Converging better response dynamics in sender-receiver games*

Stéphan Sémirat[†] and Françoise Forges[‡]

April 28, 2025

Abstract

We consider information transmission between a sender, who has finitely many types, and a receiver, who must choose a decision in a real interval. The payoffs depend on the sender's type and the receiver's decision. We assume that the payoff functions are well-behaved. We characterize the pure strategy perfect Bayesian equilibrium outcomes as incentive compatible partitions of the sender's types. We propose an algorithm, which starts from the finest partition. Then, at every step, if the current partition is not incentive compatible, a random type of the sender improves its payoff and the receiver best responds. We show that every possible run of the algorithm converges to a unique incentive compatible partition Π_*. This partition Π_* is such that any partition with more cells than Π_* is not incentive compatible, so that the algorithm determines to which extent information transmission can be effective. The partition Π_* also satisfies some refinement criteria for perfect Bayesian equilibria in sender-receiver games. Furthermore, in a discrete version of a popular class of examples (namely, if the sender's type is uniformly distributed and payoff functions are quadratic, with a constant upward bias for the sender), Π_* ex ante Pareto dominates every other incentive compatible partition.

MSC numbers: 91-08, 91A05, 91A10, 91A27.

JEL classification numbers: C72, D82, D83.

Keywords: algorithm, cheap talk, information transmission, neologism-proof equilibrium, undefeated equilibrium.

*This paper supersedes a previous one, entitled: "Forward-neologism proof equilibrium and better response dynamics". We wish to acknowledge useful comments from Andreas Blume, Archishman Chakraborty, Daniel Clark, Inga Deimen, Alexander Frug, Sidartha Gordon, Navin Kartik, Frédéric Koessler, Andrés Salamanca, Joel Sobel, Ina Taneva, Peter Vida, Leeat Yariv and anonymous referees.

[†]Université Grenoble Alpes, CNRS, INRAE, Grenoble INP, GAEL. E-mail: stephan.semirat@univ-grenoble-alpes.fr

[‡]Université Paris-Dauphine, PSL University, LEDa. E-mail: francoise.forges@gmail.com

1 Introduction

1.1 Summary

Among the assets that set effective decision-makers apart, information has emerged as fundamental in today's fast-paced, data-driven world. The ability to access, interpret, and act on reliable information is now essential for making optimal choices in a variety of contexts. The simplest situation involves two individuals: one of them is fully informed of the relevant variables (referred to as his "type"), the other must make a decision which matters to both. The situation is modelled as a sender-receiver game by allowing the informed individual to send a costless message (referred to as "cheap talk") to the decision-maker. If the latter is Bayesian, he forms a prior belief over the possible types and updates it as a function of the sender's message. There is always a "babbling" equilibrium, in which the sender adopts the same strategy whatever his type and the receiver makes an optimal decision given his prior. More interestingly, effective information transmission, in which the receiver's posterior differs from his prior, is always beneficial to the receiver, but crucially relies on the sender's incentives. The basic question, when solving a sender-receiver game, is whether it has (possibly refined) Nash equilibria in which information transmission is effective. If indeed such equilibria exist, the next question is whether they can be compared in terms of the quantity of information that is transmitted and/or the players' payoffs. These questions cannot be addressed in a systematic way (i.e., beyond nice examples) without assuming strong relationships between the players' payoffs.[1]

In a celebrated article, Crawford and Sobel (1982) introduce a model in which the sender's types and the receiver's actions both belong to a real interval. Among other assumptions, for every type, the players' payoff functions are concave, with a unique maximum; the sender is "upward biased", in the sense that, for every type, the action that is best for the sender is always higher than the receiver's optimal one. Crawford and Sobel (1982) provide a full characterization of perfect Bayesian Nash equilibrium (henceforth, PBE, a mild refinement of Nash equilibrium) outcomes. In a parametrized example known as the "uniform quadratic case", and more generally under an ad hoc "regularity condition"[2], they also identify a "most informative" PBE, which satisfies tailored refinements of PBE and ex ante Pareto dominate all other PBE. A huge number of papers builds on Crawford and

[1] As explained in (e.g.) Forges (2020), the seminal study of Aumann et al. (1968) (see also Aumann et al. (1995)), completed by Hart (1985), already illustrates the ambiguous effects of cheap talk on examples with finitely types and actions.

[2] As made precise below, the "uniform quadratic case" denotes a family of examples in which the sender's type is uniformly distributed over the unit interval while the payoff functions are quadratic and indexed by a constant positive real parameter, interpreted as the sender's "bias". The "regularity condition" is a monotonicity property of "pre-solutions", which is not formulated in terms of the parameters of the model, but is always satisfied in the uniform quadratic case.

Sobel (1982), but to the best of our knowledge, the effectiveness of information transmission, or more generally, the quantity of information transmitted at equilibrium, has only been assessed under assumptions guaranteeing a straight comparison of PBE outcomes, like in the uniform quadratic case. Furthermore, as suggested in our brief account of the literature below, only a few papers study the convergence of gradual optimization processes in Crawford and Sobel (1982) (see Olszewski (2022) and Gordon et al. (2022)).

In the current paper, we analyze a class of sender-receiver games in which, exactly as in the model of Crawford and Sobel (1982) (without any "regularity condition"), the receiver's action is a real number and the players' payoff functions are well-behaved, with an upward bias for the sender. What is much less standard in our games is that, as in Frug (2016), the sender has *finitely many types*, which can be ranked according to some order. The latter assumption makes sense when types correspond to a qualitative ranking (e.g., "excellent," "good," "fair," etc.) or when computational considerations call for discrete variables.[3] In a consistent way, we assume that the sender's set of costless messages is also finite (and at least as large as his set of types).

We propose a gradual optimization process, better denoted as an algorithm, which can be applied to any game of the class of under consideration. The algorithm possibly entails random choices at every step, but, against all odds, always converges to the same equilibrium outcome, whatever the random path followed. The bottom line is that the algorithm provides a test to determine whether effective information transmission can possibly take place at a PBE (without computing the equilibria). This test is a by-product of a much stronger property: a maximal quantity of information is transmitted at the limit equilibrium outcome reached by the algorithm. Therefore, if the algorithm converges to the "babbling" equilibrium outcome, effective information transmission cannot take place at any PBE. Moreover, the limit reached by the algorithm also satisfies some classical equilibrium selection criteria for sender-receiver games.

Our algorithm is simple enough to be described in words. To do so, we first characterize the (pure strategy) PBE outcomes as incentive compatible (IC) partitions of the sender's set of types, by making use of its finiteness. Every cell of such a partition consists of a subset of types that all send the same message at equilibrium. The receiver best responds by choosing, for every cell, the action (indeed unique under our assumptions) that maximizes his updated expected payoff, given that the sender's type belongs to that cell. The no-deviation condition of the sender takes the form of an IC condition, namely, no type can benefit from pretending to be in a cell that is not the one containing it. Given our assumptions on payoff functions, the IC partitions are "interval" partitions, in which cells

[3]See, e.g., Babichenko et al. (2023) and Condorelli and Furlan (2023) for computer programs applied to cheap talk. Our model involves a continuum of actions, but, as made clear below, we can restrict on a finite number of them.

are ordered and consist of subsets of consecutive types. While the previous characterization applies as soon as the set of messages is at least as large as the set of types, it can also be turned into a canonical representation of (pure strategy) PBE outcomes by introducing canonical messages, which consist of all nonempty subsets of types L, interpreted as "my type is in L". At a canonical equilibrium, as in the well-known "revelation principle" of mechanism design, the sender is truthful. However, in our context, he does not necessarily fully reveal his type.

The algorithm starts at the most informative partition, in which every cell consists of a single type. This partition is gradually modified, in a minimal way at every step. If, assuming that the receiver naively reacts as if the sender were telling the truth, the completely revealing partition is IC, the algorithm stops immediately. Otherwise, there is *some* sender's type that would improve its payoff by pretending to be another type. In this case, our assumptions on the payoff functions imply that the maximal type of *some* cell would prefer to be in the next cell on its right. At the next stage, such a type is taken out of its cell to join the adjacent cell it envies. This modifies two cells of the current partition, with one of them possibly becoming empty. The modified partition can be immediately interpreted as a strategy of the sender and the receiver can easily adjust his best response to the new partition, namely, on maximum two cells (no reply being necessary on the possibly empty one). The iterative process goes on by alternating *better* replies of the sender of the kind just described (in which a single type improves its payoff) and best responses of the receiver (which require an adjustment on maximum two cells).[4] At every step, a new partition is derived, but flexible as it is, the procedure could possibly lead to cycles or multiple limit partitions. We prove that *all runs of the previous algorithm converge to the same IC partition* Π_*.

The main property of the limit partition Π_* is that it is "as informative as possible", in the sense that a partition with a larger number of cells than Π_* cannot be IC. Equivalently, the number of actions of the receiver at a PBE cannot be larger than in Π_*. In particular, if Π_* is the coarsest partition (consisting of a single cell), effective information transmission is impossible.

Our characterization of PBE as IC partitions easily accommodates the *neologism-proof* equilibrium of Farrell (1993). Given an IC partition, every subset L of types that is not a cell of this partition appears as a neologism with respect to the partition. Proceeding as in Farrell (1993), a neologism L is credible if the types in L are the only ones that benefit from sending message L, taking for granted that the receiver interprets L as sent by types in L but sticks to the status quo when he does not receive message L. Existence of a neologism-proof equilibrium is a well-known issue and our framework is no

[4] All along the algorithm, the actions of the receiver are chosen among finitely many, namely, those that maximize his expected payoff, given that the sender's type belongs to some subset L, according to the probability distribution derived by Bayes rule.

exception. However, we show that Π_* is the only partition that can possibly be neologism-proof.

Mailath et al. (1993) argue that, if the receiver takes a neologism seriously, he should also revise his interpretation of the original equilibrium messages, an inference that can be understood by the sender. This kind of logic leads them to introduce another refinement of PBE, the *undefeated equilibrium*, for a class of games in which the informed player's message can be costly (i.e., signaling games). Being interested in this more general framework, they do not explicitly refer to neologisms. Doing so and using the IC partition characterization, the definition of their solution concept takes a much simpler form in our sender-receiver games: an IC partition is defeated by another IC partition if there is a cell of the latter (interpreted as a neologism for the former) that is preferred to the former by all types in this cell. In general, the notion introduced by Mailath et al. (1993) is neither weaker nor stronger than Farrell's (1993) one. We prove that the partition Π_* reached by our algorithm is undefeated. This not only guarantees the existence of an IC undefeated partition in our model but also provides a way to find such a partition via a simple algorithm, without making the list of all IC partitions and compare them pairwise, which looks like a tedious task. Another consequence of our results is that, in our model, every neologism-proof IC partition is undefeated, but the scope of this statement is limited by the fact that there may be no neologism-proof IC partition at all.

Finally, in a discrete version of Crawford and Sobel's (1982) leading class of examples, namely, if the receiver's prior belief is uniform over $\{1, 2, ..., N\}$ and the payoff functions are quadratic, with a constant bias for sender, Π_* ex ante Pareto dominates every other IC partition. To prove this, we show that the partition Π_* reached by our algorithm coincides with the one characterized by Frug (2016) under the same assumptions.

Before describing the organization of our paper, we briefly pursue the discussion of the relevant literature.

1.2 Related papers

As far as our model is concerned, the only difference with Crawford and Sobel's (1982) sender-receiver game is that the sender has finitely many types, a model first analyzed by Frug (2016). The latter paper shows that, as one would expect, the methodology used for a continuum of types does not apply to finitely many. For instance, when types belong to a real interval, IC conditions reduce to equalities to be satisfied at "cutoff types", while, in the discrete case, inequalities cannot be dispensed with. As a consequence, two different partitions might have identical cells in the discrete case, a property that does not hold in the continuous case.

Frug (2016) characterizes the ex ante Pareto dominant PBE outcome in the N type-version of the uniform quadratic case. He shows that this PBE outcome corresponds to a partition $\Pi(N)$ and gives an explicit way to compute $\Pi(N)$, which he refers to as "procedure 1", without further justification. As already mentioned above, we prove that $\Pi(N)$ coincides with the partition Π_* reached by our algorithm. This gives some foundation to Frug's (2016) partition and shows that Π_* corresponds to an ex ante Pareto dominant PBE.[5]

Sémirat and Forges (2022) consider sender-receiver games with finitely many types and well-behaved payoff functions, as in the current paper, but in which decisions necessitate the sender's approval. They establish the existence of an "equilibrium without exit" by means of converging algorithms but do not investigate the possible uniqueness of the limit equilibrium determined in this way.

As suggested in the summary above, notwithstanding the impact of Crawford and Sobel (1982), very few papers propose operational procedures (e.g., dynamic, converging ones) reaching PBE outcomes in well-behaved sender-receiver games. It is true that, in the uniform quadratic case (with a continuum of types), which is the focus of many papers, PBE outcomes are indeed easily computed and characterized as a function of the sender's (upward) bias. Motivated by *equilibrium selection*, a few recent papers (Gordon (2011), Lo and Olszewski (2022), and Gordon et al. (2022)[6]) study specific adaptive learning processes in Crawford and Sobel's (1982) sender-receiver games, typically satisfying the regularity condition.[7] By contrast, Olszewski (2022) does not require any such condition and establishes that, whatever the starting point, the best response dynamics converges to *some equilibrium*, without identifying any particular property of the limit.

To be more precise on Gordon et al.'s (2022) contribution, the authors make use of an interim best response dynamic to identify a "smallest" and a "largest" equilibrium, which correspond to the limit of a "lower" and an "upper" sequence respectively. Under the regularity condition, the two extreme equilibria coincide with the one that induces the larger number of actions in Crawford and Sobel (1982). To establish this result, Gordon et al. (2022) start by assuming that the set of messages is finite, even if the sender's type belongs to a real interval. This assumption has no effect on the set of PBE of their game but prevents the sender from using a completely revealing strategy. Gordon et al. (2022) make the further assumption that messages are ordered. Furthermore, they focus on monotonic strategies

[5]Frug (2016) also demonstrates that, if, by contrast to our assumptions, the sender's bias does not always go in the same direction, equilibria in mixed strategies may be optimal for the receiver.

[6]In the sequel, we focus on Gordon et al. (2022), our understanding being that this paper subsumes Gordon (2011) and Lo and Olszewski (2022), as well as another paper on equilibrium refinement, Kartik and Sobel (2015).

[7]With a very different motivation, Condorelli and Furlan (2024) also study the uniform quadratic version of Crawford and Sobel's (1982) sender-receiver game. They compute stationary points of memoryless independent learning algorithms playing this game, which requires to discretize the set of types as well as the set of actions. Their simulations demonstrate convergence to the most informative equilibrium.

for both players. This assumption allows for a tractable representation of actions as a function of types in their framework but is not needed in ours.

In spite of an apparent similarity between the previous result of Gordon et al. (2022) and ours, there are a number of fundamental differences. When the set of types is finite, a finite set of messages is in order and enables the sender to completely reveal his type. This is the starting point of our algorithm, which thus does not directly apply when there is a continuum of types and finitely messages.

