AUSTRALIAN
SCHOOL OF BUSINESS ™
THE UNIVERSITY OF NEW SOUTH WALES

www.business.unsw.edu.au

School of Economics
Australian School of Business
UNSW Sydney NSW 2052 Australia

http://www.economics.unsw.edu.au

Bayesian and Consistent Assessments

Carlos Pimienta

School of Economics Discussion Paper: 2009/13

The views expressed in this paper are those of the authors and do not necessarily reflect those of the School of Economic at UNSW.

BAYESIAN AND CONSISTENT ASSESSMENTS[*]

CARLOS PIMIENTA[†]

ABSTRACT. In a *Bayesian assessment* beliefs are computed from the strategy profile following Bayes' rule at positive probability information sets and for every subgame. We characterize the set of extensive-forms (extensive-form games without a payoff assignment) for which the sets of Bayesian assessments and consistent assessments coincide. In doing so we disentangle the different restrictions imposed by consistency across information sets.

1. INTRODUCTION

A sequential equilibrium (Kreps and Wilson, 1982) is a sequentially rational *consistent assessment*. The notion of consistency incorporated in the definition of sequential equilibrium provides a way of selecting beliefs at zero probability information sets. Loosely speaking, consistent beliefs must admit an explanation consisting of "small trembles" made to reach those information sets.

There is a broad theoretical literature dealing with sequential equilibrium. This partly stems from the apparently ad-hoc procedure whereby consistency selects beliefs, which urged an effort to understand better the notion of consistency and its game theoretical implications. Swinkels (1993), Battigalli (1996) and Kohlberg and Reny (1997) show that consistency is related to the game theoretical principle of strategic indepencence. If different players choose their strategies independently then their assessments must be consistent.[1]

[1] In fact they show that consistency is equivalent to a "strong" version of independence. See Swinkels (1993) and Kohlberg and Reny (1997) for different interpretations of this result.
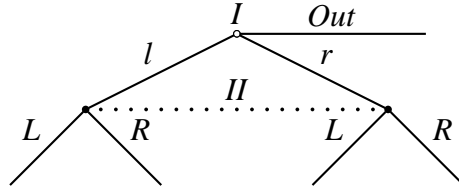
FIGURE 1.

A number of papers also offer different characterizations of consistency and/or show that, under certain conditions, sequential equilibrium is equivalent to weaker equilibrium concepts. Fudenberg and Tirole (1991) define perfect Bayesian equilibrium imposing some intuitive restrictions on beliefs and show its equivalence to sequential equilibrium in multi-period games with observable actions. Perea y Monsuwé, Jansen, and Peters (1997) provide an algebraic characterization of consistency without making use of trembles. Litan and Pimienta (2008) find the maximal class of extensive-forms such that sequential equilibrium and subgame perfection coincide in equilibrium strategies and equilibrium outcomes.

In this paper we look at those instances where consistency places no restrictions at zero probability information sets. For simplicity, we restrict ourselves to finite extensive-form games without proper subgames. This is motivated by the observation that the set of consistent assessments of a subgame coincides with the projection of the set of consistent assessments for the entire game on those coordinates corresponding to the subgame; and that every consistent assessment of the subgame of is part of some consistent assessment for the entire game. Apart from this restriction, we do not impose any further structure on the extensive-form games other than perfect recall.

To introduce the reader to the nature of the results and the characterizations that we shall derive consider the *extensive-form* of Figure 1. We refer to it as extensive-form and not as extensive-form game because it lacks a payoff assignment. (Hence, in the sequel, whenever payoffs are not yet specified we talk about extensive-forms and subforms instead of extensive-form games and subgames.) If Player *I* moves *Out* then any belief at Player *II*'s information set is consistent as it can be justified by an appropriate sequence of trembles. A similar argument holds in the extensive-form of Figure 2. If players *I* and *II* play according to (r_1, r_2) then arbitrary beliefs at Player *III*'s information set are consistent.

To formalize these ideas, we define Bayesian assessments by imposing the only requirement that beliefs at positive probability information sets are computed from the strategy profile using Bayes' rule. Clearly, every consistent assessment is a Bayesian assessment and in general, not every Bayesian assessment is consistent. Our objective is to characterize the set of extensive-forms without proper subforms such that every Bayesian assessments is consistent. Consider again the extensive-forms of Figure 1 and

Figure 2. In Figure 1 (Figure 2) whenever Player *II*'s information set (Player *III*'s information set) is reached with positive probability, Bayes' rule fully determines beliefs at that information set; and whenever that information set is reached with zero probability, arbitrary beliefs are guaranteed to be consistent.

