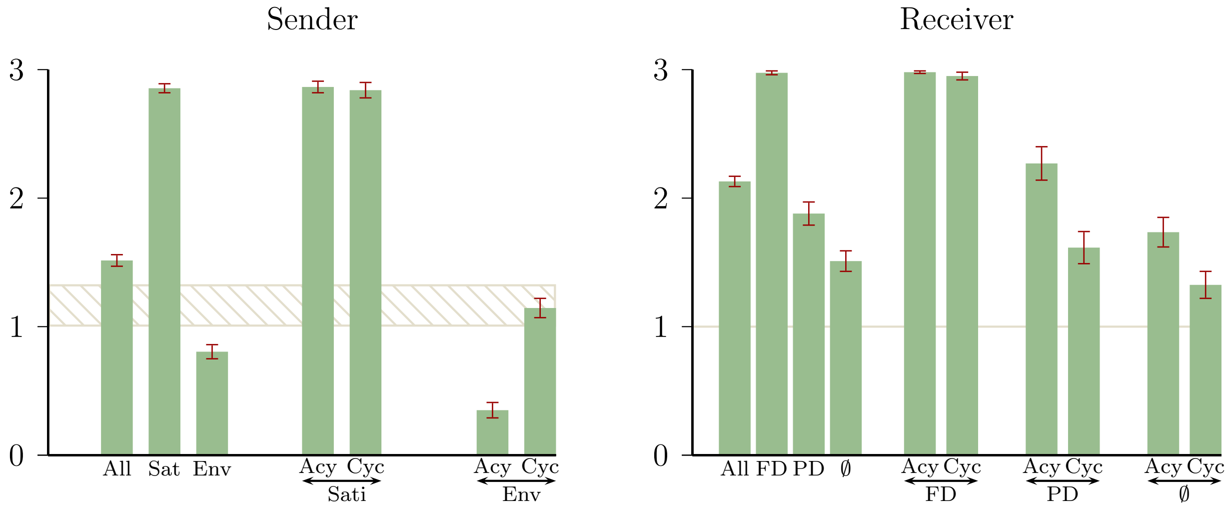


Figure 4: *Sender and receiver average payoffs.*

6.2 Receivers

Are receivers consistent? Both PBE and the first round of IEODS require consistent receiver strategies. Overall, excluding silence, after which any action is consistent, almost 99% of receivers decisions are consistent whether facing FD or PD, and whether the game is cyclic or acyclic.

The fact that receivers take evidence into account can also be seen by looking at the receivers beliefs, reported in Appendix D. Following full disclosure, the average probability that receivers put on the type disclosed was above 99% for each of the FD messages. Elicited beliefs are also consistent with evidence for PD messages, as can be seen in Table 11 of the appendix: receivers never put any weight on the type of the sender that is ruled out by evidence.

How accurate are receivers' decisions? We say that a receiver's decision is accurate if it corresponds to the optimal action associated with the true sender type. Overall, decisions are accurate 71% of the time, confirming that receivers benefit from communication. The accuracy rate jumps to 80% in acyclic games, which is closer to the prediction of FRE. In cyclic games, receivers achieve an accuracy rate of only 62%. Following full disclosure, receivers have an accuracy rate of 99% over all games. Vague messages lead to lower accuracy rates, irrespective of cyclicality. In acyclic games, the accuracy rates following partial disclosure is 76%, and it is 58% after silence. A receiver who would choose randomly with uniform weights after considering the evidence would obtain accuracy rates of 50% after partial disclosure, and 33% after silence. These rates are below the performance of receivers in acyclic games, but comparable to their performance in cyclic games, with 54% after partial disclosure, and 44% after silence. If we isolate cyclic games with a FRE (sessions 18 to 21), we obtain accuracy rates of 60% following partial disclosure, and 51% following silence. Interestingly, while receivers do perform better in these games than in other cyclic games, they perform less well than in acyclic games despite the existence of a FRE.

How long does it take receivers to act? Receivers' response times (Figure 5) yield two interesting observations. First, the response time in cyclic games is higher than in acyclic games whenever the message is vague, which suggests that cyclic games require more thinking on the receiver side. Second, response times are non-monotonic in the amount of disclosure: they are lowest for full disclosure, and highest for partial disclosure. On the face of it, this observation is slightly surprising as receivers face more possible responses after a silent message, than after partial disclosure. It suggests that deciphering the intentions of the sender following partial disclosure requires more thinking. This increased difficulty, however, does not translate into more mistakes, as the accuracy data show.

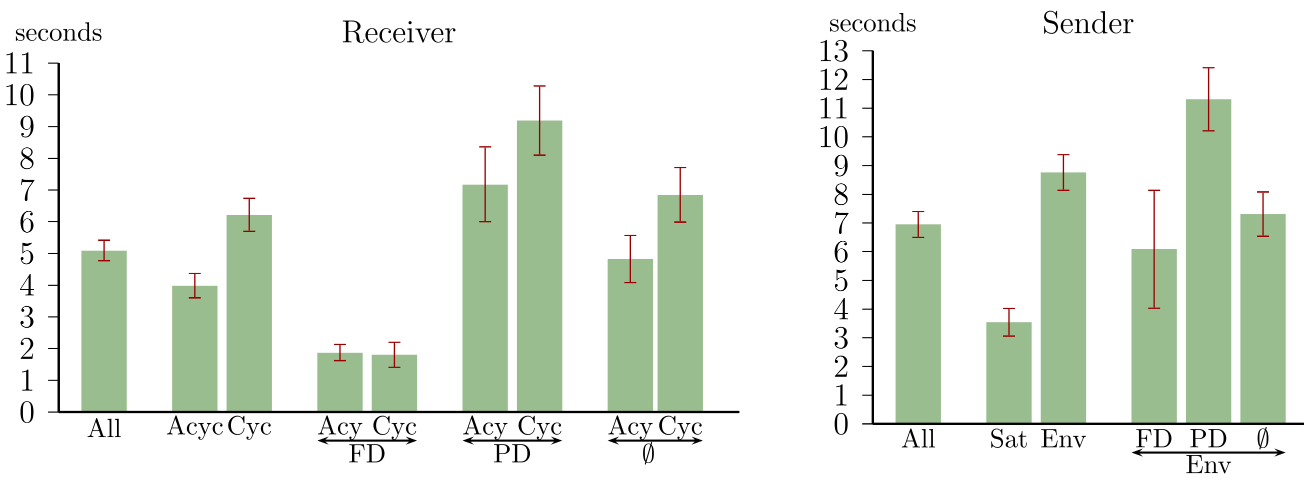


Figure 5: *Sender and receiver average response times.*

6.3 Senders

What do senders disclose? The disclosure strategies of senders are illustrated in Figure 6. Satisfied types overwhelmingly fully disclose whether the game is cyclic or acyclic. This corresponds to round 2.1 of IEODS: provided that receivers are consistent, fully disclosing when satisfied obviously dominates any other sender strategy. Envious senders, on the other hand, understand that they need to obfuscate in order to stand a chance of obtaining their first-best payoff. Indeed, their frequency of full disclosure is only 5%. This corresponds to round 2.2 in IEODS. In acyclic games, envious types do not use full disclosure more frequently than in cyclic games (resp. 5.1% and 4.5%). This shows that envious senders try to take advantage of receivers' possible lack of skepticism by obfuscating even though this would be useless in FRE. But, if receivers sometimes fail to be skeptical following vague messages (as the data indicate), it is better to be vague. Finally, envious senders use silence more often than partial disclosure (resp. 60% and 40%, conditional on using vague messages), both in cyclic and acyclic games.

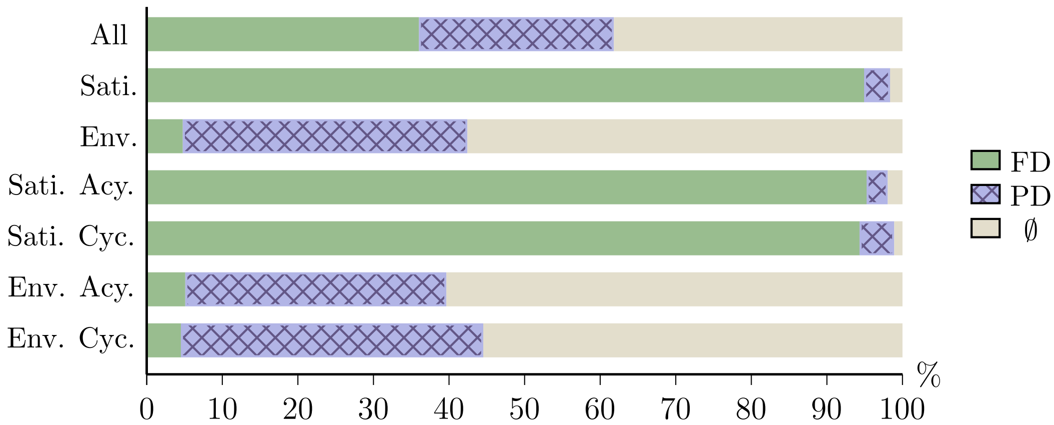


Figure 6: *Sender disclosure.*

How long does it take senders to act? Figure 5 shows that the response times of envious senders are higher than those of satisfied senders. In addition, the response times of envious senders exhibit the same non-monotonicity in the amount of disclosure as those of receivers: they are highest for partial disclosure, and lowest for silence and full disclosure. Envious senders rarely fully disclose. Interestingly, partial disclosure requires more thinking time than silence, which might explain why envious senders often prefer silence to partial disclosure.