Gordon et al.'s (2022) dynamic process is driven by interim best responses of the sender and best responses of the receiver, which potentially lead to a major change of the outcome (namely, actions as a function of types) at every step. By contrast, our algorithm only makes use of specific *better* responses of the sender, in which a single (appropriately chosen) type improves its payoff. We do consider best responses for the receiver but given the minimal change in the sender's strategy at every step, the receiver's strategy is also minimally modified at every step. When there is a continuum of types, the interval partitions summarizing the outcome are described in terms of the limit points of the successive intervals, denoted as "cutoffs" (see, e.g., Olszewski (2022) and Gordon et al. (2022)). Using this terminology, a typical step of our algorithm simply shifts a single cutoff to the right.

Gordon et al. (2022) relate the limit of their dynamic with a selection criterion introduced by Chen et al. (2008), "no incentive to separate" (NITS): if there is a unique equilibrium outcome satisfying NITS, a property that holds under the regularity condition, every interim best response sequence converges to an equilibrium with this outcome. We show that the limit partition Π_* satisfies NITS. Gordon et al. (2022) further show that, in their model, a procedure of iterated deletion of weakly dominated strategies leads to the same equilibrium selection as their best response dynamics. To identify dominated strategies, Gordon et al. (2022) need to keep track of a monotonic version of the strategic form game. By contrast, what makes our analysis tractable is the formulation in terms of outcomes (using partitions of the set of types) rather than strategies.

The NITS criterion is tailored to Crawford and Sobel's (1982) model. Several other equilibrium refinements have been proposed for sender-receiver games (see Kreps and Sobel (1994) for a general presentation and classic references), among which the concepts described in the summary above, namely, the neologism-proof equilibrium of Farrell (1993) and the undefeated equilibrium of Mailath et al. (1993). Mailath et al. (1993) identify a class of signaling games (containing Spence's model as a representative one) in which an undefeated equilibrium always exists; not surprisingly, our model does not pertain to this class.

Matthews et al. (1991) develop various notions of "announcement-proof (mixed) equilibrium", one of which (the "weak" one) is motivated by the same considerations as undefeated equilibria, namely, consistency of off-path messages with possible sender's equilibrium strategies. Their framework is sender-receiver games with finitely many types, finitely many actions and arbitrary payoff functions. They assume that a *rich* language is available to the sender, allowing him to make complex "announcements", with an intended interpretation for the receiver. By contrast, in a similar way as in the revelation principle, our canonical representation emphasizes the outcome of information transmission, namely, relies on elementary messages to be interpreted literally by the receiver, without trying to enter the complexity of the underlying language. While, somehow surprisingly, our approach can be reconciled with some standard treatments of neologisms in sender-receiver games, it is quite distinct from contributions in which meaning and language are essential (see, e.g., Blume (2023) and the references therein).

A recent paper by Clark (2021) illustrates that Farrell's (1993) concept is still vivid. Clark (2021) introduces the notion of "credible robust neologisms" for general signaling games. Having extended Farrell's (1993) definition to this class of games, he shows that neologism-proof equilibria are always robust neologism-proof. Unlike the incentive compatible neologisms behind undefeated equilibria, credible robust neologisms maintain the idea that the receiver sticks to his equilibrium strategy when he does not receive the neologism. The main novelty of Clark's (2021) notion is that, when the receiver observes the neologism L, he can form *any* belief over L. Hence a set of best responses of the receiver is associated with L, one for every belief. The neologism L is robustly credible if all types in L benefit from sending it, whatever the receiver's best response. By contrast, in an undefeated equilibrium, the receiver implicitly uses Bayes rule over credible neologisms, because these are sent at some (other) equilibrium.[8]

Clark (2021) establishes existence of a robust neologism-proof equilibrium in two popular classes of signaling games. However, in arbitrary sender-receiver games with finitely many types and actions, there may not be any robust neologism-proof equilibrium, including when there exists an undefeated equilibrium. It may also happen that a robust neologism-proof equilibrium is defeated.

1.3 Organization

Section 2 describes the model, namely, the sender-receiver game, and the solution concept, perfect Bayesian equilibrium (PBE). Proposition 1 characterizes pure PBE as incentive compatible (IC) partitions of the set of types. Section 3 provides a first description of our algorithm, converging to a unique IC partition Π_*. To this aim, we introduce the notions of envy and left-incentive compatibility. Example 1 illustrates the tractability of the algorithm compared to the interim best response dynamic.

Section 4 is devoted to the game theoretic properties of the limit partition Π_*. We first recall

[8]Hillas (1994) already proposes to refine sequential equilibrium in general extensive form games by requiring that off equilibrium path beliefs correspond to some equilibrium beliefs.

Farrell's (1993) notions of self-signaling set and neologism-proof equilibrium. Proposition 2 states that, if there exists a neologism-proof partition, it must coincide with Π_*. Then we turn to Mailath et al.'s (1993) undefeated equilibrium. Proposition 3 states that no IC partition can defeat Π_*. More generally, "pseudo-partitions", which, as we explain, characterize mixed PBE, cannot defeat Π_* either (Proposition 4). Section 4 goes on with Proposition 5 – Π_* satisfies NITS ("No incentive to Separate", Chen et al. (2008)) – and Proposition 6 – in the uniform quadratic case, Π_* corresponds to a PBE that ex ante Pareto dominates any other PBE.

Section 5 formally describes an algorithm generalizing the one introduced in Section 3, all runs of which are shown to converge to the partition Π_*. A key property is that Π_* dominates *every* other IC partition, according to a binary relation that we introduce, which formalizes the idea that a partition dominates another if its cells are more to the right. Theorem 1 summarizes the properties of Π_* that derive from its construction.

In Section 6, we establish the game theoretic properties of Π_* stated in Section 4. The proof of Proposition 3 enlightens the reason why Π_* is undefeated : in every other IC partition Π, the highest type of every cell would prefer to be in Π_* rather than in Π. The proof of Proposition 2 shows that Π_* is the only partition that can possibly satisfy a weaker property than neologism-proofness, which can be denoted as NSSHT for "No Self-Signaling of the Highest Types".

2 Model and solution concept

In this section, we describe the basic game and characterize its perfect Bayesian equilibrium (PBE) outcomes.

2.1 Sender-receiver game

We consider a sender-receiver game in which the sender's set of types Θ is finite and ordered: $\Theta = \{\theta_1, \ldots, \theta_N\}$, with $\theta_1 < \ldots < \theta_N$, where N is a positive integer. The prior probability distribution over Θ, $p \in \Delta(\Theta)$, is such that $p(\theta) > 0$ for every $\theta \in \Theta$. The sender's set of messages M is finite and such that $|M| \geq |\Theta|$. The receiver's set of decisions is $\mathbb{R}$. For every $\theta \in \Theta$, the sender's payoff function is $U^\theta : \mathbb{R} \to \mathbb{R}$ and the receiver's payoff function is $V^\theta : \mathbb{R} \to \mathbb{R}$.

The game, which we denote as $\mathrm{SR}(M)$, unfolds as follows: a chance move selects a type θ in Θ according to p; the sender is informed of θ and sends a message $m \in M$ to the receiver, who then chooses a decision $x \in \mathbb{R}$. The players' respective utilities are $U^\theta(x)$, $V^\theta(x)$, independently of the message m.

We assume that the functions $(\theta, x) \mapsto U^\theta(x)$ and $(\theta, x) \mapsto V^\theta(x)$ are well-behaved, namely, satisfy the following standard properties:

- *Strict concavity*:

$$\text{For every } \theta \in \Theta, \text{ the payoff functions } U^\theta \text{ and } V^\theta \text{ are twice continuously differentiable and for every } x \in \mathbb{R}, \ \frac{\partial^2 U^\theta(x)}{\partial x^2} < 0 \text{ and } \frac{\partial^2 V^\theta(x)}{\partial x^2} < 0. \tag{A0}$$

- *Single-crossing*:

$$\begin{aligned} &\text{For every } (\theta_1, \theta_2, x_1, x_2) \in \Theta^2 \times \mathbb{R}^2, \text{ with } \theta_2 > \theta_1 \text{ and } x_2 > x_1, \\ &\text{if } U^{\theta_1}(x_2) - U^{\theta_1}(x_1) \geq 0, \text{ then } U^{\theta_2}(x_2) - U^{\theta_2}(x_1) > 0, \text{ and} \\ &\text{if } V^{\theta_1}(x_2) - V^{\theta_1}(x_1) \geq 0, \text{ then } V^{\theta_2}(x_2) - V^{\theta_2}(x_1) > 0. \end{aligned} \tag{A1}$$

- *Unique maximizing arguments*:

$$\begin{aligned} &\text{For every } \theta \in \Theta, \text{ there exist a unique } x^*(\theta) \in \mathbb{R} \text{ and a unique } y^*(\theta) \in \mathbb{R} \\ &\text{such that } \left.\frac{\partial U^\theta(x)}{\partial x}\right|_{x=x^*(\theta)} = 0 \text{ and } \left.\frac{\partial V^\theta(x)}{\partial x}\right|_{x=y^*(\theta)} = 0. \end{aligned} \tag{A2}$$

- *Sender's upward bias*:

$$\text{For every } \theta \in \Theta, \ x^*(\theta) > y^*(\theta). \tag{A3}$$

As an illustration, the previous assumptions are satisfied by quadratic payoff functions of the form

$$U^\theta(x) = -(\theta + b_\theta - x)^2 \text{ and } V^\theta(x) = -(\theta - x)^2 \tag{quad}$$

with $b_\theta > 0$ and $\theta + b_\theta$ increasing with θ.

2.2 Perfect Bayesian equilibrium

A pure strategy for the sender (resp., receiver) in SR(M) is a mapping $\sigma : \Theta \to M$ (resp., $\tau : M \to \mathbb{R}$). The corresponding *outcome* is $\tau \circ \sigma : \Theta \to \mathbb{R}$. Henceforth, PBE refers to perfect Bayesian equilibrium in pure strategies.[9] Below we characterize the PBE outcomes of SR(M) in terms of "incentive compatible partitions" of Θ.

Let us define, for every $L \subseteq \Theta$ and every $x \in \mathbb{R}$:

$$V^L(x) = \sum_{\theta \in L} \frac{p(\theta)}{p(L)} V^\theta(x).$$

From assumptions (A0) and (A2), we can set

$$y^L = \arg\max_{x \in \mathbb{R}} V^L(x).$$

[9] See Section 4.2 for some hints on equilibria in mixed strategies.

In particular, for every $\theta \in \Theta$, the receiver's optimal action knowing θ is

$$y^{\theta} \stackrel{\text{def}}{=} y^{\{\theta\}} = y^*(\theta).$$

Using (A1), the maximizers $x^*(\theta_i)$ and $y^*(\theta_i) = y^{\theta_i}$, $i = 1, \ldots, N$, can be ordered with respect to the original order on Θ.

For every partition Π of Θ, $\pi \subseteq \Theta$ denotes any cell of Π; for every $\theta \in \Theta$, $\pi(\theta)$ denotes the cell of Π that contains θ. The *outcome* associated with partition Π is the mapping : $\Theta \to \mathbb{R} : \theta \mapsto y^{\pi(\theta)}$.

Definition 1. Let Π be a partition of Θ; Π is *incentive compatible* (IC) if

$$\text{For every } \theta \in \Theta, \text{ for every cell } \pi \text{ of } \Pi,\ U^{\theta}(y^{\pi(\theta)}) \geq U^{\theta}(y^{\pi}). \tag{IC}$$

Note that, from our assumptions, every IC partition is an "interval" partition, in which every cell consists of consecutive types that are lower than the types in the next one. To see this, let Π be an IC partition, let $\theta < \theta' < \theta''$, assume that θ and θ'' are in the same cell π of Π, with $y^{\pi} = y$, while θ' is in another cell π' of Π, with $y^{\pi'} = y' \neq y$. By (A1) and (IC) for θ and θ', we must have that $y < y'$. But given (A1), this contradicts (IC) for θ''.

The coarsest partition

$$\text{NR} = \{\Theta\}$$

is always IC and corresponds to the "babbling" PBE. At the other extreme, the sender's completely revealing strategy induces the finest partition

$$\text{CR} = \{\{\theta_1\}, \{\theta_2\}, \cdots, \{\theta_N\}\}.$$

Given the receiver's best response, namely, y^{θ_i} when the sender's type is θ_i, CR is not necessarily IC.

Proposition 1. *Every PBE of* $\text{SR}(M)$ *induces an IC partition of* Θ *with the same outcome. Conversely, for every IC partition of* Θ*, there exists a PBE of* $\text{SR}(M)$ *with the same outcome. Furthermore, for every IC partition of* Θ*, there exists a* truthful *PBE of* $\text{SR}(2^{\Theta}\backslash\{\varnothing\})$ *with the same outcome.*

Proof. Let (σ, τ) be a PBE of $\text{SR}(M)$. The sender's strategy σ induces a partition of Θ, namely, $\Pi(\sigma)$, with cells $\sigma^{-1}(m)$, $m \in \sigma(\Theta)$. Given σ and a message $m \in \sigma(\Theta)$, the receiver forms the posterior belief $p(\cdot \mid m) \in \Delta(\Theta)$ according to Bayes rule. He best responds to σ by choosing $\tau(m)$ so as to maximize his expected payoff given his posterior belief. Equivalently, for every cell π of the partition $\Pi(\sigma)$, he forms the posterior belief $p(\cdot \mid \pi)$ (equal to $\frac{p(\theta)}{p(\pi)}$ for $\theta \in \pi$) and makes the decision y^{π}. The IC condition for $\Pi(\sigma)$ expresses that, if the receiver uses τ, type θ cannot benefit from sending any message $m \in \sigma(\Theta)$ possibly different from $\sigma(\theta)$.

Let Π be an IC partition. Given a set of messages M such that $|M| \geq |\Theta|$, we can associate a different message $m = m(\pi)$ with every cell π of Π and set $\sigma(\theta) = m(\pi(\theta))$ for every $\theta \in \Theta$, so that $\Pi = \Pi(\sigma)$. Let then $\tau(m)$ be defined as y^{π} if $m = m(\pi)$ and as y^{π_0} otherwise, where π_0 is some cell of Π; the associated belief is $p(\cdot \mid \pi_0)$. The pair of strategies (σ, τ) defines be a PBE of $\mathrm{SR}(M)$.

Consider now the game $\mathrm{SR}(2^{\Theta}\backslash\{\varnothing\})$, in which the messages are the nonempty subsets L of Θ (and are thus canonically defined from the primitive set Θ). The intended meaning of message L is "my type is in L". Starting again from an IC partition Π, let the sender be truthful, namely let every type θ send the message $\pi(\theta)$. Let the receiver choose action y^L if message L is a cell of Π and choose y^{L_0}, for some arbitrary cell L_0 of Π, otherwise. Proceeding as in the previous paragraph, the latter strategies form a PBE of $\mathrm{SR}(2^{\Theta}\backslash\{\varnothing\})$.[10] $\square$

As an immediate corollary of Proposition 1, every PBE outcome of the original game, $\mathrm{SR}(M)$ – in which $|M| \geq |\Theta|$ – can be canonically represented as a *truthful* PBE outcome of $\mathrm{SR}(2^{\Theta}\backslash\{\varnothing\})$. In a canonical equilibrium, the receiver trusts the sender on the equilibrium path; by contrast, he does not interpret neologisms (i.e., off path messages) literally. Indeed, at this point, no condition guarantees that the sender could credibly send such a neologism (conceivable credibility conditions will be considered in Section 4). In the next sections, we formulate the analysis in terms of partitions, without referring to any sender-receiver game.

3 Basic algorithm, limit partition

This section starts with two definitions that will facilitate the step by step description of the algorithm outlined in the Introduction. A formal, slightly more general, description is proposed in Section 5.