It is not difficult to come up with examples of extensive-forms for which some Bayesian assessment is not consistent. This is the case of the extensive-form in Figure 3 and the Bayesian assessment $(Out, l_2, R, \mu(x_2) = 1)$. The reason is that consistent beliefs should place probability zero at the central node of Player *III*'s information set given that in a sequential equilibrium "correlation in defections are (partially) ruled out" (Kreps and Wilson, 1982, p. 875). That is, if Player *I* defects, it does not make a defection of Player *II* more likely.

Figure 5 contains another example. Kreps and Wilson (1982, p. 876) explain how the consistency criterion invokes the "common knowledge" principle for beliefs. Hence, any assessment where Player *I* moves *Out* and players *II* and *III* assess different relative probabilities over their left-hand and right-hand nodes is not consistent. The current work identifies the relevant characteristic shared by the extensive-forms in figures 3, 5, and any other extensive form where consistency selects a strict subset of Bayesian assessments. In identifying this characteristic we will also disentangle the different restrictions imposed by consistency across information sets.

There are practical reasons why we think that this type of result is worth exploring. Mainly, the characterizations derived here can be useful when it comes to economic applications. They are based on extensive-forms and quite easy to verify. Furthermore, they can be interpreted as a delineation of the cases where arbitrary off-equilibrium beliefs are guaranteed to be consistent. In the last section we show some modification of the the results that serve this purpose.

Additionally, this paper can help understand better how consistency brings about restriction in beliefs. While in some cases we may already have a very good understanding about how consistent beliefs are shaped, as it happens for instance when one information set comes after another like in Figure 5, in some other cases this relation may be more obscure or, at least, difficult to identify by arguments that are not context specific (see Figure 6).
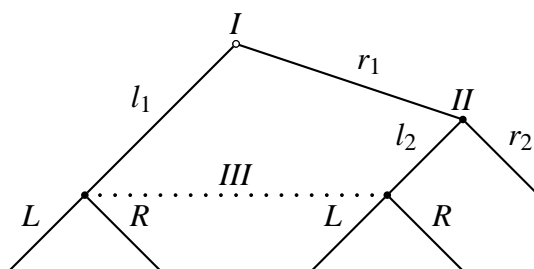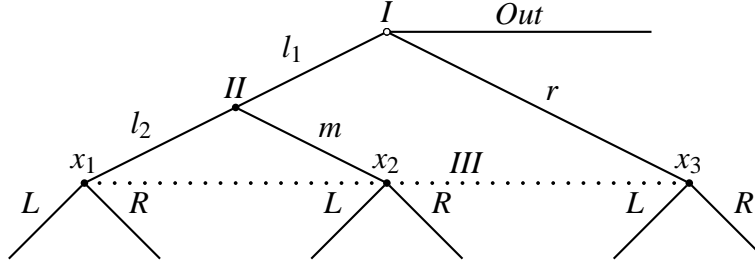

FIGURE 2.

FIGURE 3.

For this reason, a unifying explanation of the restrictions on beliefs entailed by consistency that is based solely on the characteristics of extensive-forms can be of theoretical interest.

In the next section we introduce the basic notation of extensive-form games and important definitions. Section 3 contains the results illustrated by a series of examples. Proofs are offered in Section 4. Section 5 is concerned with the relationship between sequentially rational Bayesian assessments and sequential equilibria. Section 6 concludes and shows some practical implications of the results.

2. BASIC NOTATION AND DEFINITIONS

We start by describing some necessary notation and terminology for finite extensive-form games with perfect recall. We decompose an *extensive-form game* into its *extensive-form* and the payoff assignment. Our characterizations are stated in terms of extensive-forms. As we mentioned in the introduction we focus on extensive-forms without proper *subforms*.[2]

The set of *players* is $\mathcal{N} = \{0, 1, \ldots, N\}$. Player $0 \in \mathcal{N}$ corresponds to *Nature*. We index players other than Nature with the letter $n = 1, \ldots, N$.

The set of *nodes* in the tree is represented by X and the set of final nodes by Z. The collection of *information sets* of player n is H_n. An element $h \in H_n$ represents the set of nodes that player n cannot distinguish when she has to move at h. The information set that contains node x is denoted as $h(x)$. Furthermore, $H = \bigcup_n H_n$.