7 Models of Behavior

PBE. The experiment was not designed to test PBE, and there are few metrics by which we can do so. Furthermore, all of our games have multiple equilibria. A minimal test of PBE is provided by the fraction of inconsistent receiver actions, which should by assumption be 0. In the data it is 0.62% if we consider all observations, and 1% if we exclude observations in which the sender remained silent. Payoffs provide another metric. In Appendix C, we show for each session a plot of the empirical payoffs and the range of PBE payoffs. While PBE appears to be rejected for receiver payoffs in sessions 15 and 17, for BB sender payoffs in session 7, and for YB sender payoffs in session 18, the range of PBE payoff predictions is generally consistent with the data, and otherwise not very far off. But of course, this is a very weak test, and the range of possible payoffs is often the full range of possible payoffs.

We can also get a rough idea of the validity of PBE by looking at elicited beliefs, as PBE predicts that beliefs should be accurate and players should best respond to their beliefs. In Appendix D, Table 11 and Table 12 show reported beliefs of receivers and senders for the experimental sessions in which we elicited them. While this data does show some very strong discrepancies between average beliefs and actual frequencies in some cases,[16] there are few of these and, for the receiver, they tend to correspond to messages that the senders have very rarely used. For more moderate differences, one patterns can be noticed that is suggestive of a systematic departure from PBE: senders tend to largely underestimate the degree of skepticism of receivers in acyclic games, especially following the PD message that certifies the two envious types. In Appendix E, we report the share of individual receiver and sender choices that were best responses to their reported beliefs. This share is between 65% and 98% for all sessions.

FRE. When a FRE exists, it is a focal prediction. However, FRE does not preclude multiplicity, as senders are indifferent across all messages for which they are a worst-case type. Among our games, all acyclic ones have a FRE, as well as one of our cyclic games. Based on Observation 1, we propose three metrics to assess whether FRE is matched by empirical behavior for these games. The first metric is the accuracy rate of the receivers. The second metric is the rate of skeptical choices by the receivers. The third metric is the rate of sender observations in which a sender chose a message for which she was a worst-case type.

We present these measures in Table 6 for all game categories for which a FRE exists. For acyclic games, at least 70% of observations match the prediction of FRE on all but one metric: accuracy following silent messages. The distance to FRE for this metric is mainly due to acyclic games with two envious types. Indeed, for simple acyclic games with a single envious type, 80% of all observations match FRE on one metric, and more than 93% on all others. Thus, according to our measures, FRE is a good prediction for acyclic games with a single envious type.

[16] The highest difference is 53 percentage points in session 12.

	Acyclic	Acyclic 1	Acyclic 2	Cyclic FRE	All Games with FRE
% accuracy	80.22%	95.86%	73.12%	71.69%	78.01%
– FD	99.52%	99.74%	99.34%	97.10%	98.92%
– PD	75.65%	81.82%	73.81%	60.00%	73.38%
– silence	57.85%	93.69%	50.38%	51.46%	55.78%
% skeptical	88.39%	98.45%	83.83%	54.31%	79.56%
– vague	79.32%	95.98%	75.30%	22.73%	64.19%
– PD	90.05%	93.18%	89.12%	90.77%	90.16%
– silence	72.94%	98.20%	67.67%	8.41%	52.00%
% wct message	82.26%	98.28%	75.00%	53.23%	74.74%
– satisfied	95.98%	97.39%	94.75%	100.00%	96.87%
– envious	71.41%	100.00%	64.73%	26.74%	58.66%

Table 6: *FRE metrics. If FRE was satisfied, all figures should be 100%.*

For more complicated acyclic games with two envious types, the picture is more contrasted. While at least 70% of the data is consistent with FRE on all but three metrics, the accuracy rate following silence is only 50%, which is better than chance (33%) for a receiver that considers all types to be possible after silence (round 1 receiver in the IEODS procedure), but undistinguishable from chance for a round 3 receiver who eliminates the single satisfied type and attributes the silent message to either of the two envious ones. The contrasted performance of FRE for these games is confirmed by elicited beliefs (see Appendix D), which show that at the end of the session, while receivers generally put more weight on skeptical interpretations of messages, they are far from attributing each message to a single type (FRE), and while senders tend to expect the skeptical action with higher probability, they are far from putting all weight on this action.

Finally, for cyclic games with a FRE, it is safe to say that FRE is largely rejected by the data, with compliance rates below 60% on most metrics, and as low as 8% for skepticism following silence. The low degree of skepticism following silence is especially meaningful: the only strategy that can dissuade YY and BB to pool on the silent message in this game is to skeptically attribute it to the satisfied type YB, a prediction that is clearly not matched by empirical behavior. This is confirmed by elicited beliefs which show that receivers put little weight on type YB, and senders little weight on receiver playing b.

IEODS. First, we assess the extent to which IEODS is matched by the data. Table 7 shows that the fraction of receiver observations that are consistent with IEODS is 96.58%, while it is 85.81% for sender observations. Overall, IEODS appears to predict behavior quite well. In Table 8, we break down these figures according to our game categories. The first thing to notice is that IEODS provides a good explanation of the data for all game categories, as it explains at least 90% of receiver data and 80% of sender data in each category. Note that the discrepancy between sender data and receiver data is persistent suggesting that reasoning through IEODS is harder for senders than for receivers in all of these games. To understand Table 8 further, it is useful to refer to Observation 2. In simple acyclic games (Acyclic 1), more than 90% of both sender and receiver data satisfy IEODS. For these games, IEODS uniquely selects a particular FRE (see Appendix B). Hence, in this case, IEODS makes more precise predictions than FRE since it pins down the separating equilibrium strategy that sender is using. For other acyclic games, while there exists a FRE strategy profile that survives IEODS, this is not the unique prediction of IEODS. Indeed, we found in the previous paragraph that FRE indicators are not as good for these games. In cyclic games with a FRE, no FRE strategy profile survives IEODS in our games, and indeed, the data is inconsistent with FRE for these games as we saw in Table 6, whereas IEODS remains a good match.

Second, we consider IEODS as a model of strategic thinking, and look in more details at the different rounds of elimination. Table 7 shows the fraction of the data that fail each step of each round of elimination. These figures suggest that players largely perform all steps of IEODS even though it might be slightly more difficult for senders than for receivers. However, this picture is misleading as there are many circumstances in which all possible decisions of a player comply with a given step of reasoning. For example, a receiver confronted with silence cannot be inconsistent, or a satisfied sender cannot fail rounds 2.2, 2.3 or 4. Also, round 4 eliminations occur in some games, but not in others. To get a better picture of whether players are able to perform the different parts of the IEODS reasoning, we measure the failure rates for each step of reasoning as given by the fraction of observation failing a given step when failure was possible. These figures, reported in Figure 7, provide a very different picture. For example, when confronted with a situation in which failing round 4 was possible, senders passed round 4 only 66% of the time. The general pattern confirms the immediate intuition that some steps are more difficult than others. It also shows that difficulty is not homogeneous within rounds. Just as intuition would suggest, steps 1, 2.1, 2.2 and 3.1 appear ‘easier’ to perform than steps 2.3, 3.2 and 4.

Finally, we examine IEODS as a refinement of PBE by looking at PBE that only use strategies that survive IEODS, and denote this refinement as IEODS+PBE. In simple acyclic games, as already noticed, this uniquely selects a particular FRE, and predicts a strategy profile that fits the data very well. To assess the usefulness of this refinement more generally, we look at how IEODS+PBE restricts the payoff predictions for sender types and receiver compared to PBE. This is shown in Figure 8 and Figure 9 in Appendix C. Overall, in all games, IEODS+PBE refines predictions close to average empirical payoffs.

