Forward-neologism-proof equilibrium and better response dynamics

Stéphan Sémirat, Françoise Forges

Forward-neologism-proof equilibrium and better response dynamics

Stéphan Sémirat* and Françoise Forges†

Preliminary - August 2023 ‡

Abstract

We study pure perfect Bayesian equilibria in sender-receiver games with finitely many types for the sender. Such equilibria are characterized by incentive compatible (IC) partitions of the sender's types. In the case of ordered types, real-valued decisions and well-behaved utility functions (namely, strictly concave, single-peaked, single-crossing and with an upward bias for the sender), we propose a family of iterative optimization processes that all converge to a unique IC partition Π_*. We show that Π_* is undefeated in the sense of Mailath et al. (1993). Equivalently, Π_* is forward-neologism-proof, a variant of Farrell's (1993) neologism-proof concept that we introduce. While the latter refinement (as many other ones) starts from a putative equilibrium and identifies types that would deviate if they were properly identified by the receiver, our iterative optimization processes take the opposite direction. Starting typically from a completely revealing strategy of the sender,

*Université Grenoble Alpes, CNRS, INRA, Grenoble INP, GAEL. E-mail: stephan.semirat@univ-grenoble-alpes.fr

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

‡We wish to acknowledge useful comments from Andreas Blume, Archishman Chakraborty, Daniel Clark, Inga Deimen, Sidartha Gordon, Navin Kartik, Frédéric Koessler, Andrés Salamanca, Joel Sobel, Ina Taneva and Peter Vida. We also thank the participants of various seminars (Cardiff Business School, Institut Henri Poincaré (Game Theory Seminar), One World Mathematical Game Theory Seminar, Arizona State University (ASU) and University of Arizona) and conferences and workshops (Columbia Conference in Economic Theory (New York); Colibri Focus Workshop on Strategic Communication (Graz); Behavior and Strategies in Information Design and Communication (HEC, Paris); Paris Workshop on Games, Decisions and Language; Information Transmission and Incentives (Brescia)).

types are gradually pooled as long as some of them envy decisions associated with other types. The process can be interpreted as a better response dynamics.

1 Introduction

We consider a simple model of cheap talk, in which an informed individual sends a costless message to a decision-maker. We assume that the sender has finitely many types, which can be ranked according to some order. This assumption makes sense in a number of practical situations, for instance, when the types correspond to a qualitative assessment (e.g., "excellent," "good," "fair," etc.).

Except for this, we make the same assumptions as in Crawford and Sobel (1982), namely, the receiver's actions are real-valued and the utility functions are well-behaved (strictly concave, single-peaked, single-crossing and with an upward bias for the sender). Frug (2016) considers perfect Bayesian equilibria (PBE) in this framework but does not address the question of selecting among them, which is central in the current paper.

There is typically a plethora of PBE in sender-receiver games. Many refinements of PBE, like the intuitive criterion of Cho and Kreps (1987), are useful in signaling games but have no selection power when signals have no cost. By contrast, Farrell's (1993) neologism-proof equilibrium is tailored to cheap talk games but is so demanding that it often does not exist. We show that the sender-receiver games that we consider always have a forward-neologism-proof PBE, a variant of neologism-proof PBE in which neologisms are required to be incentive compatible.

We mostly consider PBE in pure strategies. These can be characterized as incentive compatible (IC) partitions of the sender's types. More precisely, every cell of such a partition corresponds to a subset of types, which all send the same message at equilibrium. Equilibria thus induce a canonical language, in which message L means "my type is in L." The receiver reacts by choosing the unique decision that maximizes his updated expected utility, given that the sender's type lies in L. The no-deviation condition of the sender takes the form of an IC condition.

A neologism (with respect to a given equilibrium) is a message, namely a subset of types L, that is not sent at this equilibrium. All variants of neologism-proof equilibrium agree on the fact that, to be credible, a neologism L must be such that all types in L benefit from the

deviation if the receiver interprets it as sent by types in L (there are nevertheless different ways to specify the receiver's beliefs over L). Many versions of the solution concept assume that the receiver sticks to the status quo when he does not receive message L. However, if the receiver takes the neologism seriously, he should also revise his interpretation of the original equilibrium messages, an inference that can be understood by the sender. This logic leads to neologisms that are consistent with an equilibrium strategy of the sender, namely, that are cells of some IC partition. With this restriction, the receiver's beliefs over L are determined by Bayes rule. An IC partition is forward-neologism-proof if there does exist any such consistent neologism. In general, this notion is not weaker nor stronger than Farrell's (1993) one.

The previous equilibrium refinement was actually proposed by Mailath et al. (1993) under the name "undefeated equilibrium" without referring explicitly to neologisms.[1] This solution concept was defined for the class of signaling games. It takes a much simpler form in sender-receiver games, by relying on the partition characterization described above: an IC partition is defeated by another IC partition if there is a cell of the latter (interpreted as a neologism) that is preferred to the former by all types in this cell.

Thanks to this characterization, undefeated IC partitions can in principle be identified by checking, for every IC partition, whether it is defeated by any other. However this requires to make the list of all IC partitions, which is far from being tractable. Furthermore, in absence of suitable assumptions, the "defeat" relation over IC partitions can have cycles, so that there may not be any undefeated IC partition.

These difficulties can be overcome in the sender-receiver games that we study. IC partitions are then formed of "intervals," namely, subsets of consecutive types. We propose a natural iterative optimization process, which converges to an IC partition Π_*. This limit Π_* turns out to be 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.

The limit partition Π_* has further remarkable properties. First, Π_* "dominates" any other IC partition, according to a binary relation over interval partitions, which, loosely speaking, says that a partition dominates another if its cells are "more to the right." In particular, a

[1] We realized the relationship between our approach and Mailath et al.'s (1993) paper – originally entitled "Forward induction and equilibrium refinement " – after having identified the notion of forward-neologism-proof IC partition. Farrell's (1993) footnote 13 also suggests a similar notion.

partition that has more cells than Π_* cannot be IC. Furthermore, Π_* is the only IC partition that can possibly be neologism-proof in our framework.[2] Finally, Π_* satisfies the "no incentive to separate" (NITS) criterion of Chen et al. (2008).

We actually consider a *family* of iterative optimization processes, which typically start with the finest partition, namely, a fully revealing strategy of the sender, and evolve by alternating best replies of the receiver and *better* replies of the sender, in which a single type improves his utility. This way of proceeding introduces some flexibility in choosing the sender's type that improves its utility. We show that, quite unexpectedly, all versions of the algorithm converge to the same limit partition Π_*. [3]

The paper is organized as follows: below we mention some related papers; Section 2 describes the model, namely, the sender-receiver game, and the solution concepts. We first characterize pure perfect Bayesian equilibria (PBE) as incentive compatible (IC) partitions of the set of types. Then we define forward-neologism-proof (synonymously, undefeated) IC partitions and recall Farrell's (1993) notions of self-signaling set and neologism-proof equilibrium. We also introduce a dominance relation over partitions. Section 3 states our main result (Theorem 1), namely, the existence of a unique IC partition Π_* which dominates every IC partition and is undefeated. This section also sketches a proof of Theorem 1, by describing a typical better response dynamics converging to the unique partition Π_* identified in the statement. To this aim, we define the notions of envy and left-incentive compatibility. Section 4 goes on with some basic examples. Section 5 establishes stronger results than those summarized in Theorem 1 by describing a family of iterative optimization processes that all converge to the partition Π_*. Proposition 1 describes in detail the properties of Π_* as a dominating IC partition. Corollary 3 makes precise the reason why Π_* is undefeated: in every other IC partition Π, the highest type of every cell would rather be in Π_* than in Π. Section 6 explores further properties of the partition Π_*. This section is divided into three subsections. First, Section

[2] As a consequence, every neologism-proof IC partition is undefeated, but the scope of this statement is limited by the fact that more often than not, there is no neologism-proof IC partition at all.