We focus on type-ordered interval partitions, in which the cells are ordered (from left to right) according to the order on Θ.

Definition 2. Let Π be a partition of Θ and let π be a cell of Π.
Type θ *envies* cell π if $U^{\theta}(y^{\pi}) > U^{\theta}(y^{\pi(\theta)})$.
Type θ *envies* type θ' if θ envies cell $\pi(\theta')$.

In terms of Definition 2, the partition Π is IC if and only if it has no envying type.

Definition 3. A partition is *Left*-Incentive Compatible (L-IC) if no type envies a cell on its left.

[10] The fact that pure equilibria *induce* a partition of the sender's types is trivial. Proposition 1 states a stronger result, namely, that IC partitions *coincide* with equilibrium outcomes. The proof shows that the only assumption among (A0)–(A3) that is needed for the result is Assumption (A2) for the receiver, namely that he has a unique maximizing action for every belief.

Basic algorithm:[11]

- **Step 0**: Start with $\Pi_0 = \mathrm{CR} = \{\{\theta_1\}, \{\theta_2\}, \cdots, \{\theta_N\}\}$. Π_0 is L-IC.
- **Step 1**: If Π_0 is not IC, there is some type θ_i that envies some type on its right; then $\widetilde{\theta_0} = \theta_i$ also envies θ_{i+1}; merge θ_i and θ_{i+1} into $\{\theta_i, \theta_{i+1}\}$ to form Π_1. Π_1 is L-IC.
- $\cdots$
- At the end of **Step r**, Π_r is L-IC.
- **Step r+1**: If Π_r is not IC, there is some type, in some cell of Π_r, $\theta \in \pi_r^k$, that envies some cell on its right. Then the largest type in π_r^k, $\tilde{\theta}_r = \max \pi_r^k$, envies the next cell π_r^{k+1}; move $\tilde{\theta}_r$ from π_r^k to π_r^{k+1}, to form Π_{r+1}.
- $\cdots$
- A *unique IC partition* Π_* is reached, whatever the cell π_r^k chosen at every step $r+1$.

An illustration is provided below.

Example 1. Let $\Theta = \{1, \ldots, 11\}$ and p be uniform, namely, $p(\theta) = \frac{1}{11}$ for every θ. Let the payoff functions satisfy (quad), with a constant bias $b = 2$.

Here is a possible run of the algorithm:

- $\{\{1\}, \{2\}, \cdots, \{10\}, \{11\}\}$
- $\{\{1\}, \{2\}, \cdots, \{10, 11\}\}$
- $\cdots$
- $\{\{1\}, \{2\}, \{3\}, \{4\}, \{5, \cdots, 11\}\}$
- $\{\{1\}, \{2\}, \{3, 4\}, \{5, \cdots, 11\}\}$
- $\{\{1\}, \{2\}, \{3\}, \{4, \cdots, 11\}\}$
- $\{\{1\}, \{2, 3\}, \{4, \cdots, 11\}\}$
- $\{\{1, 2, 3\}, \{4, \cdots, 11\}\}$
- $\{\{1, 2\}, \{3, \cdots, 11\}\} = \Pi_*$

[11] All claims will be established in Section 5.

Let us have a closer look at the computations involved in the above run of the algorithm. At the first step, the receiver's best response to the sender's completely revealing strategy is computed. Then, at every step, the cells of the current partition are considered from the right, until a maximal type of a cell envies the next cell. Denoting the partition reached at some step as $\Pi = \{\pi^1, \ldots, \pi^{n-2}, \pi^{n-1}, \pi^n\}$, for some n in $\{1, \ldots, 11\}$, the payoff of type $\max \pi^{n-1}$ under decision $y^{\pi^{n-1}}$ is compared with its payoff under y^{π^n}. If the latter is higher than the former, $\max \pi^{n-1}$ moves to π^n. Otherwise, cell π^{n-2} is considered, with $\max \pi^{n-2}$ possibly moving to π^{n-1}, and so on. For instance, when $\{\{1\}, \{2\}, \{3, 4\}, \{5, \ldots, 11\}\}$ is reached, type 4 envies the next cell because type 4 gets a higher payoff with decision 8 (associated with $\{5, \ldots, 11\}$) than with decision 3.5 (associated with $\{3, 4\}$). In general, when the reached partition has $n = 4$ cells, then the next partition will be determined from the previous one by checking at most $n - 1 = 3$ inequalities. Furthermore, every partition is obtained by moving a single type to the adjacent cell, so that every partition can be *immediately interpreted* as a truthful strategy of the sender, in the same way as the initial partition CR, without requiring any further inference. The best response of the receiver has to be computed at every step, but this involves updating it on two "new" cells at most, no reply being needed on an empty cell.

The simplicity of the algorithm is further demonstrated by comparing it with a run of the *interim best response dynamic* in the previous example. The first step is the same as above. Then the sender's interim best response σ_1 has to be computed, which requires to determine, for *every type* i, $i = 1, \ldots, 11$, the payoff associated with *every decision* y^j, $j = 1, \ldots, 11$, and then pick the one giving the highest payoff. The result of this optimization is $\sigma_1(i) = \{i + 2\}$ for $i = 1, \cdots, 9$, $\sigma_1(10) = \sigma_1(11) = \{11\}$. Some inference is needed to determine the next partition from strategy σ_1; one can proceed as in the proof of Proposition 1: the next partition is $\{\{1\}, \{2\}, \cdots, \{8\}, \{9, 10, 11\}\}$. The receiver's optimal decision can then be updated, namely computed on the (only) new cell $\{9, 10, 11\}$. In general, however, there is no obvious way to predict the number of cells on which the receiver's best response has to be computed. For instance, there is a run of the interim best response dynamic that, at some step, say, $r - 1$, reaches the partition $\Pi_{r-1} = \{\{1\}, \{2\}, \{3\}, \{4, 5, 6\}, \{7, \cdots, 11\}\}$. An interim best response of the sender is $\sigma_r(1) = \sigma_r(2) = \{3\}$, $\sigma_r(3) = \sigma_r(4) = \sigma_r(5) = \{4, 5, 6\}$ and $\sigma_r(i) = \{7, \cdots, 11\}$ for $i = 6, \ldots, 11$, leading to the partition $\Pi_r = \{\{1, 2\}, \{3, 4, 5\}, \{6, \cdots, 11\}\}$, in which all cells differ from the ones of Π_{r-1}. Summing up, a typical step of the interim best response dynamic requires to compare the current expected payoff of *every type* with the expected payoff it would obtain under the action associated with *every cell*, to deduce the next partition from the sender's interim best response and finally, to compute the receiver's best response on *every* cell of the next partition.

As suggested by the previous example, it appears that several variants of the algorithm are possible, differing in how many types are moved at each step — from moving one type at a time as in our algorithm, to moving the maximal number of types simultaneously as in the interim best response dynamic. A natural question is to identify a computationally efficient number of types to move per step. In our view, no simple solution exists. Moving a type requires to undertake a test for envy. Hence increasing the number of types to be moved before updating the actions increases the number of tests to be undertaken at every step. While this has an advantage in saving computations of the updated actions, this can also increase the number of *failing tests* for envy. A failing test is inefficient, because tests have to be undertaken again once the actions are updated. Roughly speaking, the issue of an envy test for type $\tilde{\theta}_r$ depends on the relative distance between the sender's ideal action, $x^*(\tilde{\theta}_r)$, and the actions $y^{\pi_r^k}$ and $y^{\pi_r^{k+1}}$ associated with the cells $\pi_r^k \ni \tilde{\theta}_r$ and π_r^{k+1} in the current partition. For instance, under quadratic utilities (quad), the test succeeds if $\left(x^*(\tilde{\theta}_r) - y^{\pi_r^k}\right)^2 > \left(x^*(\tilde{\theta}_r) - y^{\pi_r^{k+1}}\right)^2$. Given $y^{\pi_r^{k+1}} > y^{\pi_r^k}$, this condition is met whenever $x^*(\tilde{\theta}_r) > y^{\pi_r^{k+1}}$. Thus, the greater the sender's bias, measured by $x^*(\tilde{\theta}_r) - y^*(\tilde{\theta}_r) > 0$ and, if type $\tilde{\theta}_r$ is among the greatest types of its cell π_r^k, bounded from below by $x^*(\tilde{\theta}_r) - y^{\pi_r^k}$, the more efficient it appears to move multiple types simultaneously. In general, however, the issue of a test may differ across types. Moreover, as the algorithm progresses, the actions move further apart and the previous inequalities are less likely to hold, making the test increasingly prone to failure. Consequently, there is no method that can be considered *a priori* more efficient at every stage of the process.

4 Game theoretic properties of the limit partition

The statements of this section show that the limit partition Π_* identified in the previous section performs well with respect to equilibrium selection criteria that can have power in sender-receiver games. The proofs, which make extensive use of the characterization of Π_* through the algorithm, are given in Section 6.

4.1 Neologism-proof partition

Let us start with Farrell's (1993) neologism-proof equilibrium. As seen in Section 2.2, in our setup, every PBE can be characterized as an IC partition Π. The associated language consists of the cells π of Π, to be interpreted as "my type is in π". Every set of types $L \subseteq \Theta$ that is not a cell of Π can be interpreted as a *neologism*.

The next definition is borrowed from Farrell (1993).[12]

[12]Farrell (1993) does not explicitly settle the case of types that are indifferent between their original equilibrium

Definition 4. Let Π be an IC partition of Θ and let L be a subset of Θ.

- L is *self-signaling* for Π if
  - for every $\theta \in L, \quad U^\theta(y^L) \geq U^\theta(y^{\pi(\theta)})$ (with at least one $>$);
  - for every $\theta \notin L, \quad U^\theta(y^L) \leq U^\theta(y^{\pi(\theta)})$.
- Π is *neologism-proof* if there is no self-signaling set for Π.

Example 2 below illustrates that, unsurprisingly, there may be no neologism-proof IC partition in our model. However, the partition Π_* is the only one that could have this property.

Proposition 2. *If there exists a neologism-proof IC partition* Π*, then* $\Pi = \Pi_*$*.*

4.2 Undefeated partition

To be self-signaling for a given equilibrium Π, a neologism L must improve the payoff of all types in L, with respect to Π, if the receiver reacts to L by making decision y^L. While this looks like a natural requirement, the second one, i.e., the types that are *not* in L should *not* benefit, with respect to Π, from sending message L, is more problematic. Indeed, to keep Π as a benchmark, one implicitly assumes that the receiver sticks to his original equilibrium strategy when a message different from L is sent. This assumption is disputable, as argued, e.g., by (Mailath et al., 1993, p. 252):

"[...] starting from a given equilibrium, adjusting the beliefs at some out-of-equilibrium information set cannot be done without simultaneously adjusting beliefs at other information sets, including some information sets on the equilibrium path. [...] Once all the subsequent adjustments are made, we must be at an equilibrium; if not, some further adjustment should be contemplated."

These insights lead Mailath et al. (1993) to compare equilibria with each other and to check whether a PBE is "defeated" by *another PBE*. Using our partition characterization, their notion of "undefeated" PBE can be simplified as follows:

Definition 5. Let Π and Π' be IC partitions of Θ.

- Π' *defeats* Π if there is a cell π' of Π' such that
  - for every $\theta \in \pi', \quad U^\theta(y^{\pi'}) \geq U^\theta(y^{\pi(\theta)})$ (with at least one $>$).
- Π is *undefeated* if Π is not defeated by any other IC partition.

Our partition characterization makes it possible to rephrase the previous definition in terms of neologisms. To see this, let Π and L be as in Definition 4; say that L is an *incentive compatible*

message and the neologism. In Definition 4, these types may send either message.

neologism if L satisfies the first requirement to be self-signaling and, in addition to this, is a cell of some IC partition Π'. Then a partition Π is undefeated if and only if there is no incentive compatible neologism for Π.

When the payoff functions are not well-behaved in the sense of Section 2 (namely, do not satisfy assumptions (A0)-(A3)), existence of an undefeated partition is not guaranteed, as can be seen on examples proposed, e.g., by Matthews et al. (1991) and Olszewski (2006).[13] However, the next result holds in our model:

Proposition 3. *The partition Π_* is undefeated.*

A trivial consequence of the two previous propositions is that, under our assumptions, an IC partition that is neologism-proof is undefeated, a property that does not hold in general.[14]

Example 1 (continued)**.** We saw that $\Pi_* = \{\{1, 2\}, \{3, ..., 11\}\}$. The other IC partitions are $\mathrm{NR} = \{\{1, \cdots, 11\}\}$ and $\Pi = \{\{1\}, \{2, \cdots, 11\}\}$. NR is defeated by Π (since type 1 prefers Π to NR) but not by Π_* (since type 2 and type 4 prefer NR to Π_*). However, Π is defeated by Π_* (since type 1 and type 2 prefer Π_* to Π).[15]

Undefeated partitions are further illustrated in the next example.

Example 2. Let $\Theta = \{1, 2, 3, 4\}$, $p(1) = \frac{1}{4}$, $p(2) = \frac{1}{4}$, $p(3) = \frac{1}{100}$, $p(4) = \frac{49}{100}$, and let the payoff functions satisfy (quad), with a constant bias $b = 0.6$.

The associated cell-contingent actions are:
$y^{\{1\}} = 1$, $y^{\{1,2\}} = 1.5$, $y^{\{1,2,3\}} = 1.53$, $y^{\{2\}} = 2$, $y^{\{2,3\}} = 2.04$, $y^{\{1,2,3,4\}} = 2.74$, $y^{\{3\}} = 3$, $y^{\{2,3,4\}} = 3.32$, $y^{\{3,4\}} = 3.984$, and $y^{\{4\}} = 4$.

There are eight interval partitions of Θ. As clear from Figure 1, in $\{\{1\}, \{2\}, \{3\}, \{4\}\}$ and $\{\{1\}, \{2\}, \{3, 4\}\}$, type 1 envies type 2. In $\{\{1\}, \{2, 3\}, \{4\}\}$, $\{\{1, 2\}, \{3\}, \{4\}\}$ and $\{\{1, 2, 3\}, \{4\}\}$, type 3 envies type 4. The partitions $\mathrm{NR} = \{\{1, 2, 3, 4\}\}$, $\Pi = \{\{1\}, \{2, 3, 4\}\}$ and $\Pi_* = \{\{1, 2\}, \{3, 4\}\}$ are the only IC ones. The partitions Π and Π_* are both "maximally informative", a phenomenon that can also arise in Crawford and Sobel's (1982) model in absence of the regularity condition (see

[13] In Example 6 of Matthews et al. (1991), there are two partially revealing IC partitions which defeat each other; both of them defeat the nonrevaling partition. In Example 5 of Olszewski (2006), there are five IC partitions; the completely revealing partition is defeated by the nonrevealing one, which in turn is defeated by any of the partially revealing ones; these defeat each other as in a three-person majority game.

[14] Consider a sender-receiver game with two types (θ_1 and θ_2) and three actions for the receiver in which both NR and CR are IC, both types prefer CR to NR but the neologism $\{\theta_i\}$ $(i = 1, 2)$ is not self-signaling according to Definition 4, because type θ_j, $j \neq i$, would benefit from it.

[15] In this example, Π_* turns out to be the only undefeated partition. However if, e.g., $|\Theta| = 8$, the IC partitions are NR and $\Pi_* = \{\{1\}, \{2, ..., 8\}\}$. Both are undefeated.