	Receiver	Sender	
Fail R1	**0.62%**	**8.58%**	**Fail R2**
		1.76%	Fail R2.1
		3.11%	Fail R2.2
		3.71%	Fail R2.3
Fail R3	**2.79%**	**5.61%**	**Fail R4**
Fail R3.1	1.54%		
Fail R3.2	1.25%		
Pass IEODS	**96.58%**	**85.81%**	**Pass IEODS**

Table 7: *Explaining the data with IEODS: each figure in the table shows the fraction of all 3690 observations that fail or pass the corresponding round.*

	Acyclic 1	Acyclic 2	Cyclic FRE	Cyclic other
Sender	92.07%	80.70%	80.15%	91.36%
Receiver	98.45%	97.03%	93.85%	96.69%

Table 8: *Fraction of observations that pass all rounds of IEODS by game category.*

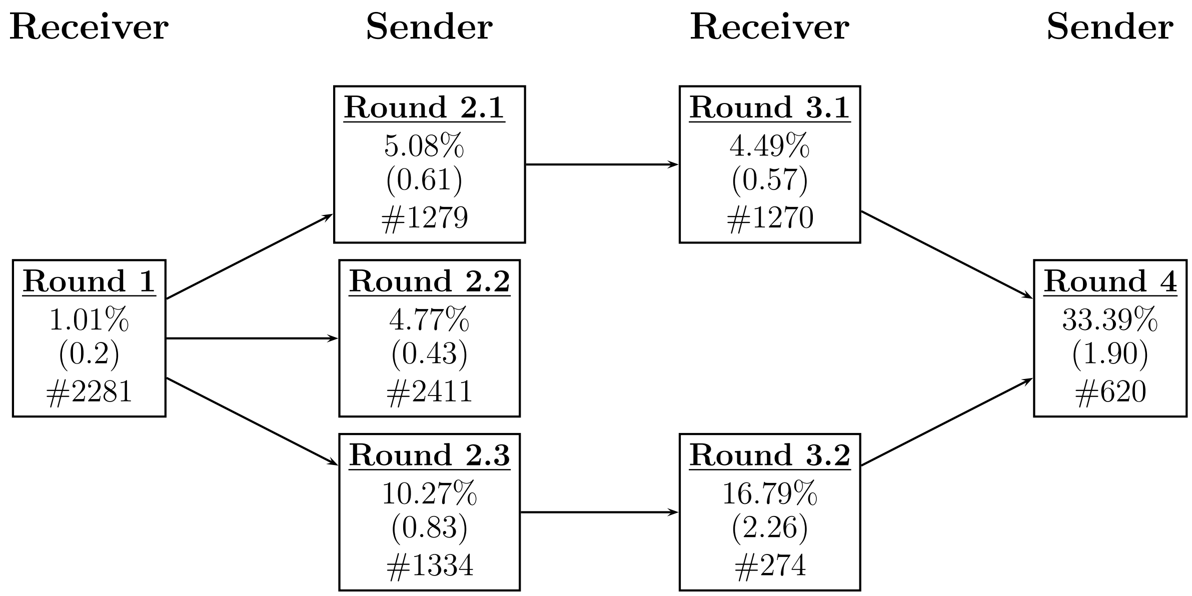


Figure 7: *Failure rates for the different rounds. The first number is the rate in percentage, the second the standard deviation in percentage points, and the third the number of observations.*

8 Conclusion

We propose a novel experimental study of disclosure games. Our design is innovative in its joint treatment of multiple games with different strategic properties. This is made possible by our graphical representation of sender's incentive which forms the basis for the definition of game, type and message categories that facilitate the discussion. We show that a non-equilibrium approach based on the iterated elimination of obviously dominated strategies generates meaningful restrictions that are well matched by the data. This procedure can also serve as an equilibrium selection device that increases the predictive power of PBE in a way that is consistent with empirical behavior in the lab for our class of games.

We find the use of IEODS as a non-equilibrium model of strategic thinking appealing because it does not rely on excessive player rationality. In particular, as pointed out in Li (2017) and Zhang and Levin (2015), it does not require contingent thinking or Bayesian updating by the players. In the games of our experiment, as in the game of Schipper and Li (2018), IEODS makes the same predictions as their concept of cautious rationalizability. Predictions do differ in some disclosure games yet to be studied in the lab.

Appendix

A Theoretical Appendix

FRE. Proceeding as in Hagenbach et al. (2014), we say that (μ, α, β) is a PBE with extremal beliefs if, in addition, it satisfies the property that for every off-path message $m \notin \mu(T)$, β_m puts probability 1 on a single type. Hagenbach et al. (2014) prove that, under Assumption 1, a FRE (μ, α, β) with extremal beliefs exists if and only if every certifiable subset of types has a worst-case type. The following lemma shows that the restriction to extremal beliefs is without loss of generality for our class of games.

Lemma 2. *If (μ, α, β) is a PBE in pure strategies, then there exists a belief system β' such that $\beta_m = \beta'_m$ for every on-path message m, and such that (μ, α, β') is a PBE with extremal beliefs.*

Proof. Consider a PBE (μ, α, β). Let m be an off-path message. Note that, by Assumption 2, any action $a \notin a^*(\mathcal{E}(m))$ yields a null payoff, while for any belief concentrated on $\mathcal{E}(m)$, there exists an action in $a^*(\mathcal{E}(m))$ that yields a strictly positive payoff. Therefore there exists $t \in \mathcal{E}(m)$ such that $\alpha(m)$ puts probability one on $a^*(t)$. Then we can define β'_m as the belief that puts mass 1 on t. Let β' be the belief system obtained by doing this operation for every off-path message m, and by letting $\beta'_m = \beta_m$ for the remaining messages. It is easy to see that (μ, α, β') is a PBE with extremal beliefs. $\square$

This proves the following proposition.

Proposition 2. *A FRE in pure strategies exists if and only if every certifiable subset of types admits a worst-case type.*

Furthermore, the proposition can be extended to mixed strategy equilibria as a corollary of the following result.

Proposition 3. *There exists a mixed strategy FRE if and only if there exists a pure strategy FRE.*

Proof. The other implication being obvious, assume that there exists a mixed strategy FRE. In such an equilibrium, each type mixes over a set of messages. These sets do not overlap, since the equilibrium is separating. And following any of these messages, which is on-path for type t, the receiver takes action $a^*(t)$. We can pick one message m^t in the set of messages sent by type t, and modify the strategy of the sender to sending m^t with probability 1 when she is of type t. The other messages become off path, but we assume that the receiver reacts to this messages with the belief that they come from t with probability 1, and thus picks action $a^*(t)$. These beliefs are consistent, and because we started from an equilibrium, they deter other types from using the same message.

Consider the messages that are off path in the initial equilibrium. Upon seeing such a message m in the initial equilibrium, the receiver forms a belief β_m with support in $\mathcal{E}(m)$. Then, given Assumption 2, her mixed response can only be to mix in some way between actions $a^*(t')$ corresponding to types $t' \in \mathcal{E}(m)$ such that $\beta_m(t') > 0$.

Suppose first that $\mathcal{E}(m)$ admits a worst-case type. Then we can change the belief β_m to a belief that puts probability 1 on a worst-case type, and that would deter any type of the sender able to send m from deviating to m.

Suppose next that $\mathcal{E}(m)$ has no worst-case type. Then any consistent belief following m leads the receiver to play some action $a^*(t')$ with $t' \in \mathcal{E}(m)$ with strictly positive probability. Since $\mathcal{E}(m)$ has no worst-case type, there exists a type $t \neq t'$ in $\mathcal{E}(m)$, that strictly prefers $a^*(t')$ to $a^*(t)$, and by Assumption 3, this type t must also prefer to $a^*(t)$ any mixture over consistent actions that puts strictly positive probability on $a^*(t')$. But this in turn implies that β_m could not have supported a FRE in mixed strategies. □

Finally, we prove the characterization of FRE in Proposition 1 which is easily extended to mixed strategy equilibria.

Proof of Proposition 1. Suppose that (ii) is satisfied. Every message is interpreted as worst-case type, so the sender has no incentive to deviate from the separating strategy. It is easy to construct the extremal belief system that supports the receiver strategy as an equilibrium, so we have a FRE.

Now assume that the sender follows the strategy prescribed by (ii) but that the receiver does not interpret every message skeptically. Then, there exists a message m after which the receiver plays $a^*(t')$ where t' is such that $t \to t'$ for some type t in $\mathcal{E}(m)$. Type t will prefer sending m than sending the prescribed $\mu(t)$ that uncovers him, making FRE impossible.

Now suppose that the receiver interprets every message skeptically, that $\mu(\cdot)$ is separating (otherwise, we trivially cant' have a FRE) but that t is not a worst-case type of $\mu(t)$. Then, for (μ, α, β) to be a FRE, it must be the case that $\alpha(\mu(t)) = a^*(t)$. This interpretation is not skeptical, a contradiction. □

IEODS. Next we prove the decomposition that characterizes the first rounds of elimination in Section 3.

Proof of Round 1. $\mathcal{M}^1 = \mathcal{M}^0$, because, when all actions are available for the receiver, all messaging strategies lead to the same possible scenarios.

By Assumption 2, the receiver gets a null payoff whenever she plays an inconsistent action. By contrast, any consistent action $a^*(t)$ yields a positive payoff in the scenario in which the message comes from t, and 0 in all other scenarios. Hence a strategy that plays an inconsistent action following some message is obviously dominated by one that replaces this action by a consistent one and is otherwise identical to the original strategy. On the other hand, a consistent strategy cannot be obviously dominated by any other as, following every message, there is a scenario in which the action prescribed by this strategy is the only one that yields a positive payoff. □

Proof of Round 2. $\mathcal{A}^2 = \mathcal{A}^1$ because $\mathcal{M}^1 = \mathcal{M}^0$.

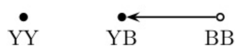
Consider a satisfied type t of the sender. Full disclosure yields to $a^*(t)$ by consistency (Round 1), and this is preferred to any other action the receiver may take by Assumption 3. Since any vague message can lead to a scenario in which the consistent receiver chooses one of these actions, vague messages are obviously dominated.