[3] Sémirat and Forges (2022) make use of similar, but more intricate, algorithms starting at a fully revealing strategy of the sender, to establish the existence of "equilibria without exit " in sender-receiver games with sender's approval. In the latter framework, nonrevealing equilibria are not relevant, because they typically involve exit. Sémirat and Forges (2022) show that the algorithms always converge but do not investigate the possible uniqueness of the limit partition.

6.1 establishes that Π_* is neologism-proof (Farrell (1993)) and satisfies NITS (No incentive to Separate, Chen et al. (2008)). The proof of the former result shows that Π_* is actually the only partition that can possibly satisfy a weaker property than neologism-proofness, which requires to resist to particular neologisms only and could be denoted as NSSHT, to indicate that the highest types of any cell could not self-signal themselves via a neologism. It turns out that, in our framework, NSSHT implies NITS. The second subsection of Section 6 shows that, in the uniform quadratic case, the partition Π_* coincides with a specific partition constructed by Frug (2016) to demonstrate that, in this particular case, the ex ante Pareto dominant PBE is partitional. Hence, in the uniform quadratic case, the partition Π_* corresponds to an ex ante Pareto optimal PBE. However, a counter-example shows that this result does not survive for more general priors. Finally, the third subsection of Section 6 deals with the extension of the properties of Π_* when mixed strategies are allowed. After having characterized mixed equilibria as IC "pseudo-partitions" of the set of types and extended our basic notions to these, we show that Π_* remains dominant and undefeated within the set of all IC pseudo-partitions, namely, when mixed strategies are allowed. A key lemma for the latter properties is that every IC pseudo-partition is dominated by an associated IC partition, a result suggested in Frug (2016) (see below).

Relationship with other papers

There is a large literature on equilibrium refinements in sender-receiver games. We will be deliberately selective. As already mentioned, Farrell (1993) and Mailath et al. (1993) provide motivations for (possibly forward-) neologism-proof equilibria; these papers contain relevant early references. The latter one identifies 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.

Three features of our approach should be kept in mind when trying to make a precise comparison with other contributions. First of all, we focus on a specific model: the discrete version of Crawford and Sobel (1982). Second, we perform our analysis in terms of equilibrium outcomes, namely, IC partitions, restricting the language to messages of the form "my type is in L." Finally, to select an undefeated equilibrium, we compare equilibria with each other, rather then testing the rationality of every equilibrium separately.

Frug (2016) proposes a first analysis of the discrete version of Crawford and Sobel (1982).

He investigates to which extent one can restrict on pure equilibria in this model. He provides an example in which a mixed equilibrium gives a higner payoff to the receiver than any partitional equilibrium. He also shows that this phenomenon cannot arise when the sender is upward biased, an assumption that is maintained throughout the current paper. More precisely, in this case, Frug (2016) proves (as Proposition 1) that every non-partitional equilibrium can be associated with a partitional one, in which the receiver uses the same number of actions and obtains a higher ex ante expected payoff.[4]

The latter property holds in our framework but does not say anything on the robustness of our own results when the sender is allowed to use a mixed strategy.[5] In Section 6.3, we extend the various tools used in this paper (starting with the equilibrium characterization) to account for mixed equilibria; we show that, when such equilibria are allowed, the unique limit of our algorithm(s) remains undefeated and still "dominates" every other equilibrium. As pointed out above, a key lemma for this result can be found in the proof of Frug's Proposition 1.

In the quadratic uniform case, Frug (2016) constructs an explicit partitional equilibrium that ex ante Pareto dominates every other equilibrium. We show that, in this particular case, the IC partition reached by our algorithm coincides with Frug's (2016) partitional equilibrium.

While our analysis is entirely formulated in terms of a canonical language, many papers refine cheap talk equilibria by relying on a "language with literal meanings." As an early example of such an approach, Matthews et al. (1991) develop various notions of "announcement-proof (mixed) equilibrium" for sender-receiver games with finitely many types and actions but arbitrary utility functions. Their "weak" notion is motivated by the same considerations as undefeated equilibria (typically, consistency of off-path messages with sender's equilibrium strategies). Blume (2023) models the idea of a pre-existing language and analyzes its role in equilibrium prediction (see also the references in this paper).[6]

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.[7] Unlike the forward neologisms

[4]Some basic ideas behind our algorithms are already in the proof of this result, which uses an iterative optimization process to transform any mixed equilibrium into a pure one. Frug (2016) mentions other related procedures, e.g., Gordon (2011) and Chen and Gordon (2015).

[5]Given our assumptions, the receiver never uses a mixed strategy at equilibrium.

[6]See also Olszewski (2006), etc.

[7]As pointed out above, in general, there is no relationship between neologism-proof equilibria and undefeated

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 a forward-neologism-proof 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 is a forward one. It may also happen that a robust neologism-proof equilibrium is not forward-neologism-proof.

Building on Gordon (2011), Kartik and Sobel (2015) and Lo and Olszewski (2022), Gordon et al. (2022) study equilibrium selection in Crawford and Sobel's (1982) original model, in which the sender's type belongs to a real interval. This is the main difference between their model and the current one. In particular, they assume, as we do, that the set of messages is finite. This assumption has no impact on the set of PBE of their game but prevents the sender from using a completely revealing strategy. Since this is the typical starting point of our algorithms, our methodology does not literally apply to their model. However, our main finding is qualitatively similar to one of their results, which identifies a "largest" equilibrium and an adjustment process converging to this equilibrium.

More precisely, Gordon et al. (2022) make the further assumption that messages are ordered and focus on monotonic strategies for both players. This allows for a tractable type-action mapping representation, which remains closer to the original strategies than the partition representation in our discrete case. Applying an interim best response dynamic, they identify a "smallest" and a "largest" equilibrium, which correspond to the limit of a "lower" and an "upper" sequence respectively. Under a suitable regularity condition introduced in Crawford

ones. However, in the class of sender-receiver games that we consider, there is a unique partition that can possibly be neologism-proof, the limit partition Π_*, which is undefeated.

[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.

and Sobel (1982), the two extreme equilibria coincide with the one that induces the larger number of actions in Crawford and Sobel (1982). As a corollary, if there is a unique equilibrium type-action mapping satisfying the "no incentive to separate" (NITS) criterion of Chen et al. (2008), a property that holds under the regularity condition mentioned above, every interim best response sequence converges to an equilibrium with this type-action mapping.

In spite of the similarity between the previous result of Gordon et al. (2022) and ours, there are a number of obvious differences. When the set of types is finite, there is no need to restrict to monotonic strategies to characterize equilibria as IC partitions.[9] Our dynamic process typically starts from specific initial conditions, namely, the sender's completely revealing strategy. It converges to the "largest" IC partition, "largest" being understood with respect to a simple binary relation over interval partitions, independently of any dynamic process.[10]

While our iterative optimization process rests on best responses of the receiver as Gordon et al.'s (2022) one, we just consider particular *better* responses of the sender, in which a single (appropriately chosen) type improves his utility. As we illustrate on Example 1 in Section 4, proceeding in this way is helpful to preserve the receiver's interpretation of messages as truthful subset of types. Our dynamic process converges to the the "largest" IC partition (according to an appropriate dominance relation), whatever the choice of the improving type at every step. It does satisfy NITS, but in our framework, other IC partitions may satisfy this criterion as well (see Section 6.1).

Gordon et al. (2022) also 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. One might want to check whether a similar property holds in our framework. The difficulty is that by reducing the analysis to (possibly IC) partitions, we implicitly focus on specific strategies of the sender and corresponding best replies of the receiver. This restricted framework is appropriate to compare equilibria with each other, which allows us to determine undefeated equilibria. Identifying dominated strategies requires to keep track of at least a simplified version of the strategic form game.[11] While such an analysis could likely be performed, there is no

[9] IC partitions characterize *pure* equilibria but the properties of our limit equilibrium remain true within mixed ones (see Section 6.3).

[10] According to this binary relation, the lowest partition is the nonrevealing one and is of course IC.

