Online Appendix For

Interim Bayesian Persuasion:First Steps

Eduardo Perez-Richet*

January 13, 2014

A Three Refinements and a Proof

In this section I define the refinements for general sender-receiver games. The type set is denoted by T, $p(\cdot)$ denotes the prior belief, and for any set $S \subseteq T$, $p(\cdot|S)$ denotes the restriction of the prior to S. Then I prove Proposition 3.

$\mathcal{R}_1$. The idea of this family of refinements is that an out-of-equilibrium action should be interpreted as an attempt by a player to signal that she would prefer to coordinate on another equilibrium in which this action is played. If the beliefs of the players in the original equilibrium do not anticipate this, the original equilibrium should be discarded. This idea was developed independently in Mailath et al. (1993) and Umbhauer (1994) with small differences. Formally, in a sender-receiver game with general action set for the sender, if T is the type set of the sender and e is the original equilibrium, consider an equilibrium e' and an action of the sender a' which is on the equilibrium path of e' but off the equilibrium path of e'. Then let $T^+ \subseteq T$ be the set of types that strictly prefer e' to e and use action a' in e', and $T^0 \subseteq T$ be the set of types who are indifferent between e and e' and use a' in e'. Then the belief of the receiver that follows the use of a' in the initial equilibrium e should be in the convex hull of $p(.|T^+)$ and

*École Polytechnique, e-mail: eduardo.perez@polytechnique.edu

$p(.|T^+ \cup T^0)$. This means that the receiver should believe that all types in T^+ send a' while types in T^0 may send a' with positive probability. If this does not hold, then e is defeated by e'. The refinement retains only undefeated equilibria. Unlike Mailath et al. (1993) or Umbhauer (1994), I do not require all types who use a' in e' to prefer e' to e, or even some best response of the receiver to her belief in following a' in e'.

This refinement raises an issue absent under the original definitions. In the model of this paper, a deviation to an information system used in a pooling equilibrium that only the high type prefers is attributed to the high type. The problem is that such an attribution would make the low type want to use this deviation as well. It does not seem to be particularly problematic in this game because, if the receiver were to attribute this deviation to both types equally, then the high type (and only the high type) would still find the deviation profitable. But the logic of this attribution may seem unsatisfying (note however that this problems is absent in any equilibrium satisfying the refinement). The next two refinements tackle this issue.

$\mathcal{R}_2$. This refinement is inspired from Myerson (1983). The difference is that the sender announces an information structure instead of a mechanism. When the sender announces a mechanism, his announcement includes the suggestion of a course of action for the receiver, and it is natural to restrict potentially destabilizing mechanisms to be incentive compatible given the beliefs they may generate. When the sender merely announces an information structure, the receiver should best respond given her beliefs, but that does not entail any natural restriction on the announcements of the sender. In order to define core outcomes, I consider a sender-receiver game, with general finite type set T for the sender, and where the sender announces an information system π (defined for the general type set T). Then an information system π is a core information system if it is an equilibrium and there is no other information system $\pi' \neq \pi$ and set $S \subseteq T$, such that for every belief $p(\cdot|S')$ of the receiver, where $S \subseteq S' \subseteq T$, any type $t \in S$ strictly prefers the outcome obtained when the receiver best responds to π' to the initial equilibrium outcome. The motivation is as follows. Suppose that π is not a core information system. Then there exists a subset of types S that would benefit from any beliefs that restricts

the prior to any superset of S. Then any type in S could credibly announce π', and tell the receiver "my type is in S." The receiver does not have to believe that the sender is indeed in S, but she should account for the fact that all types in S are strictly better off as long as she believes that they make this statement.

This refinement tackles the logical difficulty with $\mathcal{R}_1$ since a deviation must be profitable to those who initiate it if it is correctly attributed to them, but also if it is attributed to any larger set of types, thus anticipating the fact that some types may try to pool on the deviation.