Next, consider an envious type t. Full disclosure leads to $a^*(t)$ which yields a null payoff by Assumption 3. By contrast, silence may lead to any possible action by a consistent receiver, one of which with a strictly positive payoff, and others with null payoff. Therefore silence obviously dominates full disclosure, and is obviously undominated. If, instead of remaining silent, t partially discloses with some message m, then either m is advantageous for t, in which case m is not obviously dominated by silence, or m is not advantageous, in which case m yields a null payoff and is obviously dominated by silence. □

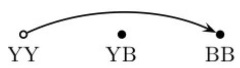
Proof of Round 3. Satisfied types cannot send vague messages so choosing the action $a^*(t)$ associated with a satisfied type t must lead to a null payoff, and if a message is still in the set of messages $\mathcal{M}^2(T)$ sent after round 2, then there must exist an envious type s for whom

m is advantageous, implying that choosing the action $a^*(s)$ yields a positive payoff in at least one possible scenario, so $a^*(t)$ is obviously dominated by $a^*(s)$ following m. Similarly, if is not advantageous for t to send m, then choosing $a^*(t)$ yields a null payoff and is obviously dominated by $a^*(s)$. $\square$

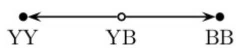
B Game by Game IEODS



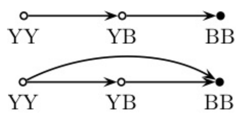
Round	Player	Type	∅	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
2	S	$\boxed{YY}$	× (2.1)	× (2.1)	•	✓	•	•
		$\boxed{YB}$	× (2.1)	× (2.1)	× (2.1)	•	✓	•
		BB	✓	•	✓	•	•	× (2.2)
3	R		c (3.1)	a, b	c (3.1)	a	b	c



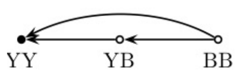
Round	Player	Type	∅	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
2	S	YY	✓	× (2.3)	•	× (2.2)	•	•
		$\boxed{YB}$	× (2.1)	× (2.1)	× (2.1)	•	✓	•
		$\boxed{BB}$	× (2.1)	•	× (2.1)	•	•	✓
3	R		a (3.1)	a, b	b, c	a	b	c



Round	Player	Type	∅	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
2	S	$\boxed{YY}$	× (2.1)	× (2.1)	•	✓	•	•
		YB	✓	✓	✓	•	× (2.2)	•
		$\boxed{BB}$	× (2.1)	•	× (2.1)	•	•	✓
3	R		b (3.1)	b (3.1)	b (3.1)	a	b	c

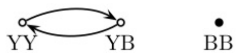


Round	Player	Type	∅	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
2	S	YY	✓	✓	•	× (2.2)	•	•
		YB	✓	× (2.3)	✓	•	× (2.2)	•
		$\boxed{BB}$	× (2.1)	•	× (2.1)	•	•	✓
3	R		a, b (3.1)	a (3.2)	b (3.1)	a	b	c
4	S	YY	✓	× (4)	•	×	•	•
		YB	✓	×	✓	•	×	•
		$\boxed{BB}$	×	•	×	•	•	✓

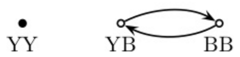


Round	Player	Type	∅	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
2	S	$\boxed{YY}$	× (2.1)	× (2.1)	•	✓	•	•
		YB	✓	✓	× (2.3)	•	× (2.2)	•
		BB	✓	•	✓	•	•	× (2.2)
3	R		b, c (3.1)	b (3.1)	c (3.2)	a	b	c
4	S	$\boxed{YY}$	×	×	•	✓	•	•
		YB	✓	✓	×	•	×	•
		BB	✓	•	× (4)	•	•	×

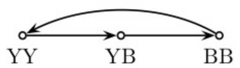
Table 9: *Iterated Elimination procedure for acyclic games. For receiver, we write surviving actions following each message (column) at the corresponding round. For each type of sender, • denotes a message that is not available, × an eliminated message, and ✓ a surviving message. Satisfied types are framed. Numbers in parenthesis indicate the step of reasoning that leads to keep/eliminate the corresponding action(s)/message.*



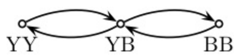
Round	Player	Type	$\emptyset$	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
		YY	$\checkmark$	$\checkmark$	$\bullet$	$\times$ (2.2)	$\bullet$	$\bullet$
2	S	YB	$\checkmark$	$\checkmark$	$\times$ (2.3)	$\bullet$	$\times$ (2.2)	$\bullet$
		$\boxed{BB}$	$\times$ (2.1)	$\bullet$	$\times$ (2.1)	$\bullet$	$\bullet$	$\checkmark$
3	R		a, b (3.1)	a, b	b, c	a	b	c



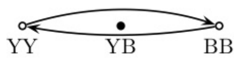
Round	Player	Type	$\emptyset$	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
		$\boxed{YY}$	$\times$ (2.1)	$\times$ (2.1)	$\bullet$	$\checkmark$	$\bullet$	$\bullet$
2	S	YB	$\checkmark$	$\times$ (2.3)	$\checkmark$	$\bullet$	$\times$ (2.2)	$\bullet$
		BB	$\checkmark$	$\bullet$	$\checkmark$	$\bullet$	$\bullet$	$\times$ (2.2)
3	R		b, c (3.1)	a, b	b, c	a	b	c



Round	Player	Type	$\emptyset$	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
		YY	$\checkmark$	$\checkmark$	$\bullet$	$\times$ (2.2)	$\bullet$	$\bullet$
2	S	YB	$\checkmark$	$\times$ (2.3)	$\checkmark$	$\bullet$	$\times$ (2.2)	$\bullet$
		BB	$\checkmark$	$\bullet$	$\times$ (2.3)	$\bullet$	$\bullet$	$\times$ (2.2)
3	R		a, b, c	a (3.2)	b (3.2)	a	b	c
		YY	$\checkmark$	$\times$ (4)	$\bullet$	$\times$	$\bullet$	$\bullet$
4	S	YB	$\checkmark$	$\times$	$\times$ (4)	$\bullet$	$\times$	$\bullet$
		BB	$\checkmark$	$\bullet$	$\times$	$\bullet$	$\bullet$	$\times$



Round	Player	Type	$\emptyset$	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
		YY	$\checkmark$	$\checkmark$	$\bullet$	$\times$ (2.2)	$\bullet$	$\bullet$
2	S	YB	$\checkmark$	$\checkmark$	$\checkmark$	$\bullet$	$\times$ (2.2)	$\bullet$
		BB	$\checkmark$	$\bullet$	$\checkmark$	$\bullet$	$\bullet$	$\times$ (2.2)



Round	Player	Type	$\emptyset$	Y	B	YY	YB	BB
1	R		a, b, c	a, b	b, c	a	b	c
		YY	$\checkmark$	$\times$ (2.3)	$\bullet$	$\times$ (2.2)	$\bullet$	$\bullet$
2	S	$\boxed{YB}$	$\times$ (2.1)	$\times$ (2.1)	$\times$ (2.1)	$\bullet$	$\checkmark$	$\bullet$
		BB	$\checkmark$	$\bullet$	$\times$ (2.3)	$\bullet$	$\bullet$	$\times$ (2.2)
3	R		a, c (3.1)	a, b	b, c	a	b	c

Table 10: *Iterated Elimination procedure for cyclic games. For receiver, we write surviving actions following each message (column) at the corresponding round. For each type of sender,* $\bullet$ *denotes a message that is not available,* $\times$ *an eliminated message, and* $\checkmark$ *a surviving message. Satisfied types are framed. Numbers in parenthesis indicate the step of reasoning that leads to keep/eliminate the corresponding action(s)/message.*

C Theoretical and Empirical Payoffs

We have computed all equilibrium payoffs for each of our games. The range for these payoffs are indicated by the grey areas in Figure 8 for sessions with acyclic games and in Figure 9 for sessions with cyclic games. The multiplicity of equilibrium payoffs comes from both the multiplicity of equilibrium sender strategies, and from the receiver's indifferences over consistent actions due to the uniform probability distribution. The payoffs range of PBE that survive IEODS is indicated by the striped areas.