[11] Being based on a canonical representation, our methodology is akin to the one that was developed for communication equilibria or more generally in mechanism design. Canonical mechanisms (with an implicit restriction to truthful equilibria) are easily compared with each other. However restriction to such mechanisms

apparent reason to expect that it would establish a link with the largest IC partition reached by our algorithm(s).[12]

2 Model and solution concepts

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 $\theta_1 < \ldots < \theta_N$, $N \in \mathbb{N}$. 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 $|M| \geq 2^{|\Theta|}$. The receiver's set of decisions is $\mathbb{R}$. For every $\theta \in \Theta$, the sender's utility function is $U^\theta : \mathbb{R} \to \mathbb{R}$ and the receiver's utility function is $V^\theta : \mathbb{R} \to \mathbb{R}$.

The game 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 utility 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*:

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

- *Single-crossing*:

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

(and their associated equilibria) is questionable to address refinement issues (see, e.g., Gerardi and Myerson (2007)).

[12] We found an example in which there are three IC partitions: the nonrevealing one, which is defeated, Π_*, the limit of our iterative procedure, and another partition Π, which is also undefeated. Both Π_* and Π_* survive to the elimination of dominated strategies (see Section 4).

- *Unique maximizing arguments*:

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

- *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 utility functions of the form

$U^\theta(x) = -(\theta + b_\theta - x)^2$, $V^\theta(x) = -(\theta - x)^2$,

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

2.2 Perfect Bayesian equilibrium

We focus on pure equilibria.[13] A pure strategy for the sender, namely, a mapping $\sigma : \Theta \to M$, induces a partition of Θ (with cells $\sigma^{-1}(m)$, $m \in \sigma(\Theta)$). A pure strategy for the receiver is a mapping $\tau : M \to \mathbb{R}$.

Given the sender's strategy σ and a signal $m \in \sigma(\Theta)$, the receiver forms the posterior belief $p(.|m) \in \Delta(\Theta)$ according to Bayes rule. Equivalently, for every cell π of the partition Π induced by σ, he forms the posterior belief $p(.|\pi)$, equal to $\frac{p(\theta)}{p(\pi)}$ for $\theta \in \pi$ and 0 otherwise.

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

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

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

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

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

[13] See Section 6.3 for comments on mixed equilibria.

Given the partition Π induced by some strategy σ of the sender, the receiver has a unique, pure best reply, which consists of choosing y^π on every cell π of Π. The partition describes a *Bayesian equilibrium* if and only if :

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

Conversely, every IC partition Π induces an equilibrium and even a *Perfect* Bayesian Equilibrium (PBE), by assuming that decision $y^{\pi'}$, for *some* cell π' of Π, is made on any out of equilibrium message (with the associated belief $p(\cdot \mid \pi')$).

Summing up, every PBE is equivalent to an IC partition of Θ. This characterization of PBE goes along with identifying canonical messages as subsets L of Θ, with the meaning "my type is in L."

Furthermore, from assumptions (A0), (A1) and (A2), every IC partition consists of "intervals" of consecutive types. To see this, let Π be an IC partition, let $\theta < \theta' < \theta''$, assume that θ and θ'' are in the same cell of Π, in which decision y is made, while θ' is in another cell, in which a distinct decision y' is made. By (A1) and (IC) for θ and θ', we must have that $y < y'$. But given (A1), this contradicts IC for θ''.

2.3 Undefeated partition, forward-neologism-proof PBE

Let us adapt Mailath et al.'s (1993) notion of "undefeated" PBE to our partition characterization. Given a partition Π of Θ, we denote as $\pi(\theta)$ the cell of Π that contains θ.

Definition 1. 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 it is not defeated by any other IC partition.

We show below that undefeated IC partitions can be interpreted as being neologism-proof in a specific sense, which we call "forward-neologism-proof."

As made clear in the previous section, every PBE defines a language, which consists of the messages that are sent at equilibrium. In our setup, a PBE is described by an IC partition Π, so that the equilibrium language is defined by the cells of Π. Every set of types $L \subseteq \Theta$ that is not a cell of Π is a *neologism*.

Let us recall the definition of Farrell (1993).[14]

Definition 2. 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.

To be self-signaling, the neologism L must first improve the utility of all types in L, if the receiver reacts by making decision y^L. The interpretation of the second requirement is that, when a message different from L is sent, the receiver sticks to his original equilibrium strategy.

As Mailath et al. (1993), we consider a forward looking receiver, who performs a thought experiment before the beginning of the game and understands that, if types in L send message L, in the hope of inducing decision y^L, then the original equilibrium messages must be sent by types that are not in L. Hence, the logic leading the receiver to modify his strategy (to y^L) on L also leads him to adjust it outside of L. Anticipating this, the sender should react so as to avoid conflict between his types, namely, in such a way that types in L do not want to imitate those in $\Theta \setminus L$ and types in $\Theta \setminus L$ do not want to imitate those in L. In other words, to be credible, the neologism L must be *incentive compatible*, given some receiver's strategy choosing y^L on L or equivalently, L must be a cell of some IC partition. Summing up, we define an IC partition Π as *forward-neologism-proof* if it is immune to *incentive compatible neologisms*. Using this definition and recalling Definition 1, Π is forward-neologism-proof if and only if Π is undefeated.[15]

2.4 Dominance relation over partitions

In this section, we introduce the last basic ingredient needed to state our result: a binary relation over partitions. Let us denote as $\max L$ (resp., $\min L$) the highest (resp., lowest)

[14]Farrell (1993) does not explicitly settle the case of types that are indifferent between their original equilibrium message and the neologism. In Definition 2, these types may send either message.

[15]Mailath et al. (1993, p. 252) argue: "[...] 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 adjustment are made, we must be at an equilibrium; if not some further adjustment should be contemplated."

element of a subset of types L. Recall that $\pi(\theta)$ denotes the cell of partition Π that contains θ.

Definition 3. Let Π and Π' be partitions of Θ.

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

In other words, Π dominates 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' (see Figure 1 below).

$$\begin{array}{lccccc} \Pi': & \ldots\}, & \{\min \pi', & \ldots\ldots\ldots\ldots\ldots, & \max \pi' & \}, \{\ldots \\ \Pi: & & \ldots\ldots\}, & \{\min \pi(\max \pi'), \ldots, & \max \pi' & , \ldots\}, \{\ldots \end{array}$$

Figure 1: $\Pi \geq \Pi'$.

The dominance relation is anti-symmetric, i.e., if $\Pi \geq \Pi'$ and $\Pi' \geq \Pi$, then $\Pi = \Pi'$, but the induced order is not complete.[16] However, the completely revealing partition

$$\mathrm{CR} = \{\{\theta_1\}, \ldots, \{\theta_N\}\}$$

dominates every other partition, while the nonrevealing partition

$$\mathrm{NR} = \{\Theta\}$$

is dominated by every other partition.

Obviously, if $\Pi \geq \Pi'$, Π' cannot have more cells than Π.

3 Statement of the main result

When assumptions (A0)-(A3) are not satisfied, existence of an undefeated partition is not guaranteed (see, e.g., Matthews et al. (1991) or Olszewski (2006).[17] By contrast, the next result holds in our model:

Theorem 1. *There exists a unique IC partition Π_* such that for every IC partition Π, $\Pi_* \geq \Pi$. This partition Π_* is undefeated.*

[16] Partitions Π and Π' might be such that some maximal element of some cell of Π is pooled with higher types in Π' than in Π, and some maximal element of some cell in Π' is pooled with higher types in Π than in Π'.

[17] 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.

Note that, by the antisymmetry of the dominance relation, an IC partition dominating any other IC partition is necessarily unique. An immediate consequence of this statement is that a partition that is finer than Π_* cannot be IC.

As part of the proof of Theorem 1, we propose *a family of natural iterative optimization processes that all converge to a unique IC partition* Π_* satisfying the properties in the statement. This is done formally in Section 5. The algorithms typically start with "initial trust", namely, with the partition $\Pi_0 = \{\{\theta_1\}, \{\theta_2\}, \cdots, \{\theta_N\}\}$ associated with the sender's completely revealing strategy and the receiver's corresponding best response, namely, y^{θ_i} if the sender's type is θ_i. If initial trust is not an equilibrium, the sender has a profitable deviation, i.e., a better response. The algorithm goes on with improvement of *some* type of the sender given the receiver's strategy. Then, the receiver adjusts, with a best response to the new strategy of the sender... and so on, alternating better responses of (*some* type of) the sender and best responses of the receiver.