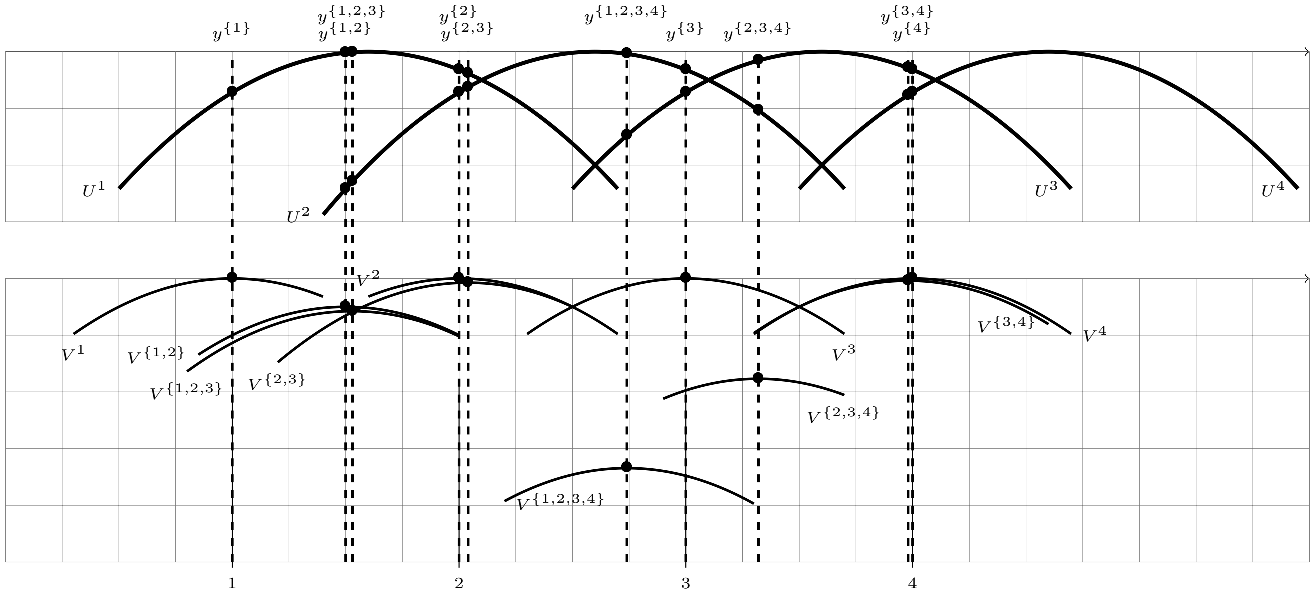

Figure 1: Payoff functions and cell contingent actions in Example 2.

Olszewski (2022)). The partition NR is defeated both by Π (thanks to the IC neologism $L = \{1\}$) and Π_* (thanks to the IC neologism $L = \{3, 4\}$). Concerning Π and Π_*, types 1 and 4 prefer partition Π_*, whereas types 2 and 3 prefer partition Π. No type except 2 prefers NR to Π_* or Π. Thus Π_* and Π are both undefeated. None of them is neologism-proof, as $L = \{4\}$ is self-signaling.

Remark: mixed strategies

A stronger version of Proposition 3 holds, namely, the partition Π_* remains undefeated even if the sender is allowed to randomize. A mixed strategy for the sender in SR(M) is a mapping $\sigma : \Theta \to \Delta(M)$, where $\Delta(M)$ denotes the set of probability distributions over M. At a PBE, given σ and a message m that is sent with positive probability by at least one type, the receiver updates his belief over Θ and chooses his best action, which according to assumptions (A0) and (A2), is uniquely defined (and can be computed in the same way as y^L above). In other words, the receiver's best response is pure. By contrast, the sender may randomize in a best response, provided he is indifferent between the various messages he sends and cannot benefit from sending a message of zero probability. Using assumptions (A0)-(A2) for the sender, mixed equilibria (including the pure ones studied up to now) can be represented by IC "pseudo - (interval) partitions" of the form

$$\Pi = \{\{\theta_1, \ldots, \theta_{i_1}\}, \{\theta'_{i_1}, \ldots, \theta_{i_2}\}, \ldots, \{\theta'_{i_k}, \ldots, \theta_N\}\},$$

in which the maximal element of the jth cell (θ_{i_j}) possibly coincides with the minimal element of the next one (θ'_{i_j}), with the understanding that a type belonging to two different pseudo-cells gets the same payoff in both of them.[16]

[16]This characterization is easily deduced from Frug's (2016) analysis. A full description includes the probability

Definition 5 readily extends to the case of an IC partition Π and an IC pseudo-partition Π'. Using this definition, we will show that

Proposition 4. *The partition Π_* is not defeated by any IC pseudo-partition.*

4.3 NITS

As a third selection criterion, let us turn to Chen et al.'s (2008) "No Incentive To Separate" (NITS), which requires that the lowest type, namely θ_1, has no incentives to signal itself (if somehow he could). Applied readily to our setting, the notion becomes:

Definition 6. Let Π be an IC partition of Θ in which the first cell, namely, $\pi(\theta_1)$, contains at least two elements. Π satisfies NITS if

$$U^{\theta_1}(y^{\theta_1}) \leq U^{\theta_1}(y^{\pi(\theta_1)}).$$

In the model of Crawford and Sobel (1982), under the already mentioned "regularity condition", NITS has a remarkable selection power. In general, as acknowledged by Chen et al. (2008), it looks like a rather weak requirement.

Proposition 5. *The partition Π_* satisfies NITS.*

4.4 Ex ante optimal PBE

As indicated in Section 1.2, Frug (2016) describes a specific partition, which, using the peculiarities of the uniform quadratic case, can be explicitly computed through what he calls "procedure 1". He shows that this partition corresponds to a (pure) PBE that ex ante Pareto dominates all PBE (including the mixed ones). We show that the partition Π_* coincides with Frug's (2016) partition, so that the next statement holds.

Proposition 6. *In the uniform quadratic case (namely, $\Theta = \{1, 2, ..., N\}$, $p(\theta) = \frac{1}{N}$ for every θ and payoff functions satisfy (quad) with a constant bias b), the partition Π_* corresponds to an ex ante Pareto-optimal PBE.*

The next example illustrates that Proposition 6 no longer holds if types are not uniformly distributed.

Example 3. Assume $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and

- θ_1 envies $\{\theta_2\}$ but not $\{\theta_2, \theta_3\}$;

distribution used by every mixing type $\theta_{i_j} = \theta'_{i_j}$ as well as the action y^{π} chosen by the receiver on every pseudo-cell π.

- θ_2 does not envy $\{\theta_3\}$, even when it is associated with θ_1, and does not envy θ_1 when associated with θ_3.

The partitions $\Pi_* = \{\{\theta_1, \theta_2\}, \{\theta_3\}\}$ and $\Pi = \{\{\theta_1\}, \{\theta_2, \theta_3\}\}$ are IC and Π is defeated by Π_*. However, the ex ante expected payoff of both players is higher in Π than in Π_* if, e.g., $b = 1$ and

- $\theta_1 = 1$, $\theta_2 = 2.5$, $\theta_3 = 6.5$,
- $p(\theta_1) = 0.9$, $p(\theta_2) = 0.0873$, $p(\theta_3) = 0.0127$.

5 General algorithm and further properties

In this section, we establish the main properties of our algorithm, formally stated as Theorem 1 at the end of the section. We first make fully precise the description sketched in Section 3, making use of Definitions 2 and 3. Then we introduce an "envy operator", denoted as "Env". Given a set E of L-IC partitions, which typically can be reached at some step r of some run of the algorithm as described in Section 3, $\mathrm{Env}(E)$ contains all partitions that can be reached at the next step. Starting from *any* L-IC partition Π_0 and applying the envy operator recursively, we reach, after finitely many steps, *the* set $\overline{\mathrm{Env}}(\Pi_0)$ of all IC partitions that can be derived from Π_0 by envy. We then show that, if Π_0 is the completely revealing partition, the set $\overline{\mathrm{Env}}(\Pi_0)$ contains a *single* IC partition, namely, the partition previously denoted as Π_*. The key property of the completely revealing partition is that it "dominates" every IC partition, according to a binary relation that we introduce in Section 5.2.

5.1 Applying "envy" to L-IC partitions

Let $\Pi = \{\pi^1, \ldots, \pi^n\}$, $n \leq N$, be an interval type-ordered partition, i.e., such that the types in π^1 are lower than the types in π^2, and so on. Such a partition Π is not IC if and only if there is some type in some cell π^k who envies some cell $\pi^{k'}$, $k' \neq k$. Denoting as $\max L$ (resp., $\min L$) the highest (resp., lowest) element of a subset of types L and using the single-crossing condition, this is equivalent to: if $k' > k$ (resp., $k' < k$), type $\max \pi^k$ envies π^{k+1} (resp., type $\min \pi^k$ envies π^{k-1}). In other words, a partition is IC if and only if no type at an edge of any cell envies an adjacent cell. In the same way, a partition is L-IC if and only if no minimal type of any cell envies the adjacent cell on its left.

Definition 7. Given a set of L-IC partitions E, $\mathrm{Env}(E)$ is the set of partitions derived from partitions Π in E as follows:

- if Π is IC, then add Π to $\mathrm{Env}(E)$;
- if $\Pi = \{\pi^1, \ldots, \pi^n\}$, $n \in \{1, \ldots, N\}$, is not IC, then for every cell π^k, $k \in \{1, \ldots, n-1\}$, in which type $\max \pi^k$ envies cell π^{k+1}:

– if π^k is a singleton, then add partition $\{\pi^1, \dots, \pi^{k-1}, \{\max \pi^k\} \cup \pi^{k+1}, \dots, \pi^n\}$ to $\mathrm{Env}(E)$;

– if π^k is not a singleton, then add partition $\{\pi^1, \dots, \pi^{k-1}, \pi^k \setminus \{\max \pi^k\}, \{\max \pi^k\} \cup \pi^{k+1}, \dots, \pi^n\}$ to $\mathrm{Env}(E)$.

The next result states that, while reducing the envy that types may have for cells on their right, the Env operator preserves L-IC. With a slight abuse of notation, we write $\mathrm{Env}(\Pi)$ for $\mathrm{Env}(\{\Pi\})$.

Lemma 1. *If* Π *is L-IC, then every partition in* $\mathrm{Env}(\Pi)$ *is also L-IC.*

Proof. Let Π be an L-IC partition. If Π is IC, and hence L-IC, then, by definition, $\mathrm{Env}(\Pi) = \{\Pi\}$ and the lemma trivially holds. Otherwise, write $\Pi = \{\pi^1, \dots, \pi^n\}$, with $n = |\Pi|$, and let $\Pi_1 \in \mathrm{Env}(\Pi)$, with $\Pi_1 = \{\pi_1^1, \dots, \pi_1^{n_1}\}$, $n_1 = |\Pi_1|$, in which, starting from Π, type $\tilde{\theta} \in \pi^{\tilde{k}}$, $\tilde{k} \in \{1, \dots, n-1\}$ has been moved to its next succeeding cell. In particular, if $\pi^{\tilde{k}} \setminus \{\tilde{\theta}\} \neq \varnothing$, then we have $n_1 = n$ and

$$\begin{cases} \pi_1^1 = \pi^1, \dots, \pi_1^{\tilde{k}-1} = \pi^{\tilde{k}-1}, \\ \pi_1^{\tilde{k}} = \pi^{\tilde{k}} \setminus \{\tilde{\theta}\}, \\ \pi_1^{\tilde{k}+1} = \pi^{\tilde{k}+1} \cup \{\tilde{\theta}\}, \\ \pi_1^{\tilde{k}+2} = \pi^{\tilde{k}+2}, \dots, \pi_1^n = \pi^n, \end{cases}$$

and, if $\pi^{\tilde{k}} \setminus \{\tilde{\theta}\} = \varnothing$, then $n_1 = n - 1$ and

$$\begin{cases} \pi_1^1 = \pi^1, \dots, \pi_1^{\tilde{k}-1} = \pi^{\tilde{k}-1}, \\ \pi_1^{\tilde{k}} = \pi^{\tilde{k}+1} \cup \{\tilde{\theta}\}, \\ \pi_1^{\tilde{k}+1} = \pi^{\tilde{k}+2}, \dots, \pi_1^{n-1} = \pi^n. \end{cases}$$

We have to show that for every $k \in \{2, \dots, n_1\}$, type $\theta = \min \pi_1^k$ does not envy its immediately preceding cell π_1^{k-1}. Since Π is L-IC, the property holds for every k such that π_1^k and π_1^{k-1} are unchanged from Π to Π_1. This occurs for every $k \in \{2, \dots, \tilde{k}-1, \tilde{k}+3, \dots, n\}$ and, if $n_1 = n-1$, for $k = \tilde{k}+2$ too. Thus it remains to show that the property also holds for $k \in \{\tilde{k}, \tilde{k}+1, \tilde{k}+2\}$ in case $n_1 = n$, and for $k \in \{\tilde{k}, \tilde{k}+1\}$ in case $n_1 = n-1$. These five cases can be described by the three cases in which θ is the minimal type of either $\pi^{\tilde{k}} \setminus \{\tilde{\theta}\}$ (and $n = n_1$), $\pi^{\tilde{k}+1} \cup \{\tilde{\theta}\}$, or $\pi^{\tilde{k}+2}$. We distinguish these three cases below. Before that, let us note that from $\tilde{\theta} = \max \pi^{\tilde{k}}$ and $\tilde{\theta} = \min(\pi^{\tilde{k}+1} \cup \{\tilde{\theta}\})$, the newly induced actions in Π_1, namely $y^{\pi^{\tilde{k}} \setminus \{\tilde{\theta}\}}$ (if $n_1 = n$) and $y^{\pi^{\tilde{k}+1} \cup \{\tilde{\theta}\}}$ satisfy

$$y^{\pi^{\tilde{k}} \setminus \{\tilde{\theta}\}} < y^{\pi^{\tilde{k}}} < y^{\pi^{\tilde{k}+1} \cup \{\tilde{\theta}\}} < y^{\pi^{\tilde{k}+1}}.$$

- If $n_1 = n$ and $\theta = \min \pi^{\tilde{k}} \setminus \{\tilde{\theta}\}$, then we can assume that $\tilde{k} \geq 2$ since otherwise there is no cell on the left of θ. Then the cell on the left of θ in Π_1 is $\pi^{\tilde{k}-1}$, and we have to show $U^\theta(y^{\pi^{\tilde{k}} \setminus \{\tilde{\theta}\}}) > U^\theta(y^{\pi^{\tilde{k}-1}})$.

- If $y^{\pi^{\tilde{k}}\setminus\{\tilde{\theta}\}} \geq x^*(\theta)$, then $x \mapsto U^\theta(x)$ is decreasing at $y^{\pi^{\tilde{k}}\setminus\{\tilde{\theta}\}}$, and therefore at $y^{\pi^{\tilde{k}}} > y^{\pi^{\tilde{k}}\setminus\{\tilde{\theta}\}}$ too. Thus $U^\theta(y^{\pi^{\tilde{k}}\setminus\{\tilde{\theta}\}}) > U^\theta(y^{\pi^{\tilde{k}}})$. Since Π is L-IC, we also have $U^\theta(y^{\pi^{\tilde{k}}}) \geq U^\theta(y^{\pi^{\tilde{k}-1}})$, and the desired inequality follows.
- If instead $y^{\pi_1^{\tilde{k}}} < x^*(\theta)$, then since θ is the minimal type of $\pi^{\tilde{k}} \setminus \{\tilde{\theta}\}$, we have $y^{\pi^{\tilde{k}-1}} < y^{\pi^{\tilde{k}}\setminus\{\tilde{\theta}\}} \leq y^\theta$, where $x \mapsto U^\theta(x)$ is increasing. Then the desired inequality also follows.