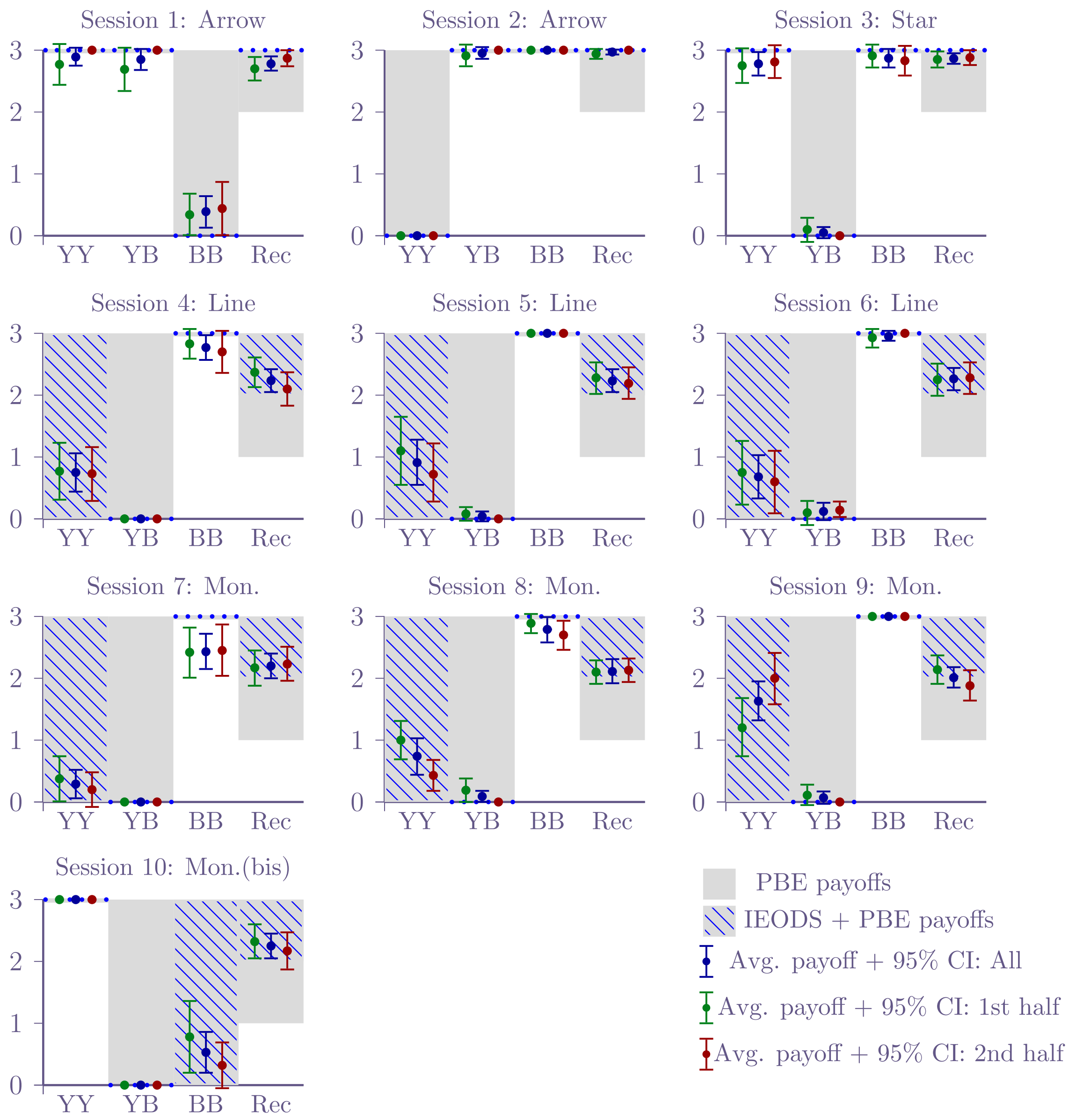


Figure 8: *PBE and IEODS + PBE predictions and actual payoffs for sessions with an acyclic game. The data points represent average payoffs and 95% confidence intervals over all periods, the 1st and 2nd half of the sessions.*

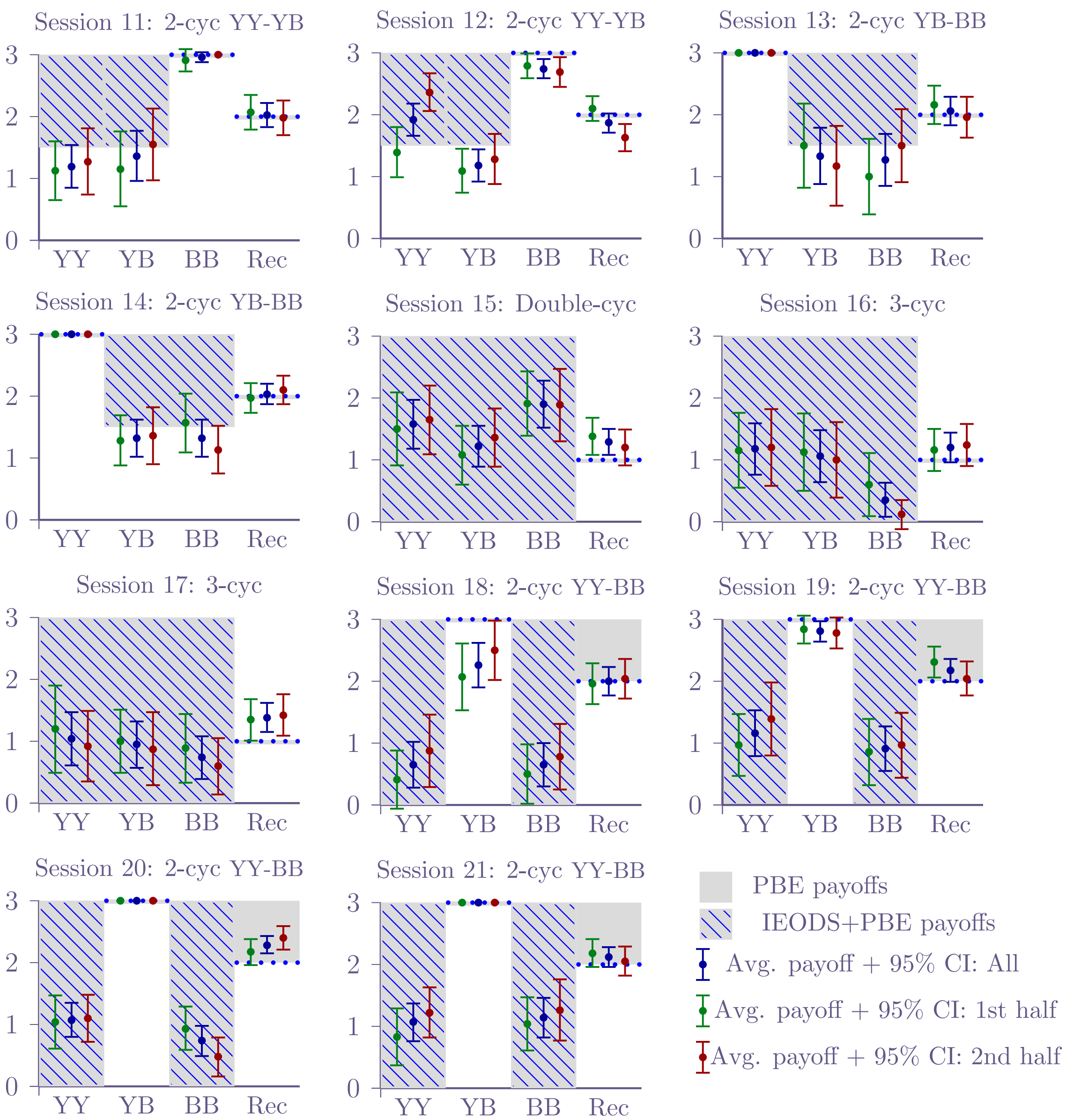
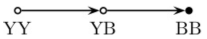


Figure 9: *PBE and IEODS+PBE predictions and actual payoffs for sessions with a cyclic game. The data points represent average payoffs and 95% confidence intervals over all periods, the 1st and 2nd half of the sessions.*

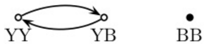
D Average Beliefs

Session 5



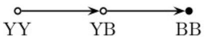
	Beliefs				Frequencies		
	YY	YB	BB		YY	YB	BB
∅	53.4	46.6	0	∅	46.7	53.3	0
	(5.6)	*(5.6)*	*(.)*				
Y	70.5	29.5	0	Y	73.9	26.1	0
	(8.1)	*(8.1)*	*(.)*				
B	0	89.5	10.5	B	0	100	0
	(.)	*(9.9)*	*(9.9)*				

Session 12



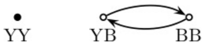
	Beliefs				Frequencies		
	YY	YB	BB		YY	YB	BB
∅	45.8	54	0.2	∅	45.2	54.8	0
	(4.0)	*(4.1)*	*(0.2)*				
Y	46.2	53.7	0	Y	53.3	46.7	0
	(2.8)	*(2.8)*	*(.)*				
B	0	65.7	34.3	B	0	12.5	87.5
	(.)	*(17.8)*	*(17.8)*				

Session 6



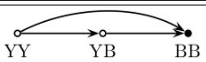
	Beliefs				Frequencies		
	YY	YB	BB		YY	YB	BB
∅	61	39	0	∅	39.5	60.5	0
	(6.4)	*(6.4)*	*(.)*				
Y	82	18	0	Y	76.7	23.3	0
	(5.0)	*(5.0)*	*(.)*				
B	0	98.9	1.1	B	0	93.3	6.7
	(.)	*(1.1)*	*(1.1)*				

Session 14



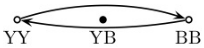
	Beliefs				Frequencies		
	YY	YB	BB		YY	YB	BB
∅	0	45.3	54.7	∅	0	48.5	51.5
	(.)	*(3.8)*	*(3.8)*				
Y	28.6	71.4	0	Y	.	.	.
	(18.4)	*(18.4)*	*(.)*				
B	0	51.1	48.8	B	0	50.8	49.2
	(.)	*(1.7)*	*(1.7)*				