The set of *actions* available at the information set h is $A(h)$. We use the terms *action* and *choice* interchangeably. We denote as $A = \bigcup_h A(h)$ the complete set of actions across information sets. Therefore, the $A(h)$'s partition A. If player n chooses action $a \in A(h)$, $h \in H_n$, when at node $x \in h$, the next node being reached is denoted by $\tau(x, a)$.

---

[2] The terms "extensive-form" and "subform" are taken from Kreps and Wilson (1982). Kreps and Wilson, however, obtain an extensive-form game from an extensive-form by specifying an assignment of payoffs to ending nodes together with a strictly positive probability measure over the set of initial nodes (i.e. Nature's initial move). We incorporate the probabilities associated to Nature's moves into the extensive-form.

Given any node, there is a unique sequence of choices that from the root of the extensive-form lead to that node. That sequence of choices is called *path to node x* and it is denoted by $\mathscr{P}(x)$.

An extensive-form game is obtained from an extensive-form by specifying for each player n a *Bernoullian utility function $u_n : Z \to \mathbb{R}$*. Once we fix an extensive-form, an extensive-form game is given by N Bernoullian utility functions and it can be seen as a point in $\mathbb{R}^{N|Z|}$.

The set of player n's *pure strategies S_n* is the set of all functions $s_n : H_n \to A$ such that $s_n(h) \in A(h)$ for all information sets h in H_n. The set of pure strategy profiles is $S = S_1 \times \cdots \times S_N$. We write $S(x)$ and $S(h)$ to denote the set of pure strategy profiles that reach, respectively, node x and information set h. The sets $S_n(x)$ and $S_{-n}(x)$ are the projections of $S(x)$ on S_n and $S_{-n} = \prod_{m \neq n} S_m$.

A *behavioral strategy profile σ* is a sequence of functions $\sigma(\cdot \mid h) : A(h) \to [0,1]$, one for each $h \in H$, satisfying $\sum_{a \in A(h)} \sigma(a \mid h) = 1$ for all h. In turn, a *system of beliefs μ* is a sequence of functions $\mu(\cdot \mid h) : h \to [0,1]$, one for each $h \in H$, satisfying $\sum_{x \in h} \mu(x \mid h) = 1$ for all h. An *assessment* is a strategy profile together with a system of beliefs (σ, μ).

We now introduce our two objects of study.

Definition 1 (Consistent Assessments). The assessment (σ, μ) is *consistent* if it is the limit point of a sequence $\{(\sigma^t, \mu^t)\}_{t=0}^{\infty}$ such that, for all t, σ^t is completely mixed (i.e. $\sigma^t(a \mid h) > 0$ for all $h \in H$ and all $a \in A(h)$) and μ^t is derived from σ^t using Bayes' rule.

Definition 2 (Bayesian Assessments). The assessment (σ, μ) is a *Bayesian assessment* if if the value of μ at information sets reached with positive probability is computed from σ using Bayes' rule.

Of course, every consistent assessment is a Bayesian assessment.

3. NON-CONSISTENT BAYESIAN ASSESSMENTS

In this section we characterize the set of extensive forms such that the set of consistent assessments is a strict subset of the set of Bayesian assessments. This is done in propositions 1 and 3. Theorem 1 will later assert that in the complement of the set laid out by the propositions the sets of Bayesian and consistent assessments coincide.

To provide a more clear intuition about the results let us introduce relative probabilities over the set S of pure strategy profiles.[3] A relative probability on S specifies the relative weight of each subset of pure strategy profiles

---

[3] Relative probabilities are equivalent to the notion of *conditional probability systems* (Myerson, 1986). In game theory conditional probability systems arise naturally from the need of specifying probabilities conditional on events that have prior probability zero. Among others, conditional probability systems have been studied by Myerson (1986); McLennan (1989a,b); Blume, Brandenburger, and Dekel (1991); Swinkels (1993); Hammond (1994); Battigalli (1996); and Kohlberg and Reny (1997)

with respect to any other subset. This includes subsets having prior probability equal to zero. A relative probability ρ on S must satisfy the following: for every subset $Q \subset S$ and all nonempty subsets $R, T \subset S$,

(i) $\rho(Q,R) \in [0,\infty]$,
(ii) $\rho(Q,Q) = 1$,
(iii) $\rho(Q,T) + \rho(R,T) = \rho(Q \cup R, T)$ if $Q \cap R = \varnothing$, and
(iv) $\rho(Q,T) = \rho(Q,S)\rho(S,T)$, whenever the product does not involve both 0 and ∞.