- Next, we consider the case in which $\theta = \tilde{\theta} = \min(\pi^{\tilde{k}} \cup \{\tilde{\theta}\})$. Since $\theta = \tilde{\theta}$ has been moved by envy, we have $U^\theta(y^{\pi^{\tilde{k}+1}}) > U^\theta(y^{\pi^{\tilde{k}}})$. In particular, since $y^{\pi^{\tilde{k}+1}} > y^{\pi^{\tilde{k}}}$, function $x \mapsto U^\theta(x)$ is not decreasing at $y^{\pi^{\tilde{k}}}$, i.e. $y^{\pi^{\tilde{k}}} \leq x^*(\theta)$. Since $\theta = \tilde{\theta}$ is the minimal type of its cell $\pi^{\tilde{k}} \cup \{\tilde{\theta}\} \neq \{\tilde{\theta}\}$ in Π_1, we also have $y^\theta < y^{\pi^{\tilde{k}}\cup\{\tilde{\theta}\}}$. Therefore $y^\theta < y^{\pi^{\tilde{k}}\cup\{\tilde{\theta}\}} < y^{\pi^{\tilde{k}}} \leq x^*(\theta)$. Moreover, the action y associated with the cell on the left of $\pi^{\tilde{k}} \cup \{\tilde{\theta}\}$ in Π_1, i.e. associated with $\pi^{\tilde{k}} \setminus \{\tilde{\theta}\}$ if $n_1 = n$ or with $\pi^{\tilde{k}-1}$ if $n_1 = n - 1$, is such that $y < y^\theta$. Then we have $y < y^\theta < y^{\pi^{\tilde{k}}\cup\{\tilde{\theta}\}} \leq x^*(\theta)$, where $x \mapsto U^\theta(x)$ is increasing, and we obtain $U^\theta(y^{\pi^{\tilde{k}}\cup\{\tilde{\theta}\}}) > U^\theta(y)$ as desired.
- Finally, if $\theta = \min \pi^{\tilde{k}+2}$, then the cell on the left of θ is $\pi^{\tilde{k}+1} \cup \{\tilde{\theta}\}$. We have $y^{\pi^{\tilde{k}+1}\cup\{\tilde{\theta}\}} < y^{\pi^{\tilde{k}+1}} < y^{\pi^{\tilde{k}+2}} \leq y^\theta$, where $x \mapsto U^\theta(x)$ is increasing. In particular, $U^\theta(y^{\pi^{\tilde{k}+1}\cup\{\tilde{\theta}\}}) < U^\theta(y^{\pi^{\tilde{k}+1}})$. Since Π is L-IC, we also have $U^\theta(y^{\pi^{\tilde{k}+2}}) \geq U^\theta(y^{\pi^{\tilde{k}+1}})$, and therefore $U^\theta(y^{\pi^{\tilde{k}+2}}) > U^\theta(y^{\pi^{\tilde{k}+1}\cup\{\tilde{\theta}\}})$ as desired.

□

Using Lemma 1, we recursively define, for any L-IC partition Π, $\text{Env}^0(\Pi) = \{\Pi\}$, and, for every $r \in \mathbb{R}$,

$$\text{Env}^{r+1}(\Pi) = \text{Env}(\text{Env}^r(\Pi)).$$

Since there are finitely many types, moving a type to the adjacent cell on its right cannot be done indefinitely. Lemma 1 thus guarantees that every sequence $(\text{Env}^r(\Pi))_{r\geq 0}$ converges to a set of *IC* partitions in a finite number of steps. Formally,

Proposition 7. *For every L-IC partition* Π, *there exists* $\overline{r} \in \mathbb{N}$ *such that every partition* $\Pi_{\overline{r}} \in \text{Env}^{\overline{r}}(\Pi)$ *is IC.*

Furthermore, from the definition of $\text{Env}(\cdot)$, if E is a set of IC partitions, $\text{Env}(E) = E$. Hence we can define

$$\overline{\text{Env}}(\Pi) = \text{Env}^{\overline{r}}(\Pi)$$

as *the* set of IC partitions derived from an L-IC partition Π by envy.

5.2 Dominance relation over interval partitions

The next definition expresses that a partition Π dominates another partition, Π', if the cells of Π are "more to the right", namely, if every maximal element of a cell of Π' is pooled with higher types in Π than Π'.

Definition 8. Let Π and Π' be interval partitions of Θ.

Π *dominates* Π' ($\Pi \geq \Pi'$) if for every cell π' of Π', $\min \pi(\max \pi') \geq \min \pi'$.

See Figure 2 below for an illustration. According to our previous notation, $\min \pi'$ and $\max \pi'$ denote, respectively, the lowest and the highest element of cell π', while $\pi(\max \pi')$ is the cell of Π which contains $\max \pi'$ and $\min \pi(\max \pi')$ is the lowest element of this cell.

$$\begin{array}{ll} \Pi' : & \ldots\}, \quad \{\min \pi', \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \max \pi'\}, \{\ldots \\ \Pi : & \qquad\quad \ldots\ldots\ldots\}, \underbrace{\{\min \pi(\max \pi'), \ldots\ldots, \max \pi', \ldots\ldots\}}_{\pi(\max \pi')}, \{\ldots \end{array}$$

Figure 2: $\Pi \geq \Pi'$

The dominance relation is anti-symmetric.[17] If partitions Π and Π' are such that $\Pi \geq \Pi'$, Π' cannot have more cells than Π.[18] Furthermore, the completely revealing partition $\text{CR} = \{\{\theta_1\}, \ldots, \{\theta_N\}\}$ dominates every other partition, while the nonrevealing partition $\text{NR} = \{\Theta\}$ is dominated by every other partition.

5.3 Dominance and envy

The following lemma shows how the envy operator (see Definition 7) preserves dominance (see Definition 8).

Lemma 2. *For every L-IC partition* Π_0 *and every* $\Pi_1 \in \text{Env}(\Pi_0)$*, if* $\Pi_0 \geq \Pi$ *for some IC partition* Π*, then* $\Pi_1 \geq \Pi$.

Proof. Let Π be an IC partition, let Π_0 be an L-IC partition such that $\Pi_0 \geq \Pi$, and let $\Pi_1 \in \text{Env}(\Pi_0)$.

Write $\Pi = \{\pi^1, \ldots, \pi^n\}$, with $n = |\Pi|$, and choose $\overline{\theta}_k = \max \pi^k$ for some $k \in \{1, \ldots, n\}$. From $\Pi_0 \geq \Pi$, we have

$$\min \pi_0(\overline{\theta}_k) \geq \min \pi^k, \tag{1}$$

[17] If $\Pi \geq \Pi'$ and $\Pi' \geq \Pi$, then Π and Π' have identical rightmost cells. By induction along the remaining rightmost cells of Π and Π', one can show that $\Pi = \Pi'$.

[18] If $\Pi \geq \Pi'$, by induction, starting at the rightmost cells, one can show that no cell of Π might contain two types of the form $\min \pi'$, with $\pi' \in \Pi'$. There are $|\Pi'|$ such types, and thus $|\Pi| \geq |\Pi'|$.

and we need to show

$$\min \pi_1(\overline{\theta}_k) \geq \min \pi^k, \tag{2}$$

where $\pi_0(\overline{\theta}_k)$ (resp., $\pi_1(\overline{\theta}_k)$) denotes the cell of Π_0 (resp., Π_1) that contains $\overline{\theta}_k$.

Denoting as $\tilde{\theta}$ the type that is moved to its next succeeding cell from Π_0 to Π_1,

- either $\pi_1(\overline{\theta}_k) = \pi_0(\overline{\theta}_k)$, in which case (2) derives from (1);
- or $\overline{\theta}_k = \tilde{\theta}$, in which case $\overline{\theta}_k$ is the lowest type of $\pi_1(\overline{\theta}_k)$, and then (2) follows from $\min \pi_1(\overline{\theta}_k) = \overline{\theta}_k = \max \pi^k \geq \min \pi^k$;
- or $\pi_1(\overline{\theta}_k) = \pi_0(\overline{\theta}_k) \cup \{\tilde{\theta}\}$, where $\tilde{\theta}$ is the highest element of the cell $\pi_0^L(\overline{\theta}_k)$, immediately on the left of $\pi_0(\overline{\theta}_k)$.

In the latter case, $\min \pi_1(\overline{\theta}_k) = \tilde{\theta}$. Then (2) can be written as

$$\tilde{\theta} \geq \min \pi^k. \tag{3}$$

Inequality (3) trivially holds whenever $\min \pi_0(\overline{\theta}_k) > \min \pi^k$. We will establish that this is always the case, namely: if $\tilde{\theta} \in \pi_0^L(\overline{\theta}_k)$ envies $\pi_0(\overline{\theta}_k)$, then $\min \pi_0(\overline{\theta}_k) > \min \pi^k$.

We show this by contradiction. Suppose $\min \pi_0(\overline{\theta}_k) = \min \pi^k$. Then $\tilde{\theta}$, the maximal element of $\pi_0^L(\overline{\theta}_k)$, is also the maximal element of cell π^{k-1}, i.e.,

$$\tilde{\theta} = \overline{\theta}_{k-1}.$$

On the one hand, from $\Pi_0 \geq \Pi$, we have

$$\min \pi_0^L(\overline{\theta}_k) \geq \min \pi^{k-1},$$

and $\pi_0^L(\overline{\theta}_k)$ consists of the highest elements of π^{k-1}. On the other hand, since $\max \pi^k = \overline{\theta}_k \in \pi_0(\overline{\theta}_k)$, we also have that π^k consists of the lowest elements of $\pi_0(\overline{\theta}_k)$. This, together with the (IC) condition on Π, prevents $\tilde{\theta}$ from strictly envying $\pi_0(\overline{\theta}_k)$ when considered in $\pi_0^L(\overline{\theta}_k)$.

Formally, we have

$$y^{\pi^{k-1}} \leq y^{\pi_0^L(\overline{\theta}_k)} \leq y^{\tilde{\theta}}, \tag{4}$$

where the second inequality follows from the maximality of $\tilde{\theta}$ in cell $\pi_0^L(\overline{\theta}_k)$, and

$$y^{\pi_0(\overline{\theta}_k)} \geq y^{\pi^k}. \tag{5}$$

From (4), $U^{\tilde{\theta}}$ is increasing at $y^{\pi^{k-1}}$ and $y^{\pi_0^L(\overline{\theta}_k)}$, so that

$$U^{\tilde{\theta}}(y^{\pi^{k-1}}) \leq U^{\tilde{\theta}}(y^{\pi_0^L(\overline{\theta}_k)}). \tag{6}$$

Since Π is IC, $U^{\tilde{\theta}}(y^{\pi^{k-1}}) \geq U^{\tilde{\theta}}(y^{\pi^k})$. Then, from $y^{\pi^{k-1}} < y^{\pi^k}$, $U^{\tilde{\theta}}$ is not increasing at y^{π^k}, and (5) implies

$$U^{\tilde{\theta}}(y^{\pi^{k-1}}) \geq U^{\tilde{\theta}}(y^{\pi_0(\overline{\theta}_k)}).$$

This, combined with (6), gives

$$U^{\tilde{\theta}}(y^{\pi_0(\overline{\theta}_k)}) \leq U^{\tilde{\theta}}(y^{\pi_0^L(\overline{\theta}_k)}),$$

and $\tilde{\theta} \in \pi_0^L(\overline{\theta}_k)$ does not envy $\pi_0(\overline{\theta}_k)$. □

5.4 Convergence of the algorithm

By applying Lemma 2 inductively, to an L-IC partition Π_0 and an IC partition Π such that $\Pi_0 \geq \Pi$, we obtain that every partition $\Pi_r \in \mathrm{Env}^r(\Pi_0)$ is such that $\Pi_r \geq \Pi$. In particular, recalling Proposition 7 and the definition of $\overline{\mathrm{Env}}(\Pi_0)$, every partition $\Pi_{\overline{r}} \in \overline{\mathrm{Env}}(\Pi_0)$ is IC and such that $\Pi_{\overline{r}} \geq \Pi$. This result is summarized in the next statement.

Corollary 1. *Let Π_0 be an L-IC partition and Π be an IC partition, such that $\Pi_0 \geq \Pi$. Every partition $\Pi_{\overline{r}} \in \overline{\mathrm{Env}}(\Pi_0)$ is IC and such that $\Pi_{\overline{r}} \geq \Pi$.*

Moreover, if an L-IC partition Π_0, such as CR, satisfies $\Pi_0 \geq \Pi$ for *every* IC partition Π, then every partition $\Pi_{\overline{r}} \in \overline{\mathrm{Env}}(\Pi_0)$ is such that $\Pi_{\overline{r}} \geq \Pi$ for *every* IC partition Π. Hence there exists at least one IC partition which dominates every other partition. But from the anti-symmetry of the dominance relation, such a dominating partition is necessarily unique. We thus obtain the following characterization of the unique limit partition reached by the algorithm when the initial L-IC partition Π_0 dominates every IC partition, e.g., $\Pi_0 = \mathrm{CR}$.

Theorem 1. *There exists a unique IC partition Π_* such that, for every IC partition Π, $\Pi_* \geq \Pi$. The partition Π_* is achieved as the unique element of $\overline{\mathrm{Env}}(\Pi_0)$, recursively derived by envy from any L-IC partition Π_0 such that for every IC partition, $\Pi_0 \geq \Pi$ (e.g., $\Pi_0 = \mathrm{CR}$).*

An immediate consequence of this statement is that a partition that is finer than Π_* cannot be IC.

One can check that, as expected, in Examples 1 and 2, Π_* dominates the other IC partitions, namely, NR and another one, denoted as Π in both examples (see Example 1 (continued)).

6 Proof of the results of Section 4

In this section, we establish that the partition Π_* identified in Theorem 1 satisfies desirable properties. We start with the most interesting one, namely, Π_* is undefeated (Proposition 3), because it implies that an undefeated PBE always exists in the class of games we consider.

6.1 Undefeated partition

To prove that no IC partition can possibly defeat Π_*, we will show that the payoff of the maximal element $\overline{\theta}$ of any cell of any IC partition Π is at least as high in Π_* as in Π. The next lemma states a preliminary step, namely, that the payoff of $\overline{\theta}$ is at least as high in $\Pi_r \in \mathrm{Env}^r(\Pi_0)$ as in Π, under the condition that $\overline{\theta}$ is also the maximal element of some cell of Π_0. This condition obviously holds for $\Pi_0 = \mathrm{CR}$.

Lemma 3. *Let Π be an IC partition, let π be a cell of Π, and let $\overline{\theta} = \max \pi$. Let Π_0 be an L-IC partition such that $\Pi_0 \geq \Pi$ and such that $\overline{\theta} = \max \pi_0$ for some cell $\pi_0 \in \Pi_0$. Let $r \in \mathbb{N}$ and let $\Pi_r \in \mathrm{Env}^r(\Pi_0)$. Then*

$$U^{\overline{\theta}}(y^{\pi_r(\overline{\theta})}) \geq U^{\overline{\theta}}(y^{\pi}),$$

with equality only if $\pi_r(\overline{\theta}) = \pi$.