Session 8



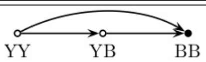
	Beliefs				Frequencies		
	YY	YB	BB		YY	YB	BB
∅	60	39.1	0.9	∅	59.4	37.6	3.0
	(5.6)	*(5.5)*	*(0.8)*				
Y	57.5	42.5	0	Y	68	32	0
	(9.2)	*(9.2)*	*(.)*				
B	0	96.5	3.5	B	0	88.9	11.1
	(.)	*(1.5)*	*(1.5)*				

Session 20



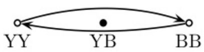
	Beliefs				Frequencies		
	YY	YB	BB		YY	YB	BB
∅	49.9	0.2	49.9	∅	48.1	0	51.9
	(0.1)	*(0.2)*	*(0.1)*				
Y	60.9	39.1	0	Y	66.7	33.3	0
	(17.0)	*(17.0)*	*(.)*				
B	0	39.1	60.9	B	0	0	100
	(.)	*(17.0)*	*(17.0)*				

Session 9



	Beliefs				Frequencies		
	YY	YB	BB		YY	YB	BB
∅	48.3	51.7	0	∅	60.7	39.3	0
	(3.8)	*(3.8)*	*(.)*				
Y	64.2	35.8	0	Y	53.9	46.1	0
	(8.2)	*(8.2)*	*(.)*				
B	0	90.8	9.2	B	0	100	0
	(.)	*(4.5)*	*(4.5)*				

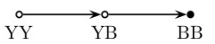
Session 21



	Beliefs				Frequencies		
	YY	YB	BB		YY	YB	BB
∅	48.6	2.8	48.6	∅	52.6	0	47.4
	(0.9)	*(1.8)*	*(0.9)*				
Y	93.6	6.4	0	Y	100	0	0
	(2.8)	*(2.8)*	*(.)*				
B	0	6.4	93.6	B	0	0	100
	(.)	*(2.8)*	*(2.8)*				

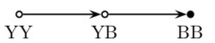
Table 11: *Receiver average beliefs and empirical frequencies.*

Session 5



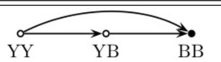
	Beliefs				Frequencies		
	a	b	c		a	b	c
∅	52.7	45	2.3	∅	56.7	43.3	0
	(4.1)	*(3.8)*	*(2)*				
Y	57.5	41.5	1	Y	95.7	4.3	0
	(10.6)	*(10.1)*	*(1)*				
B	1.1	94.1	4.8	B	0	94.1	5.9
	(1.1)	*(2.4)*	*(1.7)*				

Session 6



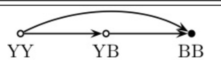
	Beliefs				Frequencies		
	a	b	c		a	b	c
∅	57.8	42.2	0	∅	65.8	32.9	1.3
	(6.6)	*(6.6)*	*(0)*				
Y	55	45	0	Y	80	20	0
	(9.5)	*(9.5)*	*(0)*				
B	0.2	97.8	2	B	0	93.3	6.7
	(0.2)	*(1.3)*	*(1.3)*				

Session 8



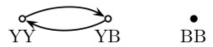
	Beliefs				Frequencies		
	a	b	c		a	b	c
∅	67.9	32	0.1	∅	70.3	28.7	1
	(8)	*(8)*	*(0.1)*				
Y	69.3	30.7	0	Y	88	12	0
	(9.8)	*(9.8)*	*(0)*				
B	0.5	96.8	2.7	B	0	83.3	16.7
	(0.5)	*(1.3)*	*(1.3)*				

Session 9



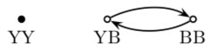
	Beliefs				Frequencies		
	a	b	c		a	b	c
∅	52.4	47.6	0	∅	52.5	47.5	0
	(8.3)	*(8.3)*	*(0)*				
Y	37.4	62.6	0	Y	69.2	30.8	0
	(12.5)	*(12.5)*	*(0)*				
B	1.7	98.3	0	B	0	91.7	8.3
	(1.7)	*(1.7)*	*(0)*				

Session 12



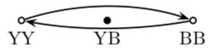
	Beliefs				Frequencies		
	a	b	c		a	b	c
∅	49	47.3	3.7	∅	26.2	73.8	0
	(4.8)	*(4.7)*	*(2.6)*				
Y	45	55	0	Y	44.2	55.8	0
	(3.1)	*(3.1)*	*(0)*				
B	0	88	12	B	0	75	25
	(0)	*(9.1)*	*(9.1)*				

Session 14



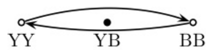
	Beliefs				Frequencies		
	a	b	c		a	b	c
∅	0	54.2	45.8	∅	3	54.5	42.5
	(0)	*(4.2)*	*(4.2)*				
Y	42.8	57.2	0	Y	.	.	.
	(17)	*(17)*	*(0)*				
B	0	53.6	46.4	B	0	46.3	53.7
	(0)	*(3.6)*	*(3.6)*				

Session 20



	Beliefs				Frequencies		
	a	b	c		a	b	c
∅	51.9	5	43.1	∅	39.2	3.8	57.0
	(3.5)	*(3.2)*	*(5.6)*				
Y	60.6	35	4.4	Y	33.3	66.7	0
	(10.4)	*(9.6)*	*(3.2)*				
B	11.9	36.9	51.2	B	0	25	75
	(8.9)	*(9.1)*	*(11.9)*				

Session 21



	Beliefs				Frequencies		
	a	b	c		a	b	c
∅	49.2	1.6	49.2	∅	47.4	9.0	43.6
	(0.7)	*(1.4)*	*(0.7)*				
Y	71	28.7	0.3	Y	66.7	33.3	0
	(13.6)	*(13.3)*	*(0.3)*				
B	0.3	28.7	71	B	40	20	40
	(0.3)	*(13.3)*	*(13.6)*				

Table 12: *Sender average beliefs and empirical frequencies.*

E Best Responses

Receivers best responses. We can use elicited beliefs to see whether receivers best responded to their reported beliefs. To that purpose, we created the indicator BR-b which, for every observation, takes value 1 if the receiver best responded to her reported beliefs in the period and the session corresponding to the observation, and 0 otherwise. We also created the corresponding indicator, BR-e, for best responses to the true empirical frequencies of the session. These two indicators allow us to decompose behavior into observations for which the receiver best responded to both her reported beliefs and the true frequencies (BR-b AND BR-e), situations in which she failed to best respond to her own beliefs (Not BR-b), and situations in which she did best respond to but was lead astray by her beliefs (BR-b AND Not BR-e). The corresponding rates when facing vague messages[17] are given for all sessions in Figure 10. The rate at which receivers failed to best respond to their belief is between 20% and 30% for most sessions, except sessions 20 and 21 for which it is lower. By contrast, the rate at which receivers best responded to inaccurate beliefs varies more widely across sessions corresponding to different games. This rate is higher in sessions with cyclic games.

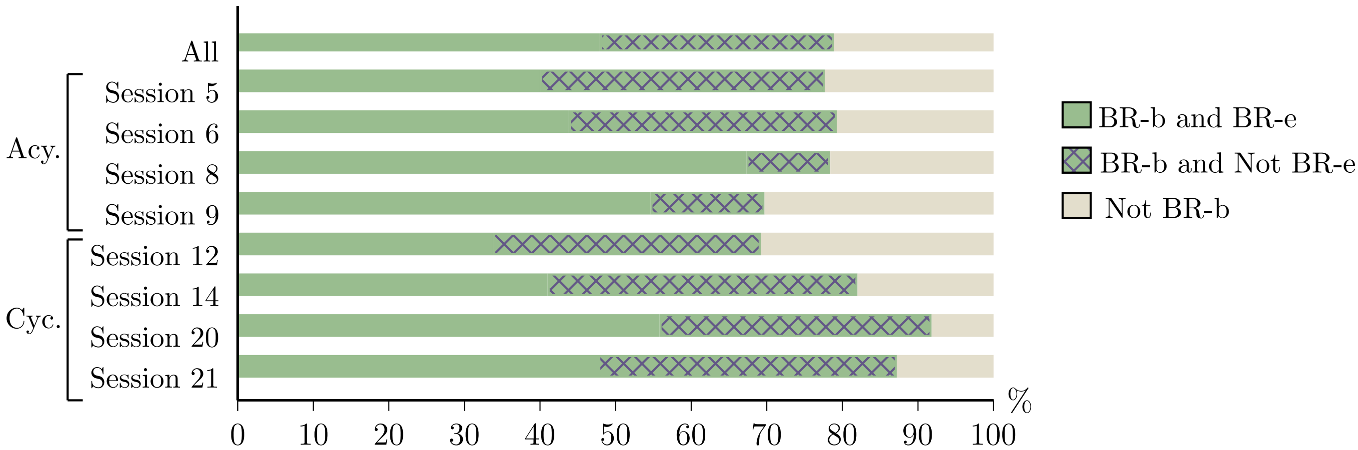


Figure 10: *Best-response rates for receivers facing with vague messages. BR-b and BR-e measures the average rate at which receivers best-responded to the elicited beliefs and to the empirical frequencies. BR-b and not BR-e measures the average rate at which receivers best-responded to the elicited beliefs but not to the empirical frequencies. Not BR-b measures the rate of failure of best-response to the elicited beliefs.*