$\mathcal{R}_3$. As in the former paragraph. I describe the refinement for a sender-receiver game with general type set T, and where the sender announces an information system π. As in Farrell (1993), I assume that statements of the kind "my type is in S" are available for every $S \subseteq T$. Consider an equilibrium e and an information system π' which is never played in e. When deviating to π', the sender can also announce that her type belongs to some set $S_0 \subseteq T$. Then let S_1 be the set of types that strictly benefit from the best response of the receiver to π' under the belief $p(\cdot|S_0)$ relative to the initial equilibrium, and so on, so that S_{k+1} is the set of types that strictly benefit from the best response of the receiver to π' under the belief $p(\cdot|S_k)$ relative to the initial equilibrium. The sequence stops if the empty set is ever reached. The types in $\bigcup_k S_k$ are those who could be tempted to use the deviation π' together with the announcement "my type is in S_0." Therefore the initial equilibrium is deemed unreasonable if it can only be supported by a belief $q \in \Delta(T)$ that does not lie in the convex hull of the set $\{p(\cdot|S_1), p(\cdot|S_2), p(\cdot|S_3), \cdots\}$. If the sequence is empty ($S_1 = \emptyset$), then all beliefs are allowed. Note the difference with Farrell (1993), which would require the existence of a set S_0 such that $S_1 = S_0$ (in Farrell (1993), the deviation is the announcement itself, whereas here it consists in a choice of a different information system accompanied with the announcement).

This refinement tackles the logical difficulty with $\mathcal{R}_1$ since a deviation must be profitable to those who initiate it if it is correctly attributed to them, but also if it is attributed to a larger set of types that may pool on the deviation. The difference with $\mathcal{R}_2$ is that $\mathcal{R}_3$ is more selective about assessing the types that would pool on the deviation.

Proof of Proposition 3. First consider $\mathcal{R}_1$. Let (π, Σ) be an information system such that the associated equilibrium outcome in $\mathcal{P}(p, \mathcal{S})$ is not in $\mathcal{H}(p, \mathcal{S})$ (so full revelation must not be available).Then consider any information system $\pi' \neq \pi$ such that the associated equilibrium outcome is in $\mathcal{H}(p, \mathcal{S})$. Then the high type must prefer the new outcome to the original. Suppose first that the low type prefers the original equilibrium outcome. Then after observing the deviation π', the receiver who, according to $\mathcal{R}_1$, is assumed to interpret it as an attempt to coordinate on the new equilibrium must believe that this message comes from the high type. If she did, however, the original equilibrium would not be an equilibrium as both types would benefit by deviating to π'. Suppose now that the low type is weakly prefers the new equilibrium. Then after observing the deviation π', the receiver must believe that she faces the high type with probability $p' \geq p$. However the original equilibrium cannot be supported by such a belief, since by deviating to π' the high type would get both a more favorable belief p' and a more favorable information system. This shows that all selected equilibrium outcomes lie in $\mathcal{H}(p, \mathcal{S})$. To show that the two sets are in fact equal, consider an information system π that leads to an equilibrium outcome in $\mathcal{H}(p, \mathcal{S})$. Since the high type cannot improve her situation, the refinement does not prevent from believing that any deviation is originated by the low type, and such beliefs clearly support the equilibrium.

Now consider $\mathcal{R}_2$. Let π be an information system such that the associated equilibrium outcome in $\mathcal{P}(p, \mathcal{S})$ is not in $\mathcal{H}(p, \mathcal{S})$. Consider another information system π' with an associated equilibrium outcome in $\mathcal{H}(p, \mathcal{S})$. Then let $S = \{H\}$. Clearly the high type prefers the outcome associated to the information system π' and the belief $p(\cdot|S)$ since the latter must put probability one on the high type. Now consider $S' = \{H, L\}$. Then $p(\cdot|S')$ is simply the prior, and since π' is in $\mathcal{H}(p, \mathcal{S})$, the best response to π' when the belief is the prior leads to a better outcome than the equilibrium associated with π. Therefore the initial equilibrium is not a core equilibrium. This proves that all core equilibrium outcomes lie in $\mathcal{H}(p, \mathcal{S})$. Now consider an information system π such that the associated equilibrium outcome lies in $\mathcal{H}(p, \mathcal{S})$. It can be supported by the belief that any other choice is due to the low type. Suppose that this equilibrium is not a core equilibrium. Then the high type would have to strictly prefer the outcome associated

with a different information system π' under the belief $p(.|T)$, which is simply the prior. But then that would contradict the fact that π together with the prior leads to a high type optimal outcome.