Standard prior probabilities are therefore given by $\rho(\cdot, S)$.

Battigalli (1996) and Kohlberg and Reny (1997) show that every consistent assessment can be generated, in a way specified below, by a relative probability defined over the set of pure strategy profiles and satisfying a strong independence property. Strong independence implies weak independence and, for our purposes, the latter concept is restrictive enough. The relative probability ρ defined over S is weakly independent if for every two nonempty Cartesian subsets $Q, R \subset S$ and every player $n \in \mathcal{N}$,[4]

$$\rho(Q_n \times Q_{-n}, R_n \times Q_{-n}) = \rho(Q_n \times R_{-n}, R_n \times R_{-n}).$$

Every consistent assessment (σ, μ) can be generated by a relative probability ρ satisfying weak independence according to:[5]

(3.1) $\qquad \sigma(a \mid h) = \rho(S(\tau(x,a)), S(h))$ for any $x \in h$; $and,$

(3.2) $\qquad \mu(x \mid h) = \rho(S(x), S(h))$.

It can be shown that perfect recall and weak independence imply that (3.1) is well defined (i.e. it does not depend on the choice of $x \in h$).

We are going to derive a condition that implies restrictions on consistent beliefs at zero probability information sets. Recall that the path to a node x, denoted by $\mathscr{P}(x)$, is the collection of actions that from the root of the extensive-form lead to that node. Consider two nodes x and y that have a common action $\bar{a} \in \mathscr{P}(x) \cap \mathscr{P}(y)$ in their respective paths. Let σ be a pure behavioral strategy profile that takes action $\bar{a}$ with probability zero and every other action in the path to x with probability one. Even though under σ both nodes x and y occur with probability zero, intuitively, node y cannot be *infinitely more likely* than node x. Let us now offer a more formal argument.

Let ρ an independent relative probability defined on S that induces the consistent assessment (σ, μ) where σ is as above. We denote s^σ the pure

---

[4] Swinkels (1993) calls this condition *individual quasi-independence*. It is the (weak) notion of independence considered by Kohlberg and Reny (1997). He uses the term *quasi-independence* for the analogous condition when more than one coordinate is allowed to change. This is the independence condition used by Battigalli (1996). Quasi-independence implies individual quasi-independence but the opposite is not true. Swinkels (1993) offers an example, credited to Myerson, at that effect.

[5] The converse is not true, i.e. not every weakly independent relative probability system generates a consistent assessment. See Kohlberg and Reny (1997) for an example.

strategy profile which is equivalent to σ. Let Player n be the player that moves at the information set $\bar{h}$ where the choice $\bar{a}$ is available. Owing to independence, the relative probability $\rho(S(x), S(y))$ must be equal to $\rho(S(x)_{-n} \times \{s_n^\sigma\}, S(y)_{-n} \times \{s_n^\sigma\})$. Moreover, $\rho(S(x)_{-n} \times \{s_n^\sigma\}, S) > 0$ because one element in $S(x)_{-n} \times \{s_n^\sigma\}$, the pure strategy profile s^σ, receives strictly positive prior probability. Therefore, $\rho(S(x)_{-n} \times \{s_n^\sigma\}, S(y)_{-n} \times \{s_n^\sigma\}) > 0$ and consequently $\rho(S(x), S(y)) > 0$. That is, node y is not infinitely more likely than node x. Now, if x and y belong to the same information set h then $S(x)$ and $S(y)$ are subsets of $S(h)$ and we obtain $\rho(S(y), S(h)) < 1$. Furthermore, if the information set h is reached with probability zero under σ the last inequality and (3.2) imply $\mu(y \mid h) < 1$.

Consider again the extensive-form of the game in Figure 3. The leftmost node and the central node in Player III's information set have a common choice, action l_1, in their respective paths. If players I and II play according to (Out, l_2) then (Out, l_2) is infinitely more likely than (Out, m). We can use independence to conclude that the profile (l_1, l_2) is infinitely more likely than (l_1, m), i.e. the leftmost node in Player III's information set is infinitely more likely than the central node. Since Player I plays Out, Player III's information set is reached with probability zero and we need to specify beliefs at her information set. It follows that consistent beliefs must assign probability zero to that central node. This restriction does not apply to Bayesian assessments.