Proof. Let $\Pi = \{\pi^1, \ldots, \pi^n\}$ be an IC partition, with $n = |\Pi|$, let $k \in \{1, \ldots, n\}$, and set $\overline{\theta}_k = \max \pi^k$. Let Π_0 be an L-IC partition such that $\Pi_0 \geq \Pi$ and such that $\overline{\theta}_k = \max \pi_0$ for some cell $\pi_0 \in \Pi_0$. Let $r \in \mathbb{N}$ and let $\Pi_r \in \mathrm{Env}^r(\Pi_0)$.

From $\overline{\theta}_k = \max \pi^k$, we have $y^{\pi^k} \leq y^{\overline{\theta}_k}$. In particular, $U^{\overline{\theta}_k}$ is increasing on $\left[y^{\pi^k}, y^{\overline{\theta}_k}\right]$. Then if $\pi_r(\overline{\theta}_k)$ is a singleton, the lemma follows (with equality only if π^k is a singleton too). Hence we can assume that $\pi_r(\overline{\theta}_k)$ is not a singleton for the remainder of the proof.

By Corollary 1, $\Pi_r \geq \Pi$. Hence the types of $\pi_r(\overline{\theta}_k)$ which are lower than or equal to $\overline{\theta}_k$ are the highest types of π^k and the lowest types of $\pi_r(\overline{\theta}_k)$. More precisely, let $L = \{\theta \in \pi_r(\overline{\theta}_k) : \theta \leq \overline{\theta}_k\}$. Equivalently, $L = \{\theta : \min \pi_r(\overline{\theta}_k) \leq \theta \leq \overline{\theta}_k\}$, so that $y^{\pi_r(\overline{\theta}_k)} \geq y^L \geq y^{\pi^k}$. In particular,

$$y^{\pi_r(\overline{\theta}_k)} \geq y^{\pi^k}. \tag{7}$$

Moreover, if $\pi_r(\overline{\theta}_k) \neq \pi^k$, the inequality is strict.

If $\overline{\theta}_k = \max \pi_r(\overline{\theta}_k)$, then $y^{\pi_r(\overline{\theta}_k)} \leq y^{\overline{\theta}_k}$, so that $U^{\overline{\theta}_k}$ is increasing on $\left[y^{\pi^k}, y^{\pi_r(\overline{\theta}_k)}\right]$, and the lemma also follows.

Otherwise, since Π_0 is such that $\overline{\theta} = \max \pi_0$, there exists $0 \leq r' < r$ and some cell $\pi_{r'} \in \Pi_{r'}$, $\Pi_{r'} \in \mathrm{Env}^{r'}(\Pi_0)$, whose maximal element is $\overline{\theta}_k$, and such that

$$U^{\overline{\theta}_k}(y^{\pi_{r'}}) < U^{\overline{\theta}_k}(y^{\pi_{r'}^+}), \tag{8}$$

where $\pi_{r'}^+ \in \Pi_{r'}$ is the cell immediately on the right of $\pi_{r'}$, and such that from step r' to $r'+1$, $\overline{\theta}_k$ is moved from $\pi_{r'}$ to $\pi_{r'}^+$, and then, from step $r'+1$ to r, cell $\pi_{r'}^+$ is possibly filled with lower types, and also possibly emptied from some of its highest types, up to achieve $\pi_r(\overline{\theta}_k)$ (see Figure 3).

$$
\begin{array}{rrcl}
\Pi_0 : & \ldots\}, \{\ldots & , \overline{\theta}_k\} & , \{\ldots. \\
 & & \vdots & \\
\Pi_{r'} : & \ldots\}, \{\ldots\ldots & , \overline{\theta}_k\} & , \overbrace{\boldsymbol{\{\ldots\ldots\ldots\}}}^{\pi^+_{r'}}, \{\ldots \\
\Pi_{r'+1} : & \ldots\}, \{\ldots\ldots\} & , \boldsymbol{\{}\overline{\theta}_k & \boldsymbol{,\ldots\ldots\ldots\}}, \{\ldots \\
 & & \vdots & \\
\Pi_{r'+k} : & \ldots\ldots\}, \boldsymbol{\{\ldots,} & \overline{\theta}_k & \boldsymbol{,\ldots\ldots\}}, \{\ldots \\
 & & \vdots & \\
\Pi_r : & \ldots\}, \boldsymbol{\{\ldots\ldots,} & \overline{\theta}_k & \boldsymbol{,\ldots\}}, \{\ldots \\
\hline
\Pi : & \ldots\}, \{\ldots\ldots\ldots, & \overline{\theta}_k\} & , \{\ldots
\end{array}
$$

Figure 3: From $\pi^+_{r'}$ to $\pi_r(\overline{\theta}_k)$ along an envy sequence initialized at Π_0.

In particular, the sequence of actions that starts at step r' with the action $y^{\pi^+_{r'}}$ that $\overline{\theta}_k$ prefers, and then, from step r' to step r, that goes on with the action associated with $\overline{\theta}_k$, that ends at $y^{\pi_r(\overline{\theta})}$, is a *strictly decreasing* sequence of actions. Therefore, we have:

$$y^{\pi_r(\overline{\theta}_k)} < y^{\pi^+_{r'}}. \tag{9}$$

Moreover, from Lemma 2, we have, by induction (since $\Pi_0 \geq \Pi$),

$$\Pi_{r'} \geq \Pi,$$

and then $\min \pi_{r'}(\overline{\theta}_k) = \min \pi_{r'} \geq \min \pi^k$. Since $\max \pi_{r'} = \max \pi^k = \overline{\theta}_k$, we obtain

$$y^{\pi^k} \leq y^{\pi_{r'}} \leq y^{\overline{\theta}_k}. \tag{10}$$

From (7) and (9), we have $y^{\pi_r(\overline{\theta}_k)} \in \left[y^{\pi^k}, y^{\pi^+_{r'}}\right)$ (with $y^{\pi_r(\overline{\theta}_k)} > y^{\pi^k}$ if $\pi_r(\overline{\theta}_k) \neq \pi^k$). Since $U^{\overline{\theta}_k}$ is single peaked, we obtain

$$U^{\overline{\theta}_k}(y^{\pi_r(\overline{\theta}_k)}) \geq \min\{U^{\overline{\theta}_k}(y^{\pi^k}), U^{\overline{\theta}_k}(y^{\pi^+_{r'}})\}, \tag{11}$$

with a strict inequality if $\pi_r(\overline{\theta}_k) \neq \pi^k$. From (10), $U^{\overline{\theta}_k}$ is increasing on $\left[y^{\pi^k}, y^{\pi_{r'}}\right]$. Hence

$$U^{\overline{\theta}_k}(y^{\pi_{r'}}) \geq U^{\overline{\theta}_k}(y^{\pi^k}). \tag{12}$$

Then from (8) and (12), we obtain

$$U^{\overline{\theta}_k}(y^{\pi^+_{r'}}) \geq U^{\overline{\theta}_k}(y^{\pi^k}).$$

In other words, $\min\{U^{\overline{\theta}_k}(y^{\pi^k}), U^{\overline{\theta}_k}(y^{\pi^+_{r'}})\} = U^{\overline{\theta}_k}(y^{\pi^k})$ and (11) gives the result. $\square$

As noticed above, when $\Pi_0 = \mathrm{CR}$, the assumptions of Lemma 3 involving Π_0 are satisfied. By applying Proposition 7, the conclusion of the lemma applies as well to the limit partition Π_*, leading to the next statement.

Corollary 2. *Let* Π *be an IC partition, let* π *be a cell of* Π*, and let* $\overline{\theta} = \max \pi$*. Then*

$$U^{\overline{\theta}}(y^{\pi_*(\overline{\theta})}) \geq U^{\overline{\theta}}(y^{\pi}),$$

with equality only if $\pi_*(\overline{\theta}) = \pi$*.*

The previous statement implies that no IC partition Π defeats Π_*, which completes the proof of Proposition 3.

The previous reasoning relies heavily on sequences $(\Pi_r)_{r\geq 0}$ that start at $\Pi_0 = \mathrm{CR}$ and converge to Π_*. The partition Π_* is undefeated because the maximal type of every cell of every IC partition is associated with higher types in Π_* by envy. Indeed going from Π_0 to Π_*, maximal types are moved to their next succeeding cell only if they prefer to be moved. This holds recursively. As a consequence, in Π_*, types are *minimally pooled* with respect to their incentives in CR. To illustrate this point, the following example shows that if condition $\overline{\theta} = \max \pi_0$ does not hold, i.e., if $\overline{\theta}$ is already pooled with higher types in Π_0, then inequality $U^{\overline{\theta}}(y^{\pi_0}) \geq U^{\overline{\theta}}(y^{\pi})$ may not hold, even if Π_0 is L-IC and $\Pi_0 \geq \Pi'$ for every IC partition Π'.

Example 4. Consider $\Theta = \{1, 3, 4, 9, 10\}$, with $p(1) = p(3) = 0.2$, $p(4) = 0.3$, $p(9) = 0.29$, and $p(10) = 0.01$. Payoff functions are $V^{\theta}(x) = -(\theta - x)^2$ and $U^{\theta}(x) = -(\theta + 0.6 - x)^2$. Partition $\Pi = \{\{1, 3\}, \{4, 9, 10\}\}$, where $y^{\{1,3\}} = 2$ and $y^{\{4,9,10\}} \simeq 6.5$, is IC. The sequence $(\Pi_r)_{r\geq 0}$ initialized at $\Pi_0 = \mathrm{CR}$ converges to $\Pi_* = \{\{1\}, \{3, 4\}, \{9, 10\}\}$. Now consider the alternative partition $\Pi'_0 = \{\{1\}, \{3, 4, 9\}, \{10\}\}$, where $y^{\{1\}} = 1$, $y^{\{3,4,9\}} \simeq 5.6$, and $y^{\{10\}} = 10$. The partition Π'_0 is L-IC, and clearly $\Pi'_0 \geq \Pi$. Let us show that it furthermore satisfies $\Pi'_0 \geq \Pi'$ for *every* IC partition Π'. Suppose, by contradiction, that there exists a cell π' of an IC partition Π' such that ($\max \pi' = 9$ and $\min \pi' > 3$) or ($\max \pi' = 4$ and $\min \pi' > 3$). In the former case, in Π' type 9 would envy $\{10\} \in \Pi'$. In the latter case, in Π' type 3 would envy $\pi' = \{4\}$. Hence there is no such Π'. Thus Π'_0 is such that $\Pi'_0 \geq \Pi'$ for every IC partition Π'. According to Proposition 7 and Theorem 1, the sequence $(\mathrm{Env}^r(\Pi'_0))_{r\geq 0}$ converges to Π_*. However, in Π type $3 = \max \pi^1$ obtains $y^{\{1,3\}} = 2$, and does not prefer cell $\pi'_0(3) = \{3, 4, 9\}$ of Π'_0, associated with $y^{\{3,4,9\}} \simeq 5.6$.

Let us turn to the proof of Proposition 4, namely, the partition Π_* cannot be defeated by any pseudo-partition.

As a first step, in the same way as Definition 5, Definition 8 readily applies to a partition Π and an IC pseudo-partition Π': Π *dominates* Π' ($\Pi \geq \Pi'$) if for every pseudo-cell π' of Π', $\min \pi(\max \pi') \geq$

$\min \pi'$. Indeed, in the previous expression, the largest element of π', namely, $\max \pi'$, is well-defined and belongs to a single cell of Π, namely, $\pi(\max \pi')$. In particular, the completely revealing partition CR dominates every IC pseudo-partition.

As a second step, we extend Lemmas 2 and 3.

Lemma 4 (Lemma 2 extended to pseudo-partitions)**.** *For every L-IC partition* Π_0 *and every* $\Pi_1 \in \mathrm{Env}(\Pi_0)$*, if* $\Pi_0 \geq \Pi$ *for some IC pseudo-partition* Π*, then* $\Pi_1 \geq \Pi$.

Proof. Let Π be an IC pseudo-partition, let Π_0 be an L-IC partition such that $\Pi_0 \geq \Pi$, and let $\Pi_1 \in \mathrm{Env}(\Pi_0)$.

Write $\Pi = \{\pi^1, \ldots, \pi^n\}$, with $n = |\Pi|$, and choose type $\overline{\theta}_k = \max \pi^k$ for some $k \in \{1, \ldots, n-1\}$, which possibly randomizes between π^k and π^{k+1}. From $\Pi_0 \geq \Pi$, we have

$$\min \pi_0(\overline{\theta}_k) \geq \min \pi^k$$

and we have to show

$$\min \pi_1(\overline{\theta}_k) \geq \min \pi^k.$$

The proof can be completed exactly in the same way as for Lemma 2. □

By applying the previous lemma inductively, we deduce that

Corollary 3. *For every IC pseudo-partition* Π*,* $\Pi_* \geq \Pi$.

Lemma 5 (Lemma 3 extended to pseudo-partitions)**.** *Let* Π *be an IC pseudo-partition, let* π *be a pseudo-cell of* Π *and let* $\overline{\theta} = \max \pi$*. Let* Π_0 *be an L-IC partition such that* $\Pi_0 \geq \Pi$ *and such that* $\overline{\theta} = \max \pi_0$ *for some cell* $\pi_0 \in \Pi_0$*. Let* $r \in \mathbb{N}$ *and let* $\Pi_r \in \mathrm{Env}^r(\Pi_0)$*. Then*

$$U^{\overline{\theta}}(y^{\pi_r(\overline{\theta})}) \geq U^{\overline{\theta}}(y^{\pi}),$$

with equality only if $\pi_r(\overline{\theta}) = \pi$ *and* $\overline{\theta}$ *does not randomize, or if* $\pi_r(\overline{\theta}) = \pi = \{\overline{\theta}\}$.

Proof. Let $\Pi = \{\pi^1, \ldots, \pi^n\}$ be an IC pseudo-partition, with $n = |\Pi|$, let $k \in \{1, \ldots, n-1\}$ and let $\overline{\theta}_k = \max \pi^k$, which possibly randomizes between cell π^k and π^{k+1}. Let Π_0 be an L-IC partition, e.g., $\Pi_0 = \mathrm{CR}$, such that $\Pi_0 \geq \Pi$ and such that $\overline{\theta}_k = \max \pi_0$ for some cell $\pi_0 \in \Pi_0$. Let $r \in \mathbb{N}$ and let $\Pi_r \in \mathrm{Env}^r(\Pi_0)$.

We have $y^{\pi^k} \leq y^{\overline{\theta}_k}$ since every type in π^k is (weakly) lower than $\overline{\theta}_k$. In particular, $U^{\overline{\theta}_k}$ is increasing on $\left[y^{\pi^k}, y^{\overline{\theta}_k}\right]$. Then, if $\pi_r(\overline{\theta}_k)$ is a singleton, the lemma follows (with equality only if π^k is a singleton too). Hence we can assume that $\pi_r(\overline{\theta}_k)$ is not a singleton for the remainder of the proof.