Now consider $\mathcal{R}_3$. Let π be an information system such that the associated equilibrium outcome in $\mathcal{P}(p, \mathcal{S})$ is not in $\mathcal{H}(p, \mathcal{S})$. Consider another information system π' with an associated equilibrium outcome in $\mathcal{H}(p, \mathcal{S})$. Suppose that the receiver deviates from the original equilibrium by choosing π' and at the same time suggests to the receiver that she is the high type, so $S_0 = \{H\}$. The receiver must realize that both types would benefit if she were to believe the suggestion, so $S_1 = \{H, L\}$, and the corresponding belief is exactly the prior. If that is indeed the receiver's belief, she will reproduce the outcome associated with π'. This outcome makes the high type strictly better of. There are two cases. First, if it is does not make the low type strictly better of, then $S_2 = \{H\}$, and the sequence generated is therefore $S_k = \{H\}$ for every even k, and $S_k = \{L, H\}$ for every odd k. Second, if both types are better of under the outcome obtained when the receiver best responds to π' with a belief equal to the prior. Then $S_2 = S_1 = \{L, H\}$, and that pins down the sequence $S_k = \{L, H\}$ for every $k \geq 1$. In both cases, the possible beliefs that support the initial equilibrium following a deviation to π' must lie in $[p, 1]$, but that clearly makes this deviation profitable for the high type, so the initial equilibrium cannot satisfy the refinement. Note that we could have used the suggestion $S_0 = \{L, H\}$ to get the same result. This proves that all equilibrium outcomes that satisfy $\mathcal{R}_3$ lie in $\mathcal{H}(p, \mathcal{S})$. Now consider an information system π such that the associated equilibrium outcome lies in $\mathcal{H}(p, \mathcal{S})$. Consider any deviation π'. If the suggestion of the receiver is $\{L\}$, then $S_1 = \emptyset$ and the belief that puts all the weight on the low type is allowed following this deviation and it supports the original equilibrium. If the suggestion is $\{L, H\}$, then the associated belief is the prior, but since the original equilibrium is high type optimal, the set S_1 is either $\{L\}$ or the empty set. If $S_1 = L$, then $S_2 = \emptyset$. In both cases, the belief that puts all the weight on the low type is allowed and supports the original equilibrium. Finally, suppose the suggestion of the receiver is $S_0 = \{H\}$. Then $S_1 = \{L, H\}$, and S_2 is either $\{L\}$ or the empty set. So the prior is a possible belief in both cases, and it supports the original equilibrium. $\square$

B Remaining Proofs

Proof of Proposition 1. Suppose that there exists a fully separating equilibrium in which the low type plays π and the high type plays $\pi' \neq \pi$. Then the high type is validated with probability 1 and the low type with probability 0. If the low type deviates to π', she is validated with probability 1 unless π' is fully revealing. So for this to be a separating equilibrium, π' must be fully revealing. But then the same outcome is obtained in a pooling equilibrium in which both types choose π', which can be supported by believing that any deviation can only be initiated by the low type. □

Proof of Proposition 2. If perfect revelation is available, the high type can ensure validation with probability 1 by deviating to full revelation, hence any equilibrium must satisfy $\nu = 1$. Any outcome in $\mathcal{P}(p, \mathcal{S})$ such that $\nu = 1$ can clearly be supported as an equilibrium which concludes the proof of 1.

Let $(1, \hat{\nu})$ be as in the proposition. If $\hat{\nu} = 1$, full revelation must be available and we are back to 1, so suppose $\hat{\nu} < 1$. I look for information systems that can generate the outcome $(1, \hat{\nu})$. The only way for the low type to be rejected with probability 1 while the high type is validated with positive probability is if the information system partially separates the two types: there must exist some signal realizations that only the high type can send, and following which the receiver validates, and some signal realizations that can be sent by both types or only the low type and following which the receiver rejects with probability 1. Furthermore, the probability that a signal that can be generated only by the high type occurs must be exactly $\hat{\nu}$. Hence there must exist an information system that proves the high type with probability $\hat{\nu}$. But then the high type can always ensure a validation probability of $\hat{\nu}$ by deviating to this information system, so any perfect Bayesian equilibrium must give her a validation probability of at least $\hat{\nu}$. Now it must also be the case that no deviation can give the high type a validation probability $\nu > \hat{\nu}$ for that would mean that $(1, \nu) \in \mathcal{P}(p, \mathcal{S})$, a contradiction. Clearly, then, every outcome in $(\rho), \nu) \in \mathcal{P}(p, \mathcal{S})$ can be supported as a pooling equilibrium.