We start our characterization with a sufficient condition for an extensive-form to have Bayesian assessments that are not consistent. It corresponds to the set of properties suggested above and in our analysis of Figure 3.

Proposition 1. *Consider an extensive-form without proper subgames. Suppose that we can find an information set h with two distinct nodes x and y and a pure behavioral strategy profile σ such that the following conditions hold:*

(i) the information set h is reached with probability zero under σ; and
(ii) there exists one action $\bar{a} \in \mathscr{P}(x) \cap \mathscr{P}(y)$ that satisfies $\sigma(\bar{a} \mid \bar{h}) = 0$.

Then the set of consistent assessments is strictly contained in the set of Bayesian assessments.

Property (i) above is clear. The sets of Bayesian and consistent assessments may differ only if some information set receives probability zero. It also shows that relative probabilities are only part of the story. Not only do we need a behavioral strategy profile that reveals a likelihood ordering between two nodes, but also one that at the same time reaches the information set containing these two nodes with probability zero.

Property (ii) captures the relevant features in Figure 3. To understand better why we need an action $\bar{a} \in \mathscr{P}(x) \cap \mathscr{P}(y)$ observe that in figures 1 and 2, where every Bayesian assessment is consistent, we cannot find two nodes in the same information set that have a common action in their respective paths. To see why we need $\sigma(\bar{a} \mid \bar{h}) = 0$, consider the extensive-form of
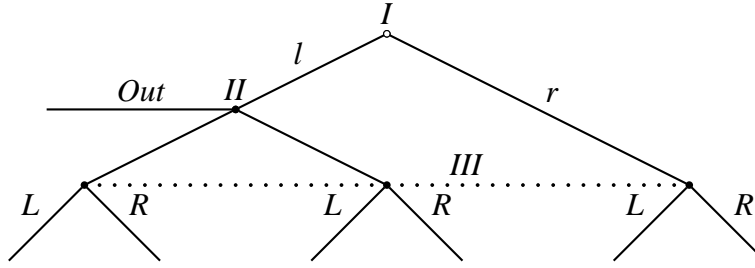
FIGURE 4.

Figure 4 and a behavioral strategy profile that assigns probability one to (l, Out). There is no restriction on how Player *III* should form her consistent beliefs. (Every conceivable belief vector at that information set is the limit of a sequence of conditional probabilities generated by an appropriately chosen sequence of trembles.) In this case, for any system of beliefs μ, the resulting assessment can be associated to a well defined independent relative probability system on the set of pure strategy profiles—and every Bayesian assessment is consistent.

It is worthwhile pointing out that it is enough to find one behavioral strategy profile that, together with the extensive-form under consideration, satisfies all three properties in Proposition 1 for us to know that there are non-consistent Bayesian assessments. However, nothing guarantees that every strategy profile satisfying these three properties is part of one of such assessments. To see this we can combine the extensive-forms in figures 3 and 4 so that both players *I* and *II* can move out. A strategy profile where, indeed, both Players *I* and Player *II* move out satisfies the three properties. However, arbitrary beliefs at player *III*'s information set are consistent.

Figure 5 is another example where the set of consistent assessment is a strict subset of the set of Bayesian assessments. Consistency implies common knowledge of beliefs. This means that Player *II* and Player *III* must have the same belief over their left-hand and right-hand nodes and that, consequently, not every Bayesian assessment is consistent. In order to see this in terms of Proposition 1 note that action r_2 belongs to the path of the two nodes in Player *III*'s information set.

The strategy profile (Out, l_2, X) is one of the strategies that Proposition 1 demands. But the behavioral strategy profile (Out, r_2, L) does not satisfy property (ii) above and yet it does not admit arbitrary Bayesian beliefs to form consistent assessments. This suggests that we should explore further how the values assumed by consistent beliefs at two different information sets relate to each other. We start this analysis studying signaling games, where it is well known that sequential equilibrium does not impose restrictions on beliefs.[6] Once Nature moves, the sender observes her type and sends a signal to the receiver. If the receiver observes a signal which is sent in equilibrium, she applies Bayes' rule to derive her beliefs about the

---

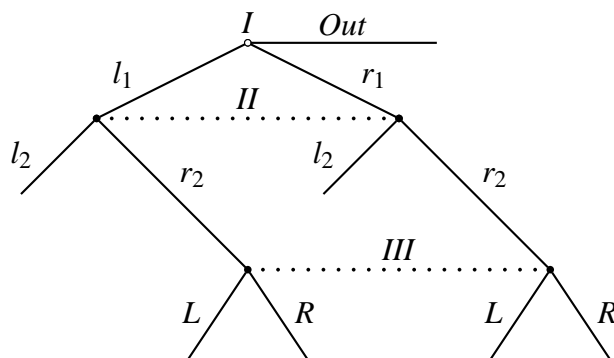[6] See, for instance, Kohlberg (1990).