The previous description gives some flexibility in choosing the better response of the sender and may thus be compatible with various sequences of partitions. However we prove that all variants of the algorithm converge to the same partition Π_*. To give a little more precise description of the iterative optimization processes, we need two additional definitions. We focus on type-ordered interval partitions, in which the cells are ordered according to the order on Θ.

Definition 4. Let Π be a partition of Θ and let π and π' be cells of Π.

Type $\theta \in \pi$ *envies* cell π' if $U^\theta(y^{\pi'}) > U^\theta(y^\pi)$.

We say that type θ envies type θ' if θ envies the cell $\pi(\theta')$ containing θ'. According to Definition 4, partition Π is IC if and only if it has no envying type.

Definition 5. A partition is *Left*-Incentive Compatible (L-IC) if no type envies a cell on its left.

We can now give a step by step description of the iterative optimization processes converging to Π_*.

- **Step 0**: Start with the completely revealing partition $\Pi_0 = \{\{\theta_1\}, \{\theta_2\}, \cdots, \{\theta_N\}\}$.

  Π_0 is L-IC.

- **Step 1**: if Π_0 is not IC, consider a type θ_i that envies a type on its right; then $\widetilde{\theta} = \theta_i$ also envies θ_{i+1}; go on with $\{\theta_i, \theta_{i+1}\}$ in Π_1. Π_1 is L-IC.

- $\cdots$
- At the end of **Step r**, Π_r is left-IC.
- **Step r+1**: if Π_r is not IC, there is some type, in some cell of Π_r, $\theta \in \pi_r^n$, that envies some cell on its right. Then the largest type in π_r^n, $\widetilde{\theta} = \max \pi_r^n$, envies the next cell π_r^{n+1}; go on by merging $\left\{\widetilde{\theta}\right\}$ with π_r^{n+1} to form Π_{r+1}.
- $\cdots$
- A *unique IC partition* Π_* is reached, whatever the type $\widetilde{\theta}$ chosen at every step.

4 Basic examples

In all the examples below, the utility functions are quadratic, with a constant bias b for the sender, i.e., $U^\theta(x) = -(\theta + b - x)^2$, $V^\theta(x) = -(\theta - x)^2$.

Example 1. Let $\Theta = \{1, \cdots, 11\}$, $b = 2$ and p be uniform, namely, $p(\theta) = \frac{1}{11}$ for every θ.[18] The partition Π_* identified in Theorem 1 is $\Pi_* = \{\{1, 2\}\{3, ..., 11\}\}$. Here is a possible run of the iterative optimization process converging to Π_*:

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

[18] The uniform quadratic case is treated in more detail in Section 6.2.

In the CR partition, every type envies the next one. The sender has many possible better responses; for instance, every type improves his utility by mimicking the next one. The variants of the optimization process at the first step correspond to choosing a particular type (between 1 and 10) and moving it to the next cell. Such a modification of the partition preserves the canonical interpretation of messages as being truthful. This need not be the case if a *best* response of the sender is considered. Starting from the CR partition, every type i between 1 and 9 best responds by pretending to be type $i+2$, while types 10 and 11 send message 11. Some inference is needed to derive the partition of Θ induced by this strategy of the sender, namely, $\{\{1\},\{2\},\cdots,\{8\},\{9,10,11\}\}$.

Using the characterization of Section 6.2, the other IC partitions are NR $=\{\{1,\cdots,11\}\}$ and $\Pi=\{\{1\}\{2,\cdots,11\}\}$. As expected, Π_* dominates NR and Π. One can check that 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).[19]

In the next example, there are three IC partitions: NR, Π_* and another partition Π, with the same number of cells as Π_*. Both Π_* and Π are undefeated. As expected Π_* dominates NR and Π.

Example 2. Let $\Theta=\{1,2,3,4\}$, $b=0.6$ and

$p(1)=\frac{1}{4}$, $p(2)=\frac{1}{4}$, $p(3)=\frac{1}{100}$, $p(4)=\frac{49}{100}$.

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. The partitions NR $=\{\{1,2,3,4\}\}$, $\Pi=\{\{1\},\{2,3,4\}\}$ and $\Pi_*=\{\{1,2\},\{3,4\}\}$ are the only IC ones. As clear from the graph, 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 partition NR is defeated by Π_*, thanks to the 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.[20]

[19] In this example, Π_* turns out to be the only undefeated partition. However if, e.g., $N=8$, the IC partitions are NR and $\Pi_*=\{\{1\}\{2,...,8\}\}$. One can check that both are undefeated.

[20] In this example, both Π_* and Π survive a tedious elimination of interim weakly dominated strategies.

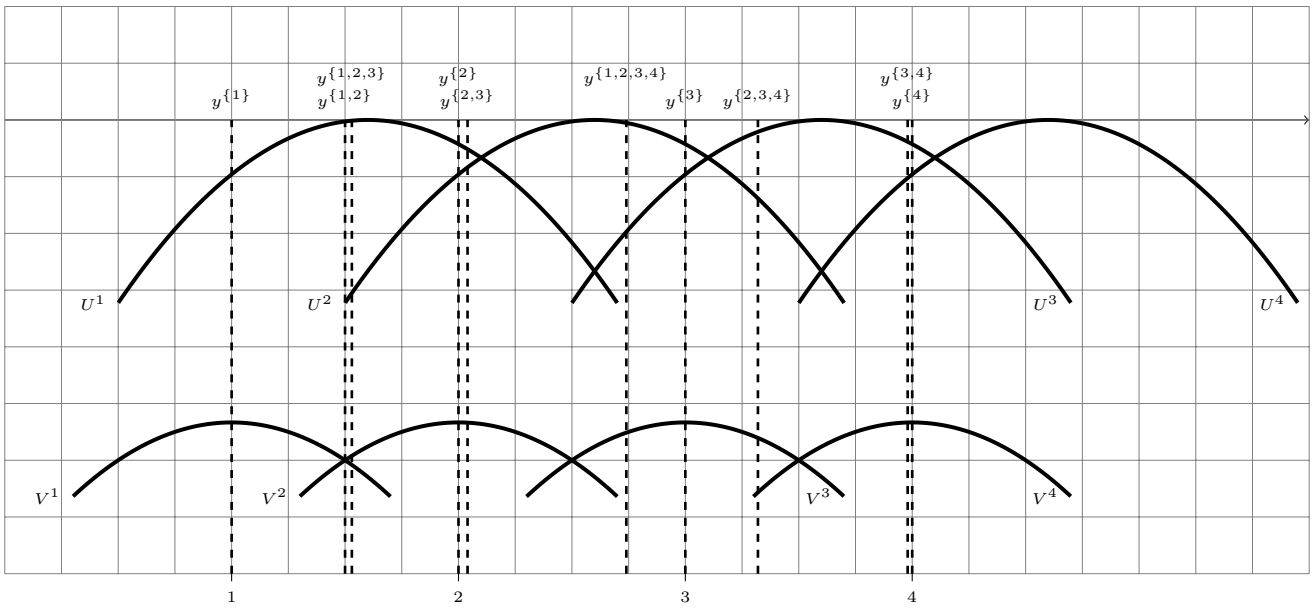

Figure 2: Utility functions and cell contingent actions in Example 2.

5 Better response dynamics (proof of the main result)

In this section, we establish Theorem 1. To do so, we propose iterative optimization processes that are a bit more general than the ones sketched in Section 3. First of all, several lemmas below only assume that the initial partition Π_0 is *L-IC*. It appears that the crucial property of the completely revealing partition is that it dominates every IC partition. Second, to emphasize the flexibility of the processes, we introduce the "envy operator" Env(.) over the set of all partitions of Θ. Given a set E of partitions, which typically can be achieved at some step r of some version of the algorithm (as described in Section 3), Env(E) contains all partitions that can be achieved at the next step.