Senders best responses. As for receivers, we can use elicited beliefs to see whether senders best responded to their reported beliefs and empirical frequencies. To that purpose, we created

[17]When facing FD, receivers best responded to beliefs and true frequencies at a rate close to 100% in all sessions.

the indicator BR-b which, for every observation, takes value 1 if the sender best responded to her reported beliefs in the period and the session corresponding to the observation, and 0 otherwise. We also created the corresponding indicator, BR-e, for best responses to the true empirical frequencies of the session. These two indicators allow us to decompose behavior into observations for which the sender best responded to both her reported beliefs and the true frequencies (BR-b AND BR-e), situations in which she failed to best respond to her own beliefs (Not BR-b), and situations in which she did best respond to but was lead astray by her beliefs (BR-b AND Not BR-e). The corresponding rates for envious senders[18] are given for all sessions in Figure 11. As for receivers, the rate at which envious senders failed to best respond to their belief is roughly between 20% and 30% for most sessions, except sessions 20 and 21 for which it is lower. By contrast, the rate at which senders best responded to inaccurate beliefs varies more widely across sessions corresponding to different games. It is higher in sessions with acyclic games.[19]

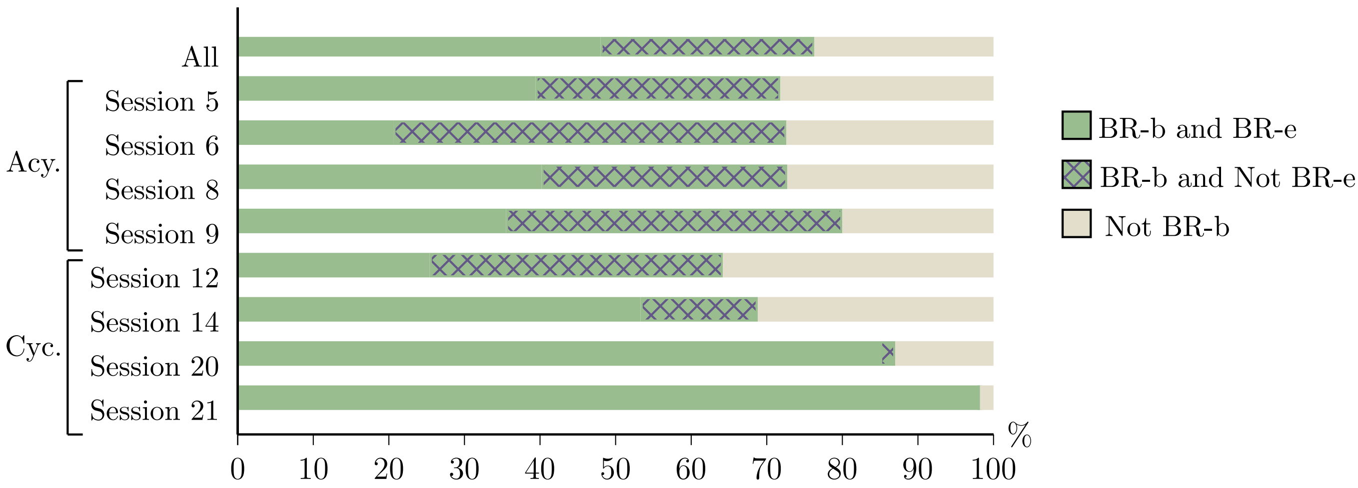


Figure 11: *Best-response rates for envious senders. BR-b and BR-e measures the average rate at which envious senders best-responded to the elicited beliefs and to the empirical frequencies. BR-b and not BR-e measures the average rate at which envious senders best-responded to the elicited beliefs but not to the empirical frequencies. Not BR-b measures the rate of failure of best-response to the elicited beliefs.*

[18]Satisfied senders best responded to beliefs and true frequencies at a rate close to 100% in all sessions.

[19]Session 12 is an exception, but this can be explained by the unbalance in the true frequencies of a and b actions following silence which can be noted in Appendix D, Table 12. Senders failed to pick up on this unbalance which they could have benefitted from, and instead best-responded to their reported beliefs.

F Screenshots

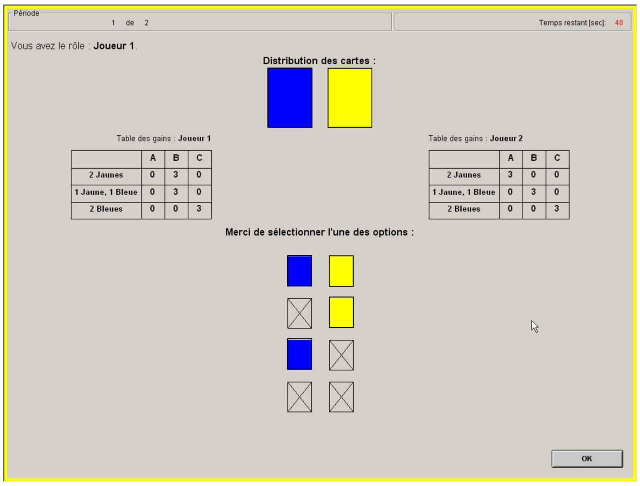


Figure 12: *Senders' screen.*

Figure 13: *Sender belief elicitation.*

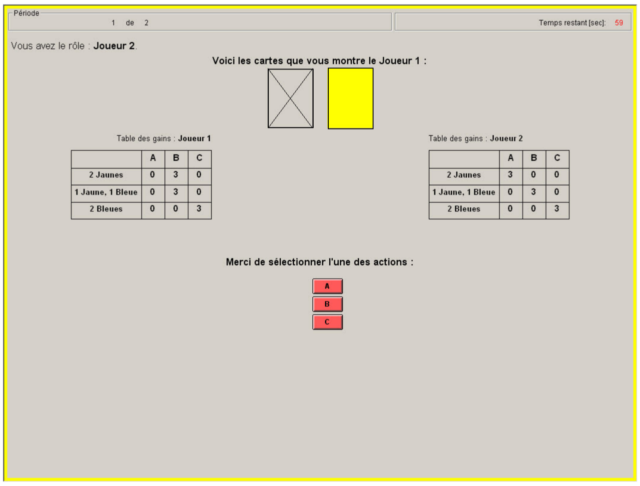


Figure 14: *Receivers' screen.*

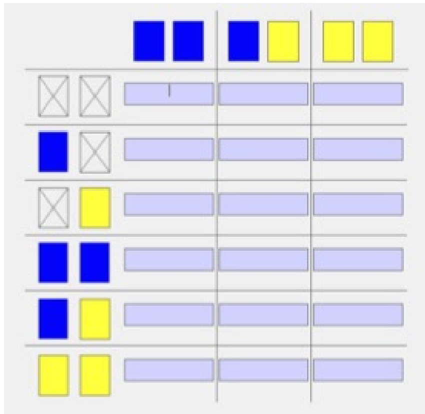


Figure 15: *Receiver belief elicitation.*

G Instructions (translated in English)

Thank you for agreeing to participate in this experiment. It is part of a research program on information transmission.

During the experiment, you will have to be perfectly focused and we ask you to turn cell phones off. It is important that you do not talk to any other participant during the whole duration of the experiment. At any point, do not hesitate to ask us if you have any question about the proceedings of the experiment.

In this experiment, you will be given the opportunity to make points. The more points you make, the larger the amount of money you will receive at the end of the experiment. For your coming to the lab, a 5 euros show-up fee will be added to your final gains.

Proceedings of the experiment

Participants' roles. Before the start of the experiment, a computer will randomly assign a role to each of you: half of you will be **Player 1** and the other half will be **Player 2**. You will keep this role for the whole duration of the experiment. The experiment is made up of 20 rounds (plus 2 trial rounds during which no points can be made).

A round of play. Each round of play proceeds as follows. First, the computer randomly creates pairs of subjects made up of a Player 1 and a Player 2. In each round, you will thus be matched with a new participant whose identity will never be revealed.

A pair of cards is then randomly drawn by the computer: either two Yellow cards are drawn, or two Blue cards are drawn, or one Yellow card and one Blue card are drawn. Each of these three pairs has the same chance to be drawn.

Next, the two cards are shown to Player 1 who then perfectly knows their two colors. The two cards are not shown to Player 2. Once Player 1 has seen the two cards, he has the possibility to show some cards to Player 2.

- Being Player 1, you can decide either to show the two cards you have seen, to show only one of them (and to choose which one) or to show none. What you will show is exactly what Player 2 will see on his computer screen.
- Being Player 2, you will observe the cards that Player 1 has decided to show you (if any) and you will choose one of 3 actions: A, B or C.