FIGURE 5.

type of the sender. If a signal is not sent in equilibrium, the receiver has no restrictions whatsoever to form her beliefs about the type of the sender upon receiving that signal. Note well that every pair of nodes that belong to the same information set have completely different paths from each other because "identical" signals can come from different types.

Consider the extensive-form of a slightly modified signaling game in Figure 6.[7] After the sender learns her type, and before she sends a signal, she can end the game. As it is the case in a standard signaling game, no pair of nodes that belong to the same information set have a common action in their respective paths. Nonetheless, there are actions that are common to the paths to nodes that belong to different information sets. That is, $C_1 \in \mathscr{P}(x_1) \cap \mathscr{P}(y_2)$ and $C_2 \in \mathscr{P}(x_2) \cap \mathscr{P}(y_1)$. Consider the behavioral strategy profile $(F_1, F_2, U_1, D_2, R_U, L_2)$ which assigns probability zero to C_1 and C_2. As mentioned previously, if a strategy profile assigns positive probability to all the actions leading to a node but one, which is also in the path to a second node, then the underlying independent relative probability must consider the (set of pure strategy profiles leading to the) first node as infinitely more likely than the second node. In terms of the example this means $\rho(S(x_1), S(y_2)) = \rho(S(x_2), S(y_1)) = \infty$ which in turn implies that we cannot have $\rho(S(y_1), S(x_1)) = \rho(S(y_2), S(x_2)) = \infty$ because otherwise a node must be infinitely more likely than itself. From this argument we obtain that a consistent assessment with a behavioral strategy profile as above cannot display beliefs such that $\mu(y_1 \mid h_1) = \mu(y_2 \mid h_2) = 1$. This restriction does not apply to Bayesian assessments.

The relevant features of the previous example are generalized in the next proposition into a new sufficient condition on extensive-forms so that the set of consistent assessments is a strict subset of the set of Bayesian assessments.

Proposition 2. *Consider an extensive-form without proper subforms. Suppose that we can find two distinct information sets h_1, h_2, four distinct nodes*

---

[7] The extensive-form of this game is a simplified version of the one in Figure 235.1 in Osborne and Rubinstein (1994).
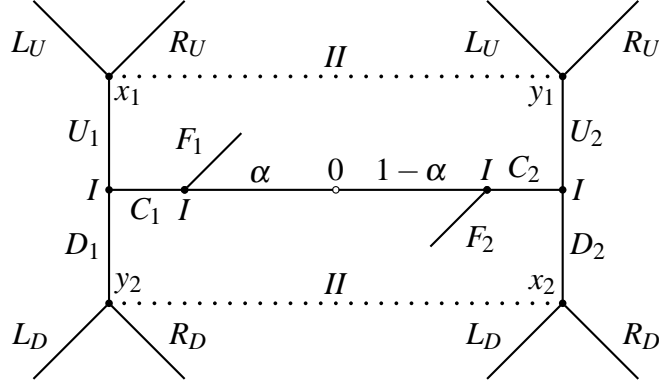
FIGURE 6.

x_1, y_1, x_2, y_2 and a pure behavioral strategy profile σ such that the following conditions hold:

(i) x_1, $y_1 \in h_1$ and x_2, $y_2 \in h_2$;

(ii) the information sets h_1 and h_2 are reached with probability zero under σ; and

(iii) we can find actions $\bar{a}_1 \in \mathscr{P}(x_1) \cap \mathscr{P}(y_2)$ and $\bar{a}_2 \in \mathscr{P}(x_2) \cap \mathscr{P}(y_1)$ that satisfy $\sigma(\bar{a}_1 \mid \bar{h}_1) = \sigma(\bar{a}_2 \mid \bar{h}_2) = 0$.

Then the set of consistent assessments is strictly contained in the set of Bayesian assessments.