5.1 Left-incentive compatibility along the algorithm

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 *iff* there is some type in some cell π^k who envies some cell $\pi^{k'}$, $k' \neq k$. From the single-crossing condition, this is equivalent to: if $k' > k$ (resp., $k' < k$) then type $\max \pi^k$ envies π^{k+1} (resp., type $\min \pi^k$ envies π^{k-1}). Thus, the partition is IC *iff* no type at an edge of a cell envies an adjacent cell.

Starting from an L-IC partition (see Definition 5), the following lemma allows us to *recursively* eliminate envy of some types for cells on their right.

Lemma 1. *If* $\Pi = \{\pi^1, \ldots, \pi^{|\Pi|}\}$ *is L-IC, then every partition* Π_1*, derived from* Π *by merging some envying type* $\max \pi^k$*,* $k \in \{1, \ldots, |\Pi| - 1\}$*, with the cell* π^{k+1} *it envies on its right, is also L-IC.*

Proof. Let $\Pi = \{\pi^1, \ldots, \pi^{|\Pi|}\}$ be L-IC, and let $\Pi_1 \in \text{Env}(\Pi)$, with $\Pi_1 = \{\pi_1^1, \ldots, \pi_1^{|\Pi_1|}\}$, in which type $\tilde{\theta} \in \pi^k$ is pooled with its next succeeding cell. In particular, if $\pi^k \setminus \{\tilde{\theta}\} \neq \varnothing$, then

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

and then $|\Pi_1| = |\Pi|$; and, if $\pi^k \setminus \{\tilde{\theta}\} = \varnothing$, then

$$\begin{cases} \pi_1^1 = \pi^1, \ldots, \pi_1^{k-1} = \pi^{k-1}, \\ \pi_1^k = \pi^{k+1} \cup \{\tilde{\theta}\}, \\ \pi_1^{k+1} = \pi^{k+2}, \ldots, \pi_1^{|\Pi|-1} = \pi^{|\Pi|}, \end{cases}$$

and then $|\Pi_1| = |\Pi| - 1$.

The lemma obtains since Π satisfies (L-IC), and $y^{\pi^{k+1} \cup \{\tilde{\theta}\}} \leq y^{\pi^{k+1}}$ (because $\tilde{\theta} = \min(\pi^{k+1} \cup \{\tilde{\theta}\})$ and $y^{\pi^k \setminus \{\tilde{\theta}\}} \leq y^{\pi^k}$ (because $\tilde{\theta} = \max \pi^k$): for the minimal types of the cells of Π_1, the action moves never add envy of an immediately preceding cell. □

Definition 6. Given a set of L-IC partitions E, $\text{Env}(E)$ is the set of partitions derived from partitions Π in E as follows: if Π is IC, then $\Pi \in \text{Env}(E)$; if Π is L-IC (but not IC), then merge one type $\max \pi$, for some $\pi \in \Pi$, with the cell it envies on its right, and add the resulting partition to $\text{Env}(E)$. Do this for every type in Π which envies the cell on its right.

Using Lemma 1, we recursively define, from an 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, merging a type to the cell on its right cannot be done indefinitely. Hence Lemma 1 also guarantees that every sequence $(\text{Env}^r(\Pi))_{r \geq 0}$ converges to a set of *IC* partitions in a finite number of steps.

Corollary 1. *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 Env(.), if E is a set of IC partitions, then $\text{Env}(E) = E$. 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.

In the next section, we identify a condition on partitions Π ensuring that $\overline{\text{Env}}(\Pi)$ contains a unique partition.

5.2 Dominance along the algorithm

Recall the binary relation $\geq$ defined over partitions (see Definition 3). The following lemma shows how dominance is preserved by the envy operator defined in Definition 6.

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

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

and we want to show

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

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

Let $\tilde{\theta}$ denote the type that is pooled to its next succeeding cell from Π_0 to Π_1. Then we have:

- 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) results 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, we have $\min \pi_1(\overline{\theta}_k) = \tilde{\theta}$. Then (2) might be written

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

Inequality (3) trivially holds whenever $\min \pi_0(\overline{\theta}_k) > \min \pi^k$. We show that this is always the case, that is: 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 do 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.,

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

Then, on the one hand, from $\Pi_0 \geq \Pi_1$, we have

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

and we obtain that $\pi_0^L(\overline{\theta}_k)$ 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 situation, and the (IC) condition on Π, prevents $\tilde{\theta}$ to strictly envy the action associated with $\pi_0(\overline{\theta}_k)$ when considered in $\pi_0^L(\overline{\theta}_k)$.

Formally, we respectively have

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

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

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

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

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

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

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

This combined with (6) gives

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

and $\tilde{\theta} \in \pi_r^L(\overline{\theta}_k)$ does not strictly envy the action associated with $\pi_r(\overline{\theta}_k)$. □

5.3 Convergence of the algorithm

By recursion, a direct consequence of Lemma 2 is that if $\Pi_0 \geq \Pi$, then every partition $\Pi_r \in \text{Env}^r(\Pi)$, $r \geq 0$, is such that $\Pi_r \geq \Pi$. In particular, every IC partition $\Pi_{\overline{r}} \in \overline{\text{Env}}(\Pi)$ is such that $\Pi_{\overline{r}} \geq \Pi$.

Corollary 2. *For every IC partition* Π *and every L-IC partition* Π_0 *such that* $\Pi_0 \geq \Pi$*, every IC partition* $\Pi_{\overline{r}} \in \overline{\text{Env}}(\Pi_0)$ *is such that* $\Pi_{\overline{r}} \geq \Pi$*.*

Moreover, if a partition Π_0, such as CR, satisfies $\Pi_0 \geq \Pi$ for *every* IC partition Π, then every IC partition $\Pi_{\overline{r}} \in \overline{\text{Env}}(\Pi_0)$ 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 unique. We obtain the following proposition – a detailed version of the first part of Theorem 1.

Proposition 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* $\overline{\text{Env}}(\Pi_0)$*, recursively derived by envy from any L-IC partition* Π_0 *such that for every IC partition,* $\Pi_0 \geq \Pi$ *(e.g.,* $\Pi_0 = \text{CR}$*).*

5.4 Undefeated partition

Let Π be an IC partition and Π_0 be an L-IC partition such that $\Pi_0 \geq \Pi$. The next lemma shows that the maximal element $\overline{\theta}$ of every cell of Π weakly prefers the cell containing it in any partition $\Pi_r \in \text{Env}^r(\Pi_0)$, provided that $\overline{\theta}$ is also the maximal element of a cell of Π_0 (a property that automatically holds for $\Pi_0 = \text{CR}$). The result is straightforward if $\overline{\theta}$ is the maximal element of a cell of Π_r, since in this case, $\Pi_r \geq \Pi$ implies that $\overline{\theta}$ is associated with (weakly) lower types in Π_r than in Π, resulting in a lower action in Π than in Π_r, in a region (on the left of $x^*(\overline{\theta})$) where $x \mapsto U^{\overline{\theta}}(x)$ increases. If instead $\overline{\theta} < \max \Pi_r$, then the result requires that $\overline{\theta} = \max \pi_0$, $\pi_0 \in \Pi_0$. This assumption implies that at some step r' between 0 and r, type $\overline{\theta}$ is merged *by envy* to its next succeeding cell. Since at that step r', type $\overline{\theta}$ is also the maximal element of a cell of $\Pi_{r'}$, which dominates Π, $\overline{\theta}$ also prefers the action taken in its next succeeding cell in $\Pi_{r'}$ *to the action taken on* $\overline{\theta}$ *in* Π. Then, from $r'+1$ on, $\overline{\theta}$ keeps preferring the cell containing it, up to partition Π_r.

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* $\Pi_r \in \text{Env}^r(\Pi_0)$*. 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 $\Pi = \{\pi^1, \ldots, \pi^{[\Pi]}\}$ be an IC partition, let $k \in \{1, \ldots, |\Pi|\}$, 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 $\Pi_r \in \text{Env}^r(\Pi_0)$.

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 suppose that $\pi_r(\overline{\theta}_k)$ is not a singleton for the remainder of the proof.