From Lemma 4, we have, by induction, $\Pi_r \geq \Pi$. From $\Pi_r \geq \Pi$, the types of $\pi_r(\overline{\theta}_k)$ which are lower than or equal to $\overline{\theta}_k$ are the highest types of π^k and the lowest types of $\pi_r(\overline{\theta}_k)$. Hence the action associated with these types is higher than y^{π^k}, and lower than $y^{\pi_r(\overline{\theta}_k)}$. In particular,

$$y^{\pi_r(\overline{\theta}_k)} \geq y^{\pi^k}. \tag{13}$$

Moreover, if $\pi_r(\overline{\theta}_k) \neq \pi^k$, or if $\pi_r(\overline{\theta}_k) = \pi^k$ and $\overline{\theta}_k$ randomizes, the inequality is strict. From now on, the proof proceeds exactly in the same way as the proof of Lemma 3. □

Proceeding as for Proposition 3, we conclude that Proposition 4 holds.

6.2 Neologism-proof partition

In this section, we establish Proposition 2, namely, that the partition Π_* identified in Theorem 1 is the only one that can be neologism-proof (see Definition 4) in our model.

Proof. Suppose that $\Pi = \{\pi^1, \ldots, \pi^n\}$ is a neologism-proof IC partition, with $n = |\Pi|$. We shall show that every cell of Π also is a cell of Π_* by induction, starting from the last cell of Π. To this end, we shall consider a specific path of the algorithm, starting at $\Pi_0 = \text{CR}$ and reaching Π.

Formally, given $(\Pi_r)_{r\in\mathbb{N}}$ such that for every $r \geq 0$, $\Pi_r \in \text{Env}^r(\text{CR})$, we define a condition $P(t)$, $t \in \{0, \ldots, n\}$, as follows:

P(t) : There exists a rank $r = r[t] \in \mathbb{N}$ and a partition $\Pi_r \in \text{Env}^r(\text{CR})$ such that

$$\Pi_r = \{\underbrace{\{\theta_1\}, \ldots, \{\theta_{k_r-1}\}}_{\text{or } = \varnothing \text{ if } t=n}, \underbrace{\pi_r^{n_r-t}, \ldots, \pi_r^{n_r-1}}_{\text{or } = \varnothing \text{ if } t=0}\}, \tag{14}$$

where $n_r - 1 = |\Pi_r|$, $k_r = n_r - t = |\Pi_r| + 1 - t$, and:

(i) if $t > 0$, cells $\pi_r^{n_r-t}$,...,$\pi_r^{n_r-1}$ are the last cells of Π, and

(ii) if $0 < t < |\Pi|$, type $\theta_{k_r} = \min \pi_r^{n_r-t}$ does not prefer $y^{\theta_{k_r-1}}$ to its Π_r-associated action $y^{\pi_r(\theta_{k_r})}$, i.e.:

$$U^{\theta_{k_r}}(y^{\theta_{k_r-1}}) \leq U^{\theta_{k_r}}(y^{\pi_r(\theta_{k_r})}).$$

Note that if condition $P(t)$ holds at $t = n$, then, from (14) and (i), the partition achieved at rank $r[t] = r[n]$ is Π, and therefore $\Pi = \Pi_*$.

We now show by induction that condition $P(t)$ holds for every $t \in \{0, \ldots, n\}$.

Base case. Condition $P(0)$ holds with $r[0] = 0$, because then (14) holds, because $\Pi_{r[0]} = \Pi_0 = \{\{\theta_1\}, \ldots, \{\theta_N\}\}$, and properties (i) and (ii) are irrelevant when $t = 0$.

Induction step. Given $0 \le t \le n-1$ such that condition $P(t)$ holds, with associated rank $r = r[t]$ and integers n_r and k_r, let us denote by

$$\pi = \{\theta_{k_r-\ell_r}, \dots, \theta_{k_r-1}\}$$

the cell of Π containing θ_{k_r-1}, with $\ell_r \ge 1$. Then the induction step is obtained if we show:

(i) there is a continuation path of the envy driven algorithm that starts from Π_r and achieves partition

$$\Pi_{r'} = \{\underbrace{\{\theta_1\}, \dots, \dots, \{\theta_{k_r-\ell_r-1}\}}_{\text{or } = \varnothing \text{ if } t=n-1}, \pi, \underbrace{\pi_r^{n_r-t}, \dots, \pi_r^{n_r-1}}_{\text{or } = \varnothing \text{ if } t=0}\} \tag{15}$$

at some step $r' = r[t+1] \in \mathbb{N}$, $r' \ge r$, and

(ii) if $t < n-1$, type $\theta_{k_{r'}} = \theta_{k_r-\ell_r} = \min \pi$ does not prefers $y^{\theta_{k_r-\ell_r-1}} = y^{\theta_{k_{r'}-1}}$ to y^π (note that $t < n-1$ and thus $k_r - \ell_r > 1$).

If π is a singleton, both properties (i) and (ii) are obtained at $r' = r[t+1] = r[t]$. Indeed, (i) partition $\Pi_{r'}$ as defined in (15) is already achieved at $r = r[t]$, and (ii) if $k_r - 1 > 1$, type $\theta_{k_{r'}} = \theta_{k_r-1} = \min \pi$ does not prefer $y^{\theta_{k_{r'}-1}} = y^{\theta_{k_r-2}}$ to $y^\pi = y^{\theta_{k_{r'}}}$, because $U^{\theta_{k_{r'}}}$ is increasing at $y^{\theta_{k_{r'}}}$ and $y^{\theta_{k_{r'}-1}} < y^{\theta_{k_{r'}}}$. Hence we can assume that π is not a singleton for the remainder of the proof.

In order to obtain (i) and (ii) as stated above when π is not a singleton, we will use the assumption that Π is neologism-proof on particular subsets T of π. For this, we define, for $\ell, \ell' \in \{1, \dots, \ell_r\}$, $\ell \le \ell'$:

$$T_{\ell,\ell'} = \{\theta_{k_r-\ell'}, \dots, \theta_{k_r-\ell}\}.$$

In particular, we have $T_{1,\ell_r} = \pi$, and for every $\ell \in \{2, \dots, \ell_r\} \neq \varnothing$ (recall that π is not a singleton): $T_{1,\ell-1} = \{\theta_{k_r-(\ell-1)}, \dots, \theta_{k_r-1}\}$ consists of the $\ell - 1$ highest types of π, whereas $T_{\ell,\ell_r} = \{\theta_{k_r-\ell_r}, \dots, \theta_{k_r-\ell}\}$ consists of the remaining lower types of π.

We claim that (i) and (ii) hold as long as the following property (P) holds.

(P): For every $\ell \in \{1, \dots, \ell_r\}$, if type $\theta_{k_r-\ell}$ is pooled with the type of π that are lower that itself, i.e. if it is associated with $y^{T_{\ell,\ell_r}}$, then it prefers $y^{T_{\ell',\ell-1}}$ for every $\ell' \in \{1, \dots, \ell-1\}$.

We now prove the claim, and next, we prove that (P) holds.

Proof of the claim. Starting from the singletons $\{\theta_{k_r-\ell_r}\}$, …,$\{\theta_{k_r-1}\}$ reached as cells of Π_r, property (P) guarantees that there is a continuation path of envy driven moves among types $\theta_{k_r-\ell_r}$, …, θ_{k_r-1} that achieves cell π. The corresponding envy driven continuation path is defined as follows: at every step, the highest type of the lowest cell is moved to its next succeeding cell (see Figure 4 for an illustration). This continuation path completes the proof of the claim concerning (i). To get (ii), note that the last move of the continuation path is necessarily the one in which $\theta_{k_r-\ell_r}$, associated with $y^{\theta_{k_r-\ell_r}}$, is moved by envy to cell $T_{1,\ell_r-1} = \{\theta_{k_r-(\ell_r-1)}, \dots, \theta_{k_r-1}\}$. In particular, $U^{\theta_{k_r-\ell_r}}(y^{\theta_{k_r-\ell_r}}) < U^{\theta_{k_r-\ell_r}}(y^{T_{1,\ell_r-1}})$. From this inequality, from $y^{\theta_{k_r-\ell_r}} < y^\pi = y^{T_{1,\ell_r}} < y^{T_{1,\ell_r-1}}$ and from $U^{\theta_{k_r-\ell_r}}$ single-peaked at $x^*(\theta_{k_r-\ell_r}) \geq y^{\theta_{k_r-\ell_r}}$, we can deduce that $U^{\theta_{k_r-\ell_r}}(y^\pi) > U^{\theta_{k_r-\ell_r}}(y^{\theta_{k_r-\ell_r}})$. Since moreover, $y^{\theta_{k_r-(\ell_r+1)}} < y^{\theta_{k_r-\ell_r}}$ and $U^{\theta_{k_r-\ell_r}}$ increasing on $(-\infty, y^{\theta_{k_r-\ell_r}}]$, we also obtain $U^{\theta_{k_r-\ell_r}}(y^\pi) > U^{\theta_{k_r-\ell_r}}(y^{\theta_{k_r-(\ell_r+1)}})$. This completes the proof of the claim concerning (ii).

$$
\begin{array}{rlcl}
r = r[t], \Pi_r: & \{\dots\} & \{\theta_{k_r-4}\}, \{\theta_{k_r-3}\}, \{\theta_{k_r-2}\}, \{\theta_{k_r-1}\} & \{\dots\} \\
& \{\dots\} & \{\theta_{k_r-4}, \theta_{k_r-3}\}, \{\theta_{k_r-2}\}, \{\theta_{k_r-1}\} & \{\dots\} \\
& \{\dots\} & \{\theta_{k_r-4}\}, \{\theta_{k_r-3}, \theta_{k_r-2}\}, \{\theta_{k_r-1}\} & \{\dots\} \\
& \{\dots\} & \{\theta_{k_r-4}, \theta_{k_r-3}, \theta_{k_r-2}\}, \{\theta_{k_r-1}\} & \{\dots\} \\
& \{\dots\} & \{\theta_{k_r-4}, \theta_{k_r-3}\}, \{\theta_{k_r-2}, \theta_{k_r-1}\} & \{\dots\} \\
& \{\dots\} & \{\theta_{k_r-4}\}, \{\theta_{k_r-3}, \theta_{k_r-2}, \theta_{k_r-1}\} & \{\dots\} \\
r' = r[t+1], \Pi_{r'}: & \{\dots\} & \underbrace{\{\theta_{k_r-4}, \theta_{k_r-3}, \theta_{k_r-2}, \theta_{k_r-1}\}}_{=\pi} & \{\dots\}
\end{array}
$$

Figure 4: The continuation path achieving π, if $|\pi| = 4$, in which each moves results from property (P).

Proof of (P). Formally, (P) might be written as:

(P) For every $\ell \in \{1, \dots, \ell_r\}$, for every $\ell' \in \{1, \dots, \ell - 1\}$,

$$U^{\theta_{k_r-\ell}}(y^{T_{\ell,\ell_r}}) < U^{\theta_{k_r-\ell}}(y^{T_{\ell',\ell-1}}). \tag{16}$$

First, note that for every $\ell \in \{2, \dots, \ell_r\}$,

$$y^{T_{\ell,\ell_r}} < y^\pi < y^{T_{1,\ell-1}}, \tag{17}$$

because, as noted above, T_{ℓ,ℓ_r} consists of lower types of π, and $T_{1,\ell-1}$ consists of higher types of π.

Since $U^{\theta_{k_r-1}}$ is increasing at $y^{\theta_{k_r-1}} = y^{T_{1,1}}$, inequality $y^\pi < y^{T_{1,1}}$ implies that type θ_{k_r-1} prefers $y^{T_{1,1}}$ to y^π. If θ_{k_r-2} did not prefer $y^{T_{1,1}}$ to y^π, then either when $t = 0$, or when $t \geq 1$, the set $T_{1,1}$ would be self-signaling in Π. Indeed, in the latter case, by the induction hypothesis (ii), θ_{k_r} does not prefer $y^{\theta_{k_r-1}}$ to $y^{\pi(\theta_{k_r})}$. Thus θ_{k_r-2} prefers $y^{T_{1,1}}$ to y^π, i.e., the following inequality holds at $\ell = 2$:

$$U^{\theta_{k_r-\ell}}(y^\pi) < U^{\theta_{k_r-\ell}}(y^{T_{1,\ell-1}}). \tag{18}$$

Let us now show, by a similar argument, that (18) holds inductively along $\ell = 2, \dots, \ell = \ell_r - 1$ with regard to the set $T_{1,\ell-1}$ and the type $\theta_{k_r-\ell}$. More precisely, for every $\ell \in \{2, \dots, \ell_r\}$, type θ_{k_r} (if $t \geq 1$)

does not prefer $y^{T_{1,\ell-1}}$ to $y^{\pi(\theta_{k_r})}$ because $U^{\theta_{k_r}}$ is increasing on $(-\infty, y^{\theta_{k_r}}]$, and $y^{T_{1,\ell-1}} \leq y^{\theta_{k_r-1}} < y^{\theta_{k_r}}$ and, by the induction hypothesis (ii), θ_{k_r} does not prefer $y^{\theta_{k_r-1}}$ to $y^{\pi(\theta_{k_r})}$. Thus, if there is some $\ell \in \{2, \ldots, \ell_r\}$ such that (18) does not hold, then there is a greatest $\ell \in \{2, \ldots, \ell_r\}$, namely $\overline{\ell}$, such that (18) holds for every $\ell \in \{2, \ldots, \overline{\ell}\}$ and $T_{1,\overline{\ell}-1}$ would be self-signaling. Since Π is neologism-proof, for every $\ell \in \{2, \ldots, \ell_r\}$, $T_{1,\ell-1}$ is not self-signaling, and thus (18) holds.

Now given $\ell \in \{2, \ldots, \ell_r\}$, the second inequality in (17) and inequality (18) imply that $U^{\theta_{k_r-\ell}}$ is increasing at y^{π}. Then it is increasing on $(-\infty, y^{\pi}]$ and the first inequality in (17) implies

$$U^{\theta_{k_r-\ell}}(y^{T_{\ell,\ell_r}}) < U^{\theta_{k_r-\ell}}(y^{\pi}). \tag{19}$$

Together with (18), we obtain

$$U^{\theta_{k_r-\ell}}(y^{T_{\ell,\ell_r}}) < U^{\theta_{k_r-\ell}}(y^{T_{1,\ell-1}}). \tag{20}$$

Now for every $\ell' \in \{1, \ldots, \ell-1\}$, $T_{\ell',\ell-1}$ consists of lower types of $T_{1,\ell-1}$, and every type in $T_{\ell',\ell-1}$ is higher than every type in T_{ℓ,ℓ_r}, and therefore

$$y^{T_{\ell,\ell_r}} < y^{T_{\ell',\ell-1}} \leq y^{T_{1,\ell-1}}.$$

Since moreover, $\theta_{k_r-\ell}$ is the greatest type of T_{ℓ,ℓ_r} and is lower than every type in $T_{\ell',\ell-1}$, we have $y^{T_{\ell,\ell_r}} \leq y^{\theta_{k_r-\ell}} < y^{T_{\ell',\ell-1}}$. Therefore

$$y^{T_{\ell,\ell_r}} \leq y^{\theta_{k_r-\ell}} < y^{T_{\ell',\ell-1}} \leq y^{T_{1,\ell-1}}.$$

Then inequality (16) follows: if $y^{T_{\ell',\ell-1}} \leq x^*(\theta_{k_r-\ell})$, then $U^{\theta_{k_r-\ell}}$ is increasing on $[y^{T_{\ell,\ell_r}}, y^{T_{\ell',\ell-1}}]$, which gives (16), and otherwise $U^{\theta_{k_r-\ell}}$ is decreasing on $[y^{T_{\ell',\ell-1}}, y^{T_{1,\ell-1}}]$, and thus

$$U^{\theta_{k_r-\ell}}(y^{T_{\ell',\ell-1}}) \geq U^{\theta_{k_r-\ell}}(y^{T_{1,\ell-1}}),$$

and (20) gives the result. $\square$

The result established above is actually stronger than Proposition 2. Indeed, the fact that Π is neologism-proof is only used to establish inequality (18), i.e.,

$$U^{\theta_{k-\ell}}(y^{\pi}) < U^{\theta_{k-\ell}}(y^{\{\theta_{k-(\ell-1)}, \ldots, \theta_{k-1}\}}) \tag{*}$$

for every cell $\pi = \{\theta_{k-|\pi|}, \ldots, \theta_{k-1}\} \in \Pi$ such that $|\pi| \geq 2$ and for every $\ell \in \{2, \ldots, |\pi|\}$. In particular, condition (*) already implies that $\Pi = \Pi_*$.