Remark 1.   (i) If $h_1 = h_2$ then we can apply Proposition 1. However, it is important that $x_1 \neq y_1$ and $x_2 \neq y_2$. In particular, $\bar{a}_1 \in \mathscr{P}(x_1) \cap \mathscr{P}(x_2)$ and $\bar{a}_2 \in \mathscr{P}(x_2) \cap \mathscr{P}(y_1)$ together with $\sigma(\bar{a}_1 \mid \bar{h}_1) = \sigma(\bar{a}_2 \mid \bar{h}_2) = 0$ do not guarantee that some Bayesian assessments is not consistent. Note as well that the last two inclusions do not imply $\mathscr{P}(x_1) \cap \mathscr{P}(y_1) \neq \varnothing$, hence, we cannot apply Proposition 1 in this case.

(ii) Having two actions $\bar{a}_1$ and $\bar{a}_2$ is also necessary. Suppose that in the extensive-form of Figure 6 we delete action C_2 and replace the two consecutive information sets of Player *I* by a single information set where Player *I* has available the three choices F_2, U_2 and D_2. With this modification every Bayesian assessment is consistent.

In our signaling game the two zero probability information do not come one after the other as it occurs, for instance, in the extensive-form of Figure 5. We have seen that this extensive-form satisfies the conditions of Proposition 1, therefore, we already know that some Bayesian assessments are not consistent. We can now see that this extensive-form also satisfies the conditions of Proposition 2. Indeed, if Player *I* moves *Out*, then Player *II*'s information set and Player *III*'s information set are reached with probability zero. The left-hand node in *II*'s information set and the left-hand node
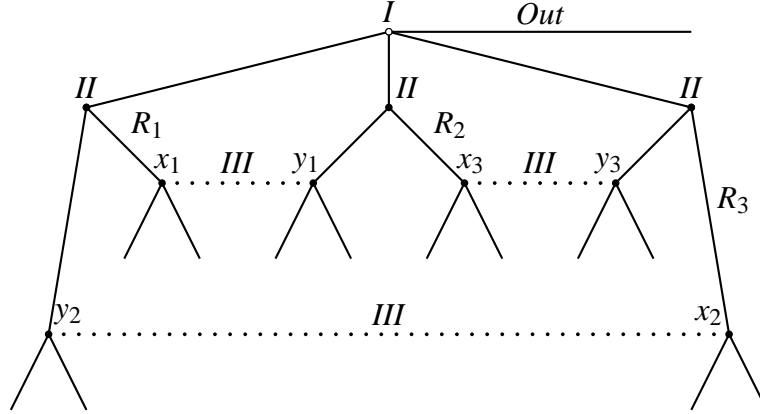
FIGURE 7.

in *III*'s information set have action l_2 in their respective paths. The analogous is true for the right-hand nodes and action r_2. Note, nevertheless, that if moving *Out* was not a possible action the resulting extensive-form would satisfy the conditions of Proposition 1, but not the conditions of Proposition 2.

Similar arguments to those in our previous two examples are also valid when three or more information sets are involved. Figure 7 illustrates this with three information sets. Consider that players *I* and *II* play according to (Out, R_1, R_2, R_3). We must specify beliefs at the three information sets of Player *III*. The following pairs of nodes have an action in their respective paths that the previous profile attaches probability zero: (x_1, y_2), (x_2, y_3) and (x_3, y_1). Weak independence implies that, for each of these pairs, the pure strategy profile leading to the first node is infinitely more likely than the pure strategy profile leading to the second node. It follows that a system of beliefs such that $\mu(y_1 \mid h_1) = \mu(y_2 \mid h_2) = \mu(y_3 \mid h_3) = 1$ is not consistent. Again, this restriction does not apply to Bayesian assessments.

The extensive-form of Figure 7 is such that no two information sets receiving probability zero come one after another. Figure 8 contains an extensive-form where for some information sets this is the case. For reasons analogous to those discussed in the previous examples, the assessment $(Out, L_1, L_2, \mu(y_1 \mid h_1) = 1, \mu(y_2 \mid h_2) = 1, \mu(y_3 \mid h_3) = 1, \mu(y_4 \mid h_4) = 1)$ is not consistent.[8]

The general result, which subsumes Proposition 2, is the following:[9]

Proposition 3. *Consider an extensive-form without proper subforms. Suppose that we can find $K \geq 2$ distinct information sets $h_1, \ldots, h_K$, $2K$ distinct*

---

[8] One small difference is that in this example $S(y_1)$ and $S(x_2)$ are of the same "order of magnitude", i.e. neither set is infinitely more likely than the other because L_1 is chosen by the strategy profile with probability one. The same is true for $S(y_3)$ and $S(x_4)$.