	A	B	C
2 Yellow	10	0	0
1 Yellow, 1 Blue	0	2	0
2 Blue	3	3	3

Table of gains: Player 1

	A	B	C
2 Yellow	5	5	5
1 Yellow, 1 Blue	0	10	0
2 Blue	0	0	1

Table of gains: Player 2

Points in each round. The number of points that you will make at the end of each round will depend on the color of the two cards *and* on the action chosen by Player 2. The exact number of such points will be shown in two tables of gains that you will observe during the whole round of play. Here is an example of such tables (these tables are not the one you will be playing with):

Table 1 gives the number of points that Player 1 can make and Table 2 the number of points that Player 2 can make. Each line corresponds to one of the three possible draw of cards and each column corresponds to one of the three possible actions (A,B or C). Thus, if one of the two cards is Yellow and the other is Blue and if Player 2 chooses action B, Player 1 makes 2 points and Player 2 makes 10 points.

During the 22 rounds of play, the tables of gains will be kept unchanged. However, two new cards are drawn in the beginning of each round of play.

Feedbacks at the end of a round. At the end of each round, you will be informed of all the choices that have been made. Being Player 1, you will learn which action has been chosen by Player 2 and how many points you made. Being Player 2, you will learn which cards had been drawn and how many points you made. Every player will also be shown how many points have been made by the other player. You will then move on to the next round of play.

Value of the points in euro.

At the end of the experiment, you will have played 22 rounds. The 2 first rounds are only trial ones and do not bring any points. Points made on each of the following 20 rounds are added and transformed in euros at the following rate: 5 points = 1 euro.

Good Luck !

Comprehension test

Before the experiment starts, we ask you to answer the following questions to make sure the instructions are clear for everyone. You will not be evaluated regarding your answers to these questions, and the experimentalist will give you the answers orally once everyone is done.

1. During the whole experiment, you will play with the same partner. True / False
2. Your role will be different in each round of play. True / False
3. In each round, new tables of gains will be proposed. True / False
4. Being Player 1, you have to show at least one card to Player 2 for him to be able to choose an action. True / False
5. Being Player 1, if your two cards are blue, you can show a yellow card to Player 2. True / False
6. Being Player 2, if Player 2 has not shown you any card, you will never know which cards he saw. True / False
7. The tables of gains are the ones given above. Being Player 2, if Player 2 shows you a yellow card, you will make 5 points whatever the action you choose. True / False
8. The tables of gains are the ones given above. Being Player 1, if your two cards are blue, you will make 3 points whatever the action of Player 2. True / False
9. The tables of gains ate the ones given above. Being Player 2, if the two cards are yellow, choosing action A will make you earn 10 points. True / False
10. If you make 60 points in total, you will earn 12 euros (+ 5 euros show-up fee). True / False

References

Ashlagi, I. and Y. A. Gonczarowski (2016): "Stable Matching Mechanisms are not Obviously Strategy-Proof," Working paper.

Bade, S. and Y. A. Gonczarowski (2017): "Gibbard-Satterthwaite Success Stories and Obvious Strategyproofness," Working paper.

Battaglini, M. and U. Makarov (2014): "Cheap Talk with Multiple Audiences: an Experimental Analysis," *Games and Economic Behavior*, 83, 147–164.

Benndorf, V., D. Kübler, and H.-T. Normann (2015): “Privacy Concerns, Voluntary Disclosure of Information, and Unraveling: an Experiment,” *European Economic Review*, 75, 43–59.

Bernheim, B. D. (1984): “Rationalizable Strategic Behavior,” *Econometrica*, 52, 1007–28.

Blume, A., D. V. DeJong, G. R. Neumann, and N. E. Savin (2002): “Learning and Communication in Sender-Reciever Games: An Econometric Investigation,” *Journal of Applied Econometrics*, 17, 225–247.

Blume, A., Y.-G. Kim, and G. B. Sprinkle (1998): “Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games,” *American Economic Review*, 88, 1323–40.

Brown, A. L., C. F. Camerer, and D. Lovallo (2012): “To Review or not to Review? Strategic Thinking at the Box Office,” *American Economic Journal: Microeconomics*, 4, 1–26.

Cai, H. and J. T.-Y. Wang (2006): “Overcommunication in strategic information transmission games,” *Games and Economic Behavior*, 56, 7–36.

Camerer, C. F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*, New York: Russel Sagee Foundation: Princeton University Press.

Camerer, C. F., T.-H. Ho, and J. K. Chong (2004): “A Cognitive Hierarchy Model of Games,” *Quarterly Journal of Economics*, 119, 861–98.

Costa-Gomes, M. A. and V. P. Crawford (2006): “Cognition and Behavior in Two-Person Guessing Games: An Experimental Study,” *American Economic Review*, 96, 1737–68.

Costa-Gomes, M. A., V. P. Crawford, and B. Broseta (2001): “Cognition and Behavior in Normal-Form Games: An Experimental Study,” *Econometrica*, 69, 1193–1235.

Crawford, V. and J. Sobel (1982): “Strategic Information Transmission,” *Econometrica*, 50, 1431–1451.

Crawford, V. P. (1995): “Adaptive Dynamics in Corodination Games,” *Econometrica*, 63, 103–43.

——— (2003): “Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentations of Intentions,” *American Economic Review*, 93, 133–149.

Crawford, V. P., M. A. Costa-Gomes, and N. Irriberi (2013): “Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence and Applications,” *Journal of Economic Literature*, 51, 5–62.

Dickhaut, J. W., M. Ledyard, A. Mukherji, and H. Sapra (2003): “Information Management and Valuation: An Experimental Investigation,” *Games and Economic Behavior*, 44, 25–53.

Dickhaut, J. W., K. A. McCabe, and A. Mukherji (1995): “An Experimental Study of Strategic Information Transmission,” *Economic Theory*, 6, 389–403.

Dranove, D. and G. Z. Jin (2010): "Quality Disclosure and Certification: Theory and Practice," *Journal of Economic Literature*, 48, 935–963.

Dye, R. A. (1985): "Disclosure of Nonproprietary Information," *Journal of Accouning Research*, 23, 123–145.

Forsythe, R., R. M. Isaac, and T. R. Palfrey (1989): "Theories and Tests of "Blind Bidding" in Sealed-Bid Auctions," *Rand Journal of Economics*, 20, 214–238.

Giovannoni, F. and D. J. Seidmann (2007): "Secrecy, Two-Sided Bias and the Value of Evidence," *Games and Economic Behavior*, 59, 296–315.

Grossman, S. J. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24, 461–483.

Hagenbach, J., F. Koessler, and E. Perez-Richet (2014): "Certifiable Pre-Play Communication: Full Disclosure," *Econometrica*, 82, 1093–1131.

Jin, G. Z., M. Luca, and D. Martin (2015): "Is No News (Perceived as) Bad News? An Experimental Investigation of Information Disclosure," Working paper.

King, R. R. and D. E. Wallin (1991a): "Market induced Information Disclosure: An Experimental Markets Investigation," *Contemporary Accounting Research*, 8, 170–197.

——— (1991b): "Voluntary Disclosure when Seller's Level of Information is Unknown," *Journal of Accounting Research*, 29, 96–108.

Lai, E. K., W. Lim, and J. T.-Y. Wang (2015): "An Experimental Analysis of Multidimensional Cheap Talk," *Games and Economic Behavior*, 91, 114–144.

Li, S. (2017): "Obviously Strategy-Proof Mechanisms," *American Economic Review*, 107, 3257–3287.

Luca, M. and J. Smith (2015): "Strategic Disclosure: the Case of Business School Rankings," *Journal of Economic Behavior and Organization*, 112, 17–25.

Mathis, J. (2008): "Full Revelation of Information in Sender-Receiver Games of Persuasion," *Journal of Economic Theory*, 143, 571–584.

Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391.

Nagel, R. (1995): "Unraveling in Guessing Games: An Experminental Study," *American Economic Review*, 85, 1313–26.

Okuno-Fujiwara, A., M. Postlewaite, and K. Suzumura (1990): "Strategic Information Revelation," *Review of Economic Studies*, 57, 25–47.

Pearce, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–50.

Pycia, M. and P. Troyan (2018): "Obvious Dominance and Random Priority," Working paper.

Rappoport, D. (2017): "Evidence and Skepticism in Verifiable Disclosure Games," Working paper.

Schipper, B. C. and Y. X. Li (2018): "Strategic Reasoning in Persuasion Games: an Experiment," Working paper.

Seidmann, D. J. and E. Winter (1997): "Strategic Information Transmission with Verifiable Messages," *Econometrica*, 65, 163–169.

Stahl, D. O. and P. W. Wilson (1994): "Experimental Evidence on Players' Models of Other Players," *Journal of Economic Behavior and Organizations*, 25, 309–27.

——— (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218–54.

Troyan, P. (2016): "Obviously Strategy-Proof Implementation of Top Trading Cycles," Working paper.

Van Zandt, T. and X. Vives (2007): "Monotone Equilibria in Bayesian Games of Strategic Complementarities," *Journal of Economic Theory*, 134, 339–360.

Vespa, E. and A. J. Wilson (2016): "Communication with Multiple Senders: An Experiment," *Quantitative Economics*, 7, 1–36.

Wang, J. T.-Y., M. Spezio, and C. F. Camerer (2010): "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games," *American Economic Review*, 100, 984–1007.

Zhang, L. and D. Levin (2015): "Bounded Rationality and Robust Mechanism Design: an Axiomatic Approach," *American Economic Review*, 107, 235–39.