By Corollary 2, $\Pi_r \geq \Pi$. 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 $\pi_{r'} \in \Pi_{r'}$, $\Pi_{r'} \in \text{Env}^{r'}(\Pi_0)$, 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).

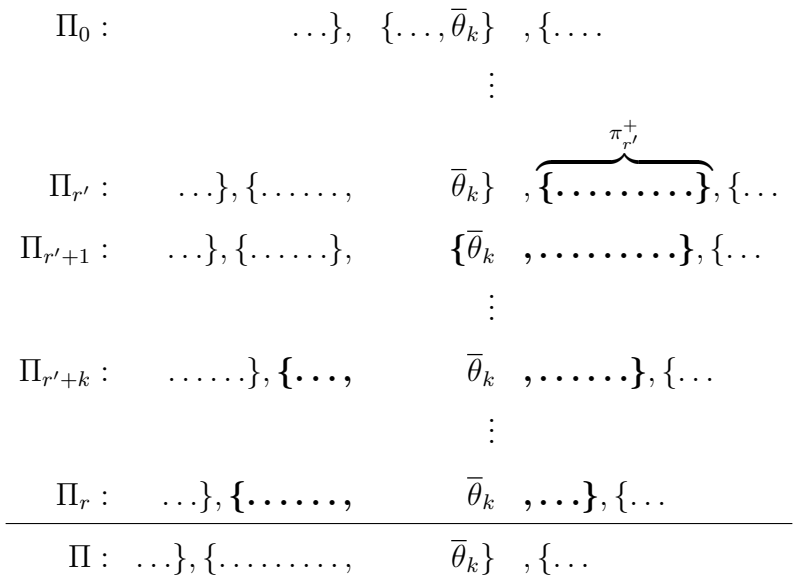

Figure 3: From $\pi^+_{r'}$ to $\pi_r(\overline{\theta}_k)$ along an envy sequence initialized at Π_0.

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 recursion (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. □

Let us take $\Pi_0 = \text{CR}$ (the completely revealing partition) in Lemma 3. The assumptions involving Π_0 are satisfied since *every* type is the maximal type of a cell of Π_0. The conclusion of the lemma thus applies to every IC partition Π and (by Corollary 1) to the limit partition Π_*, i.e., every type that is maximal in some cell π of Π has higher utility in Π_* than in Π. This is the content of the next statement.

Corollary 3. *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 $\pi_*(\overline{\theta}) = \pi$.

As an immediate consequence, no IC partition Π defeats Π_*, which establishes the second part of Theorem 1.

The previous reasoning relies heavily on sequences $(\Pi_r)_{r\geq 0}$ that start at $\Pi_0 = \text{CR}$ 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_*. Indeed going from Π_0 to Π_*, maximal types are pooled with the next succeeding cell only if they prefer to be pooled. The feature runs recursively along the full partitions. 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 3. Consider $\Theta = \{1, 3, 4, 9, 10\}$, with $p(1) = p(3) = 0.2$, $p(4) = 0.3$, $p(9) = 0.29$, and $p(10) = 0.01$. Utility 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 $\Pi_0 = \text{CR}$ 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 1, the sequence $(\text{Env}^r(\Pi_0'))_{r\geq 0}$ 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 $y^{\{3,4,9\}} \simeq 5.6$.

6 Further properties

6.1 Relationship with other refinement criteria

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

Proposition 2. *Assume there exists a neologism-proof IC partition* Π. *Then* $\Pi = \Pi_*$.

Proof. Suppose that $\Pi = \{\pi^1, \ldots, \pi^{n_\Pi}\}$, $n_\Pi \in \{1, \ldots, N\}$, is a neologism-proof IC partition. We shall show that every cell of Π also is a cell of Π_* by recursion, starting from the last cell of Π. To that end, we shall consider a specific path of the envy driven algorithm initialized at $\Pi_0 = \text{CR}$, i.e., a sub-sequence $(\Pi_{r[t]})_{t\in\mathbb{N}}$, $t \mapsto r[t]$ increasing, of a sequence of partitions $(\Pi_r)_{r\in\mathbb{N}}$ such that $\Pi_0 = \text{CR}$ and for every $r \geq 0$, $\Pi_{r+1} \in \text{Env}(\Pi_r)$, that achieves Π.

Formally, given $(\Pi_r)_{r\in\mathbb{N}}$ as defined above, we define a condition $P(t)$, $t \in \{0, \ldots, |\Pi|\}$, 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

$$\Pi_r = \{\underbrace{\{\theta_1\}, \ldots, \ldots, \{\theta_{k_r-1}\}}_{\text{or } = \varnothing \text{ if } t = |\Pi|}, \underbrace{\pi_r^{n_r-t}, \ldots, \pi_r^{n_r-1}}_{\text{or } = \varnothing \text{ if } t = 0}\}, \tag{13}$$

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

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

(ii) 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 $t = |\Pi|$, then, from (13) and (i), the partition achieved at rank $r[t] = r[|\Pi|]$ is Π, and therefore $\Pi = \Pi_*$.

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

Initialization. Condition $P(0)$ trivially holds, because (13) holds at $\Pi_0 = \text{CR} = \{\{\theta_1\}, \ldots, \{\theta_N\}\}$, and properties (i) and (ii) are irrelevant when $t = 0$.

Heredity. Given $0 \leq t < |\Pi|$ such that condition $P(t)$ holds, with associated rank $r = r[t]$ and integers n_r and k_r, let us denote by

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

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

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

$$\Pi_{r'} = \{\underbrace{\{\theta_1\}, \ldots, \ldots, \{\theta_{k_r-\ell_r-1}\}}_{\text{or } = \varnothing \text{ if } t = |\Pi|-1}, \pi, \underbrace{\pi_r^{n_r-t}, \ldots, \pi_r^{n_r-1}}_{\text{or } = \varnothing \text{ if } t = 0}\} \tag{14}$$

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

(ii) 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 $t < |\Pi|$ 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 (14) 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. To that end, we define, for $\ell, \ell' \in \{1, \ldots, \ell_r\}$, $\ell \leq \ell'$:

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

In particular, we have $T_{1,\ell_r} = \pi$, and for every $\ell \in \{2, \ldots, \ell_r\} \neq \varnothing$ (recall that π is not a singleton): $T_{1,\ell-1} = \{\theta_{k_r-(\ell-1)}, \ldots, \theta_{k_r-1}\}$ consists of the $\ell - 1$ highest types of π, whereas $T_{\ell,\ell_r} = \{\theta_{k_r-\ell_r}, \ldots, \theta_{k_r-\ell}\}$ 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 $\ell \in \{1, \ldots, \ell_r\}$, if type $\theta_{k_r-\ell}$ is pooled with the lowest types of π, i.e. if it is associated with $y^{T_{\ell,\ell_r}}$, it envies any pooled next succeeding set of types, i.e., it envies $y^{T_{\ell',\ell-1}}$ for every $\ell' \in \{1, \ldots, \ell-1\}$.

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 ends 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 $\theta_{k_r-\ell_r}$ is associated with $y^{\theta_{k_r-\ell_r}}$ and is moved by envy to cell $T_{1,\ell_r-1} = \{\theta_{k_r-(\ell_r-1)}, \ldots, \theta_{k_r-1}\}$. 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 that, 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 achieves the proof of the claim concerning (ii).

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

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 $\ell \in \{1, \ldots, \ell_r\}$, for every $\ell' \in \{1, \ldots, \ell - 1\}$,

$$U^{\theta_{k_r-\ell}}(y^{T_{\ell,\ell_r}}) < U^{\theta_{k_r-\ell}}(y^{T_{\ell',\ell-1}}). \tag{15}$$

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

$$y^{T_{\ell,\ell_r}} < y^\pi < y^{T_{1,\ell-1}}, \tag{16}$$

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 recursion 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 $\ell = 2$:

$$U^{\theta_{k_r-\ell}}(y^{\pi}) < U^{\theta_{k_r-\ell}}(y^{T_{1,\ell-1}}). \tag{17}$$

Let us now show, by a similar argument, that (17) holds recursively along $\ell = 2, \ldots, \ell = \ell_r - 1$ with regard to the set $T_{1,\ell-1}$ and the type $\theta_{k_r-\ell}$. More precisely, for every $\ell \in \{2, \ldots, \ell_r\}$, 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 recursion 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 (17) does not hold, then there is a greatest $\ell \in \{2, \ldots, \ell_r\}$, namely $\bar{\ell}$, such that (17) holds for every $\ell \in \{2, \ldots, \bar{\ell}\}$ and $T_{1,\bar{\ell}-1}$ 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 (17) holds.