Let us rewrite condition (*) for an arbitrary IC partition Π, by relabelling types with respect to the original order over Θ :

For every cell $\pi = \{\theta_{k+1}, \ldots, \theta_{k+|\pi|}\}$ of Π, $k \in \{0, \ldots, N-2\}$, such that $|\pi| \geq 2$, for every $\ell \in \{1, \ldots, |\pi|-1\}$,

$$U^{\theta_{k+\ell}}(y^{\pi}) < U^{\theta_{k+\ell}}(y^{\{\theta_{k+\ell+1}, \ldots, \theta_{k+|\pi|}\}}). \tag{**}$$

Thinking of Π as a putative equilibrium to be tested and recalling Definition 4, condition (**) says that, in every cell π of Π containing at least two types, no neologism consisting of the highest types of π, namely, of the form $\{\theta_{k+\ell+1}, \ldots, \theta_{k+|\pi|}\}$, can be self-signaling, because the preceding type, namely, $\theta_{k+\ell}$, would benefit from the neologism as well.[19]

Summing up, condition (**) could be referred to as "No Self-Signaling of the Highest Types" (NSSHT). The proof of Proposition 2 shows that every neologism-proof IC partition satisfies NSSHT and that the partition Π_* is the only one that can possibly satisfy NSSHT.

6.3 NITS

In this section, we establish Proposition 5.

Proof. Let us consider a sequence $(\Pi_r)_{r\geq 0}$ that starts at $\Pi_0 = \mathrm{CR}$, satisfies $\Pi_{r+1} \in \mathrm{Env}(\Pi_r)$, and thus converges to Π_*. Let π_*^1 be the cell containing $\theta_1 = \min \Theta$. We must show that

$$U^{\theta_1}(y^{\theta_1}) \leq U^{\theta_1}(y^{\pi_*^1}). \tag{21}$$

If $\pi_*^1 = \{\theta_1\}$, then (21) holds. Otherwise, there is some step r such that $\pi_r^1 = \{\theta_1\}$ and type θ_1 prefers action $y^{\pi_r^2}$ to $y^{\pi_r^1} = y^{\theta_1}$, i.e.,

$$U^{\theta_1}(y^{\theta_1}) < U^{\theta_1}(y^{\pi_r^2}). \tag{22}$$

Then $\pi_{r+1}^1 = \pi_r^1 \cup \{\theta_1\}$ and $y^{\pi_{r+1}^1}$ lies in the interval $(y^{\theta_1}, y^{\pi_r^2})$. Let us show that

$$U^{\theta_1}(y^{\theta_1}) < U^{\theta_1}(y^{\pi_{r+1}^1}). \tag{23}$$

If $y^{\pi_{r+1}^1} \leq x^*(\theta_1)$, (23) follows from the fact that U^{θ_1} increasing on $[y^{\theta_1}, y^{\pi_{r+1}^1}]$. If $y^{\pi_{r+1}^1} > x^*(\theta_1)$, U^{θ_1} is decreasing on $[y^{\pi_{r+1}^1}, y^{\pi_r^2}]$, so that $U^{\theta_1}(y^{\pi_r^2}) < U^{\theta_1}(y^{\pi_{r+1}^1})$, which also implies (23) from (22). This completes the proof of (23). Now, from π_{r+1}^1 to π_*^1, only the highest type of the first cell of every reached partition is possibly moved to its next succeeding cell. Accordingly,

$$y^{\theta_1} < y^{\pi_*^1} \leq y^{\pi_{r+1}^1}.$$

Let us establish (21). If $y^{\pi_*^1} \leq y^*(\theta_1)$, (21) follows from the fact that U^{θ_1} is increasing on $[y^{\theta_1}, y^{\pi_*^1}]$. If $y^{\pi_*^1} > y^*(\theta_1)$, U^{θ_1} is decreasing on $[y^{\pi_*^1}, y^{\pi_{r+1}^1}]$, so that $U^{\theta_1}(y^{\pi_*^1}) \geq U^{\theta_1}(y^{\pi_{r+1}^1})$, which also implies (21) from (23). $\square$

[19] According to condition (**), the neologism $\{\theta_{k+\ell+1}, \ldots, \theta_{k+|\pi|}\}$ is not self-signaling because it does not satisfy the *second* requirement in Definition 4, which, as argued before, relies on the receiver's inertia. Note that, under our assumptions, condition (**) also implies that the neologism satisfies the *first* requirement to be self-signaling, namely, all types in $\{\theta_{k+\ell+1}, \ldots, \theta_{k+|\pi|}\}$ prefer $y^{\{\theta_{k+\ell+1}, \ldots, \theta_{k+|\pi|}\}}$ to y^{π}. This clearly holds for the base case $\ell = |\pi| - 1$, and condition (**) gives the induction step $((\ell) \implies (\ell - 1))$.

Note that NSSHT implies NITS. Indeed, if the lowest cell $\pi = \{\theta_1, \ldots, \theta_{|\pi|}\}$, with $|\pi| \geq 2$, of an IC partition Π is such that the lowest type θ_1 satisfies inequality (**), i.e., $U^{\theta_1}(y^\pi) < U^{\theta_1}(y^{\pi\setminus\{\theta_1\}})$, then, because $y^\pi < y^{\pi\setminus\{\theta_1\}}$, it must be that U^{θ_1} is increasing at y^π, and therefore $U^{\theta_1}(y^{\theta_1}) < U^{\theta_1}(y^\pi)$, i.e., the partition Π satisfies NITS.

However, compared with NITS, NSSHT (viewed as a refinement criterion) looks too strong, since it may remove every equilibrium. As the following example illustrates (see also Example 2 in Section 4), the partition Π_* may not satisfy NSSHT (but is the only one which could satisfy it).

Example 5. Let us consider the uniform quadratic case with two types, i.e., $\Theta = \{1, 2\}$. The receiver's optimal actions are $y^1 = 1$, $y^{\{1,2\}} = 1.5$ and $y^2 = 2$. If $0 < b \leq 0.5$, $\Pi_* = \text{CR} = \{\{1\}, \{2\}\}$ is neologism-proof and thus satisfies NSSHT. If $b > 0.5$, $\Pi_* = \text{NR} = \{\{1, 2\}\}$ is the only IC partition. If $0.5 < b \leq 0.75$, the neologism $\{2\}$ is self-signaling, because type 1 weakly prefers $y^{\{1,2\}}$ to y^2. In this case, no IC partition satisfies NSSHT. However, if $b > 0.75$, type 1 strictly prefers y^2 to $y^{\{1,2\}}$, so that the neologism $\{2\}$ is not self-signaling anymore: the partition Π_* is neologism-proof and satisfies NSSHT.

6.4 Ex ante optimal PBE

In this section, we establish Proposition 6 by showing that, in the uniform quadratic case, the partition Π_* reached by our algorithm coincides with the ex ante Pareto optimal partition proposed by Frug (2016), henceforth denoted as Frug's partition.

Proof. For simplicity, we focus on the case where $4b$ is an integer and set $k_0 = 4b - 2$. An interval partition, $\Pi = \{\pi^1, \ldots, \pi^n\}$, with $n = |\Pi|$, is then IC if and only, for every j,

$$| \pi^j | + k_0 \leq | \pi^{j+1} | \leq | \pi^j | + k_0 + 4. \tag{24}$$

If $k_0 \leq 0$, i.e., if $b \leq \frac{1}{2}$, Frug's partition and Π_* are both equal to CR. Let $k_0 \geq 1$ and set $x_{j+1} = 1 + jk_0$, for $j = 0, 1, \ldots$ Recalling that $\Theta = \{1, 2, ..., N\}$, let $k = k(N)$ be such that

$$\sum_{j=1}^{k} x_j \leq N < \sum_{j=1}^{k+1} x_j.$$

Then, let $q \in \{0, \cdots, k_0\}$ and $r \in \{0, \cdots, k-1\}$ be such that

$$N - \sum_{j=1}^{k} x_j = qk + r. \tag{25}$$

Frug's partition $\Pi(N)$ consists of the interval partition of $k = k(N)$ cells π^j, $j = 1, \ldots, k$, containing $| \pi^j |$ consecutive types:

$$| \pi^j | = \begin{array}{ll} x_j + q & j = 1, \ldots, k - r \\ x_j + q + 1 & j = k - r + 1, \ldots, k. \end{array}$$

We show that $\Pi(N) = \Pi_*$ by induction on N. Assume that $\Theta = \{1, 2, ..., N+1\}$ and let $\Pi(N+1)$ be Frug's associated partition. By assumption, when the set of types is $\{2, ..., N+1\}$, our algorithm converges to Frug's partition, which, with a slight abuse of notation, is denoted as $\Pi(N)$. Let us perform a run of the algorithm over $\{1, ..., N+1\}$; after a few steps, the partition $\{\{1\}, \Pi(N)\}$ is reached (again with a slight abuse of notation).

Case 1: The number of cells of $\Pi(N)$ is k and the number of cells of $\Pi(N+1)$ is $k+1$, namely,

$$\sum_{j=1}^{k} x_j \leq N < N+1 = \sum_{j=1}^{k+1} x_j.$$

This implies

$$N - \sum_{j=1}^{k} x_j = k_0 k.$$

Hence,

$$\begin{aligned} &| \quad \pi^j(N) \mid= x_j + k_0 = 1 + jk_0 \quad j = 1, \dots, k \\ &| \quad \pi^j(N+1) \mid= x_j = 1 + (j-1)k_0 \quad j = 1, \dots, k+1. \end{aligned}$$

In other words, Frug's partition satisfies $\Pi(N+1) = \{\{1\}, \Pi(N)\}$. But the same happens with our algorithm, since type 1 does not envy the first cell of $\Pi(N)$, which contains $1 + k_0$ types (see (24)).

Case 2: $\Pi(N)$ and $\Pi(N+1)$ have the same number of cells k, namely,

$$\sum_{j=1}^{k} x_j \leq N < N+1 < \sum_{j=1}^{k+1} x_j.$$

Then, in (25), $q \in \{0, \cdots, k_0 - 1\}$. The $(k-r)th$ cell of $\Pi(N)$ contains $x_{k-r} + q$ types, while the $(k-r)th$ cell of $\Pi(N+1)$ contains $x_{k-r} + q + 1$ types. All the other cells of $\Pi(N)$ contain the same number of types as the corresponding cell in $\Pi(N+1)$.

Let us run the algorithm from $\{\{1\}, \Pi(N)\}$. At the first step, type 1 envies the first cell of $\Pi(N)$, which contains $x_1 + q < 1 + k_0$ types (see (24)) and thus joins it. At the second step, there are k cells, the first one contains $x_1 + q + 1$ types. The last type in this cell envies the next cell, which contains $x_2 + q < x_1 + q + 1 + k_0$ types (see again (24)) and thus joins it. There is no envy in the first cell, but the last type of the second cell envies the next one. We can go on moving the last type of a cell to the next one until this last type reaches the $(k-r)th$ cell of $\Pi(N)$ and joins it. Then there is no envy anymore, and the final partition, namely, Π_*, is $\Pi(N+1)$. $\square$

References

Aumann, Robert J, Michael Maschler, and Richard E Stearns, "Repeated games of incomplete information: An approach to the non-zero-sum case," *Report of the US Arms Control and Disarmament Agency, ST-143*, 1968, pp. 117–216.

_ , _ , and _ , *Repeated games with incomplete information*, MIT press, 1995.

Babichenko, Yakov, Inbal Talgam-Cohen, Haifeng Xu, and Konstantin Zabarnyi, "Algorithmic Cheap Talk," *arXiv preprint arXiv:2311.09011*, 2023.

Blume, Andreas, "Meaning in communication games," Technical Report, Working Paper, University of Arizona 2023.

Chen, Ying, Navin Kartik, and Joel Sobel, "Selecting cheap-talk equilibria," *Econometrica*, 2008, *76* (1), 117–136.

Clark, Daniel, "Robust neologism proofness," Technical Report, Working Paper, Massachusetts Institute of Technology 2021.

Condorelli, Daniele and Massimiliano Furlan, "Cheap Talking Algorithms," *arXiv preprint arXiv:2310.07867*, 2023.

Crawford, Vincent P. and Joel Sobel, "Strategic information transmission," *Econometrica*, 1982, pp. 1431–1451.

Farrell, Joseph, "Meaning and credibility in cheap-talk games," *Games and Economic Behavior*, 1993, *5* (4), 514–531.

Forges, Françoise, "Games with incomplete information: from repetition to cheap talk and persuasion," *Annals of Economics and Statistics*, 2020, (137), 3–30.

Frug, Alexander, "A note on optimal cheap talk equilibria in a discrete state space," *Games and Economic Behavior*, 2016, *99*, 180–185.

Gordon, Sidartha, "Iteratively stable cheap talk," *Department of Economics, Ecole des Sciences Politiques, Paris*, 2011.

_ , Navin Kartik, Melody Pei-yu Lo, Wojciech Olszewski, and Joel Sobel, "Effective communication in cheap talk games," Technical Report, Working Paper, University of California-San Diego 2022.

Hart, Sergiu, "Nonzero-sum two-person repeated games with incomplete information," *Mathematics of Operations Research*, 1985, *10* (1), 117–153.

Hillas, John, "Sequential equilibria and stable sets of beliefs," *Journal of Economic Theory*, 1994, *64* (1), 78–102.

Kartik, Navin and Joel Sobel, "Effective communication in cheap talk games," Technical Report, Working Paper, University of California-San Diego 2015.

Kreps, David M and Joel Sobel, "Signalling," *Handbook of game theory with economic applications*, 1994, *2*, 849–867.

Lo, Melody Pei-yu and Wojciech Olszewski, "Learning in cheap talk games," Technical Report, Working Paper, Northwestern University 2022.

Mailath, George J, Masahiro Okuno-Fujiwara, and Andrew Postlewaite, "Belief-based refinements in signalling games," *Journal of Economic Theory*, 1993, *60* (2), 241–276.

Matthews, Steven A, Masahiro Okuno-Fujiwara, and Andrew Postlewaite, "Refining cheap-talk equilibria," *Journal of Economic Theory*, 1991, *55* (2), 247–273.

Olszewski, Wojciech, "Rich language and refinements of cheap-talk equilibria," *Journal of Economic Theory*, 2006, *128* (1), 164–186.

_ , "A result on convergence of sequences of iterations with applications to best-response dynamics," *Mathematics of Operations Research*, 2022, *47* (3), 2333–2343.

Sémirat, Stéphan and Françoise Forges, "Strategic information transmission with sender's approval: The single-crossing case," *Games and Economic Behavior*, 2022, *134*, 242–263.