To prove the last point, note by 1. and 2. that there cannot exist any information system

such that the high type can prove her type with positive probability. Therefore any outcome in $\mathcal{P}(p, \mathcal{S})$ can be supported as an equilibrium if the receiver believes that any deviation comes from the low type exclusively. □

C D1: An Example

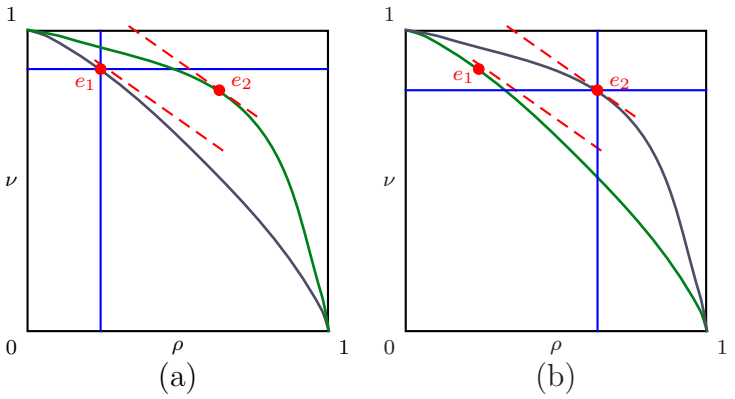

Figure 1: Example – The high type optimal equilibrium is e_1, but the only equilibrium that satisfies D1 is e_2. (a) shows why e_1 cannot satisfy D1, while (b) shows why e_2 satisfies D1.

In this example the sender only has two information systems π_1 and π_2 available, and the sets of possible outcomes generated by these information systems under all beliefs, $\cup_{\beta\in[0,1]}\mathcal{P}(\beta, \pi_1)$ and $\cup_{\beta\in[0,1]}\mathcal{P}(\beta, \pi_2)$, are represented respectively by the lower and the upper curve on Figure 1. The receiver is pro validation, and her indifference sets over outcomes are represented by the dashed red lines. So the two possible equilibria are e_1 and e_2. Consider e_1 and a deviation in which the sender announces π_2 instead of π_1. The best responses of the receiver that are preferred by the high type given this deviation correspond to the outcomes that lie on the portion of the higher curve that is above the horizontal blue line that goes through e_1 in panel (a), while those that are preferred by the low type are the ones that lie on the portion of the higher curve that is to the left of the vertical blue line. Clearly, according to D1, the receiver should attribute the deviation to the high type, but e_1 cannot be supported if that is the case. Now consider e_2 and a deviation in which the sender announces π_1. The best responses of the receiver that are preferred by the high type given this deviation are the ones that lie on the

portion of the lower curve that is above the horizontal blue line in panel (b), while those that are preferred by the low type are the ones that lie on the portion of the lower curve that is to the left of the vertical blue line. According to D1, the receiver should attribute the deviation to the low type, and since this belief is compatible e_2, this equilibrium passed the test. However it is easy to see that e_1 is the unique high type optimal equilibrium.

References

Farrell, J. (1993): "Meaning and Credibility in Cheap Talk Games," *Games and Economic Behavior*, 5, 514–531.

Mailath, G. J., M. Okuno-Fujiwara, and A. Postlewaite (1993): "Belief-Based Refinements in Signaling Games," *Journal of Economic Theory*, 60, 241–276.

Myerson, R. B. (1983): "Mechanism Design by an Informed Principal," *Econometrica*, 51, 1767–1797.

Umbhauer, G. (1994): "Forward Induction, Consistency, Preplay Communication and Epsilon Perturbations," Mimeo Beta, Univesité Louis Pasteur, Strasbourg, France.