Now given $\ell \in \{2, \ldots, \ell_r\}$, the second inequality in (16) and inequality (17) 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 (16) implies

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

Together with (17), we obtain

$$U^{\theta_{k_r-\ell}}(y^{T_{\ell,\ell_r}}) < U^{\theta_{k_r-\ell}}(y^{T_{1,\ell-1}}). \tag{19}$$

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 below 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 (15) 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 (15), 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 (19) gives the result. □

Remark 1. *As a consequence of Theorem 1 and Proposition 2, in our model, an IC partition that is neologism-proof is undefeated, a property that does not hold in general.*[21]

The result established above is actually stronger than Proposition 2. Indeed, the fact that Π is neologism-proof is only used to establish inequality (17), i.e.,

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

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 $\ell \in \{1,\ldots,|\pi|-1\}$,

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

Thinking of Π as to a putative equilibrium to be tested and recalling Definition 2, 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.[22]

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.

Let us turn to the "No Incentive to Separate"(NITS) criterion of Chen et al. (2008), which was developed for sender-receiver games with a continuum of types and requires that the lowest type has no incentives to signal itself (if somehow he could). Applying the criterion readily to our setting, an IC partition $\Pi = \{\pi^1,\ldots,\pi^{|\Pi|}\}$ (in which $|\pi^1| \geq 2$) satisfies NITS if

$$U^{\theta_1}(y^{\pi^1}) > U^{\theta_1}(y^{\theta_1}).$$

[21] 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 2, because type θ_j, $j \neq i$, would benefit from it.

[22] 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 2, which, as argued before, relies on the receiver's inertia. Note that, under our assumptions, condition (**) 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}.

We establish below that the partition Π_* satisfies the latter property.

Proposition 3. *The partition Π_* satisfies NITS.*

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

$$U^{\theta_1}(y^{\theta_1}) \leq U^{\theta_1}(y^{\pi_*^1}). \tag{20}$$

If $\pi_*^1 = \{\theta_1\}$, (20) 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.,

$$U^{\theta_1}(y^{\theta_1}) < U^{\theta_1}(y^{\pi_r^2}). \tag{21}$$

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

$$U^{\theta_1}(y^{\theta_1}) < U^{\theta_1}(y^{\pi_{r+1}^1}). \tag{22}$$

If $y^{\pi_{r+1}^1} \leq x^*(\theta_1)$, (22) 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 (22) from (21). This ends the proof of (22). 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 show (20). If $y^{\pi_*^1} \leq y^*(\theta_1)$, (20) 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 (20) from (22). □

Remark 2. *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_*, which is the only candidate to NSSHT, may not satisfy this condition.

Example 4. Let us consider the uniform quadratic case (as in Section 4) 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.

By contrast with NITS, which focuses on the lowest cell of a partition, NSSHT prevents the "incentives to separate" of every lower set of types of *every cell* of the underlying partition Π. This feature looks more relevant in a discrete type setting, because the border effects that necessarily occur between cells in the continuous case may have no counterpart in the discrete one.[23] However, as highlighted by the previous example, condition (**), may remove every partition, even if it just applied to the first cell. Hence, one would like to identify a criterion stronger than NITS but weaker than NSSHT, which, as the iterative optimization processes considered above, would only select the partition Π_*. Such a project is beyond the scope of the present paper, and we leave it for future research.

6.2 Ex ante optimal PBE

In the uniform quadratic case, namely, if $\Theta = \{1, 2, ..., N\}$, $p(\theta) = \frac{1}{N}$ for every θ and

$$U^\theta(x) = -(\theta + b - x)^2 \text{ and } V^\theta(x) = -(\theta - x)^2,$$

Frug (2016) proposes a specific procedure to construct an equilibrium partition and then establishes that this partition corresponds to an ex ante Pareto optimal equilibrium. We establish below that Frug's procedure leads to the partition Π_* identified in Theorem 1, so that:

Proposition 4. *In the uniform quadratic case, partition Π_* corresponds to an ex ante Pareto-optimal PBE.*

For simplicity, we focus on the case where $4b$ is an integer and set $k_0 = 4b - 2$. Let $\Pi = \{\pi^1, \ldots, \pi^{|\Pi|}\}$ be an interval partition. Recalling that IC holds as soon as no type at an

[23] For instance, with a finite set of types, in case of a small bias of the sender (relative to the distance between the types), several IC partitions may have the same first cell, irrespective of the way types are pooled afterwards. This cannot happen in the continuous case.

edge of a cell envies an adjacent cell, Π is IC if and only if the highest type in π^j does not envy π^{j+1} and the lowest type in π^{j+1} does not envy π^j, namely, if and only, for every j,

$$\mid \pi^j \mid + k_0 \leq \mid \pi^{j+1} \mid \leq \mid \pi^j \mid + k_0 + 4. \tag{23}$$

6.2.1 Frug's (2016) procedure

If $k_0 \leq 0$, i.e., if $b \leq \frac{1}{2}$, Frug's (2016) procedure leads to the fully revealing partition, i.e., $\{\{1\}, ..., \{N\}\}$. This is the starting partition in a possible version of our algorithm. If $b \leq \frac{1}{2}$, no type θ envies type $\theta + 1$ (since $\theta + 1 \geq \theta + 2b$) and the algorithm stops at step 1.

Henceforth, we thus assume $k_0 \geq 1$. To describe Frug's (2016) procedure, let us set

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 1 + k_0 \\ &\cdots \\ x_{j+1} &= x_j + k_0 \quad = 1 + jk_0. \end{aligned}$$

Let $k \equiv k(N)$ 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

$$N - \sum_{j=1}^{k} x_j = qk + r. \tag{24}$$

Frug's (2016) equilibrium partition $\Pi \equiv \Pi(N)$ is an interval partition of k cells π^j, $j = 1, \ldots, k$, containing $\mid \pi^j \mid$ types:

$$\begin{aligned} \mid \pi^j \mid = \quad & x_j + q \qquad j = 1, \ldots, k - r \\ & x_j + q + 1 \quad j = k - r + 1, \ldots, k. \end{aligned}$$

Condition (23) clearly holds.

6.2.2 Proof of Proposition 4

In this section, we show that a fully specified version of the algorithm described in Section 3 converges to Frug's (2016) partition. This version (illustrated on Example 1) consists of choosing, at every step, the envying type $\widetilde{\theta}$ in the cell that is "most on the right".[24]

[24]Then, as the algorithm always prescribes, $\tilde{\theta}$ is chosen as the highest type in the cell and is moved to the next cell. There is flexibility only in the choice of the cell containing an envying type.

We proceed by induction on N. Let the set of types be $\{1, 2, ..., N+1\}$ and let $\Pi(N+1)$ be Frug's associated partition. By the induction assumption, our algorithm reaches Frug's partition when the set of types is $\{2, ..., N+1\}$. Let us call this partition $\Pi(N)$, with a little abuse of notation. Let us perform the algorithm over $\{1, ..., N+1\}$; after a few steps, one reaches the partition $\{\{1\}, \Pi(N)\}$ (again with a little 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,

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

In other words, Frug's partitions satisfy $\Pi(N+1) = \{\{1\}, \Pi(N)\}$. 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 (23)).

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 (24), $q \in \{0, \cdots, k_0 - 1\}$. 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 (23)) 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 (23)) 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 is exactly $\Pi(N+1)$.

Example 5. To illustrate the proof above, let us take $b = 2$ as in Example 1 (equivalently, $k_0 = 6$). If $N \leq 7$, then $\Pi(N)$ contains a single cell (nonrevealing equilibrium). If $N = 8$, then $\Pi(8) = \{\{1\}\{2, ..., 8\}\}$. This illustrates case 1 above. For $N = 11$ and $N = 12$, $\Pi(N)$ contains two cells: $\Pi(11) = \{\{1, 2\}\{3, ..., 11\}\}$. To go from $N = 11$ to $N = 12$, by induction, one starts with $\{\{1\}\{2, 3\}\{4, ..., 12\}\}$; type 1 envies $\{2, 3\}$; after one step, the algorithm converges to $\{\{1, 2, 3\}\{4, ..., 12\}\} = \Pi(12)$. This illustrates case 2.

The next example illustrates that Proposition 4 no longer holds if types are not uniformly distributed. In this case, the partition Π_* identified in Theorem 1 may not be the receiver's ex ante preferred one.

Example 6. Assume $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and

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

Then $\Pi_* = \{\{\theta_1, \theta_2\}, \{\theta_3\}\}$ and $\Pi = \{\{\theta_1\}, \{\theta_2, \theta_3\}\}$ are IC partitions.[25] The receiver's loss $\sum_{i \in \{1,2,3\}} p_i (y^{\Pi(\theta_i)} - \theta_i)^2$ is greater at Π_* than at Π if, e.g., $b = 1$ and

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

6.3 Mixed strategies

In this section, we sketch how our analysis can be extended to mixed strategies. First, mixed PBE can be characterized as IC "pseudo-partitions."Then Definitions 1 and 3 can be extended to show that the partition Π_* identified in Theorem 1 dominates every IC pseudo-partition and cannot be defeated by any IC pseudo-partition.

A mixed strategy for the sender is a mapping $\sigma : \Theta \to \Delta(M)$, where $\Delta(M)$ denotes the set of probability distributions over M. At an equilibrium, 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.

[25] Partition Π is defeated by Π_* since $\{\theta_3\}$ is an IC neologism.

By contrast, the sender may randomize in a best response, provided he is indifferent between the various messages he sends. Of course, he cannot benefit from sending a message of zero probability. Using assumptions (A0) and (A2) again, the sender can only randomize between *two* different messages.[26]

From (A1), mixed equilibria (including the pure ones studied along the previous sections) can be represented by IC "pseudo - (interval) partitions", with "pseudo - cells"

$$\sigma^{-1}(m) = \{\theta \in \Theta : \sigma(\theta)(m) > 0\},$$

such that, for $m \neq m'$, $\sigma^{-1}(m) \cap \sigma^{-1}(m')$ is not necessarily empty but contains at most one type. IC pseudo-partitions thus take the form

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

with maximal element of the jth cell (θ_{i_j}) equal the minimal element of the next one (θ'_{i_j}) when type $\theta_{i_j} = \theta'_{i_j}$ mixes between the messages sent respectively by the previous types and the next ones.[27] Note that an IC pseudo-partition can contain singletons. By assumption (A3), a singleton can mix with the next pseudo-cell but not with the previous one. Hence an IC pseudo-partition contains at most N pseudo-cells.

Let Π be a partition and Π' an IC pseudo-partition. Definition 3 readily applies to Π and Π', namely, 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, $\pi(\max(\pi'))$.

The next result is established in Frug (2016) (in the proof of Proposition 1).

Lemma 4. *Let Π be an IC pseudo-partition. There exists an IC partition $P(\Pi)$, with the same number of cells as Π, such that $P(\Pi) \geq \Pi$.*

Since Π_* dominates every IC partition, we get the following property, which extends the first part of Theorem 1.

Corollary 4. *For every IC pseudo-partition Π, $\Pi_* \geq \Pi$.*

[26] See Frug (2016) for a more detailed analysis of mixed equilibria.

[27] The complete description includes, for every mixing type $\theta_{i_j} = \theta'_{i_j}$, the probability distribution over the two corresponding cells.

We go on by showing that the second part of Theorem 1 extends as well, i.e., Π_* is not defeated by any IC pseudo-partition. Proceeding as above for Definition 3, Definition 1 is a way to describe when an IC pseudo-partition Π' defeats an IC partition Π. Indeed, every type's utility is well-defined in an IC pseudo-partition (a type belonging to two different pseudo-cells gets the same utility in both of them). Adopting this definition, we establish below that Lemmas 2 and 3 go through. A prerequisite follows from Lemma 4: the completely revealing partition, CR, dominates every IC pseudo-partitions, in the same way as it dominates every partition.

Lemma 5 (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 a pseudo IC 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 $\Pi = \{\pi^1, \ldots, \pi^{n_\Pi}\}$, with $n_\Pi = |\Pi|$, and choose type $\overline{\theta}_k = \max \pi^k$ for some $k \in \{1, \ldots, n_\Pi - 1\}$, that 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 proceeds exactly in the same way as for Lemma 2. □

Lemma 6 (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 $\Pi = \{\pi^1, \ldots, \pi^{[\Pi|}\}$ be an IC pseudo partition, let $k \in \{1, \ldots, |\Pi| - 1\}$ and let $\overline{\theta}_k = \max \pi^k$ that 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 suppose that $\pi_r(\overline{\theta}_k)$ is not a singleton for the remainder of the proof.

From Lemma 5, we have, by recursion, $\Pi_r \geq \Pi$. 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 greater than y^{π^k}, and lower than $y^{\pi_r(\overline{\theta}_k)}$. In particular,

$$y^{\pi_r(\overline{\theta}_k)} \geq y^{\pi^k}. \tag{25}$$

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, the proof proceeds exactly in the same way than the proof of Lemma 3. □

Proceeding as in Section 2.3, we obtain that

Corollary 5. *The partition Π_* is not defeated by any IC pseudo-partition.*

We conclude with some general remarks about the previous extensions.

Remark 3. *The extension of Definition 3 used above is appropriate to define when a partition dominates a pseudo-partition. More generally, the dominance relation can be defined between pseudo-partitions. A way to do it is to rank the cells of any pseudo-partition from the end, by giving rank N to the last one, $N-1$ to the next to last one, and so on. Given a pseudo-partition Π, let $R(\pi, \Pi)$ denote the rank of the pseudo-cell π in Π. We now define the rank $r(\theta, \Pi)$ of type θ in pseudo-partition Π as the expected rank of the pseudo-cells containing θ. More precisely, if type θ belongs to a single pseudo-cell, namely, if type θ uses a pure strategy, $r(\theta, \Pi) = R(\pi(\theta), \Pi)$; if type θ belongs to two adjacent pseudo-cells π_j and π_{j+1} and sends the associated messages with probability ρ and $1-\rho$ respectively, then $r(\theta, \Pi) = \rho R(\pi_j, \Pi) + (1-\rho)R(\pi_{j+1}, \Pi)$. For every pseudo-partition Π, the mapping $r(\cdot, \Pi) : \Theta \to \mathbb{R}$ takes values between 1 and N and is increasing. One can then define the dominance relation between IC pseudo-partitions Π' and Π as follows: $\Pi \geq \Pi'$ if for every θ, $r(\theta, \Pi) \leq r(\theta, \Pi')$.*

Remark 4. *While the literal extension of Definition 1 provides a way to check whether a pure equilibrium can be defeated by a mixed one in our model, the interpretation of such a literal extension in terms of neologisms looks problematic, because neologisms consist of messages rather than strategies. The formulation of an appropriate notion of forward-neologism-proofness for mixed equilibria is beyond the scope of the current paper, which focuses on a specific class of cheap talk games.*

References

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

Chen, Ying and Sidartha Gordon, "Information transmission in nested sender–receiver games," *Economic Theory*, 2015, *58* (3), 543–569.

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

Cho, In-Koo and David M Kreps, "Signaling games and stable equilibria," *The Quarterly Journal of Economics*, 1987, *102* (2), 179–221.

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

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.

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

Gerardi, Dino and Roger B Myerson, "Sequential equilibria in Bayesian games with communication," *Games and Economic Behavior*, 2007, *60* (1), 104–134.

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.

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.

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.

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