[9] Nonetheless, we present propositions 2 and 3 separately to facilitate the exposition of the results. On a technical note, the proof of the Proposition 3 is done by induction (see page 18), and Proposition 2 corresponds to the initial case.
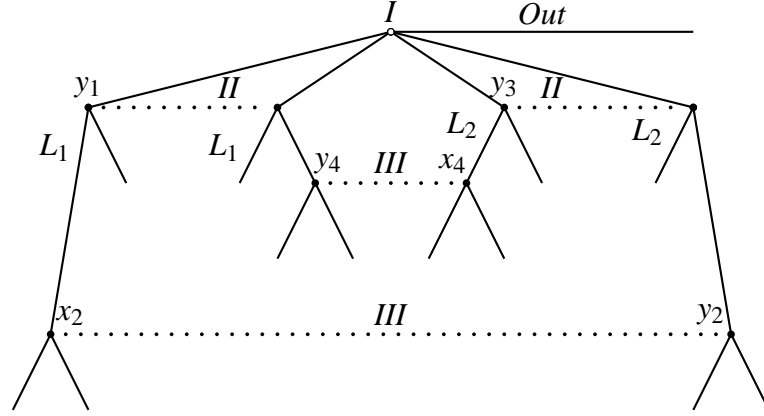
FIGURE 8.

nodes $x_1, y_1, \ldots, x_K, y_K$ and a pure behavioral strategy profile σ such that the following conditions hold:

  (i) *for each $i = 1, \ldots, K$ nodes x_i and y_i belong to h_i;*
 (ii) *the information sets $h_1, \ldots, h_k$ are reached with probability zero under σ; and*
(iii) *for each $i = 1, \ldots, K - 1$ we can find and action $\bar{a}_i \in \mathscr{P}(x_i) \cap \mathscr{P}(y_{i+1})$ that satisfies $\sigma(\bar{a}_i \mid \bar{h}_i) = 0$, likewise, there is an action $\bar{a}_K \in \mathscr{P}(x_K) \cap \mathscr{P}(y_1)$ such that $\sigma(\bar{a}_K \mid \bar{h}_K) = 0$.*

Then the set of consistent assessments is strictly contained in the set of Bayesian assessments.

Remark 2.   (i) Proposition 2 corresponds to $K = 2$.
  (ii) It is embedded in the statement of the proposition that we have to find K information sets *and an order* of those information sets such that the condition are true. The conditions do not need to hold for every possible order.
 (iii) Again, it is important that the $2K$ nodes be distinct. However, if some of the K information sets are not different it would only mean that the conditions are satisfied for an integer strictly smaller than K.
 (iv) If the conditions are satisfied for some value of K it does not follow that they are also satisfied for some integer smaller than K. See for instance, Figure 7 where $K = 3$ and Figure 8 where $K = 4$. In both cases the conditions do not hold for smaller values of K.

   Theorem 1 states that the conditions in propositions 1 and 3 are not only sufficient but also necessary.

Theorem 1. *Consider an extensive-form without proper subforms where some Bayesian assessments is not consistent. Then the extensive-form satisfies either the conditions listed in Proposition 1 or the conditions listed in Proposition 3.*

Intuitively, for a fixed behavioral strategy profile, if a zero probability choice is in the path to two different nodes then we are losing freedom to choose consistent beliefs because the same tremble is associated to two different nodes. The fact that the conditions in propositions 1 and 3 are not satisfied means that we always have enough freedom so that arbitrary beliefs are consistent.

Take a behavioral strategy profile σ and two nodes x and y. Suppose that there is at least one choice in the path to x that is not in the path to y and at least one choice in the path to y that is not in the path to x. Suppose further that none of these two choices is taken with positive probability under σ. Based only in this information about σ and the structure of the extensive-form we cannot derive a definite likelihood ordering between nodes x and y. Recurring to an argument based on trembles, we can fine-tune the trembles associated to the two actions pinpointed before to make one of the sequences of trembles necessary to reach one of the nodes as likely as we want with respect to the other. If x and y are the only nodes in the same information set h and this is reached with probability zero under σ then we can freely choose consistent beliefs at h.

If the conditions in Proposition 1 are not satisfied the above argument holds even if we have more than two nodes in h and for any σ. So consistent beliefs at h can be freely chosen. But this choice of beliefs could, in principle, constrain the set of consistent beliefs available for a second zero probability information set h'. This can happen when one of the choices whose tremble we fine-tuned before is in the path to some node in h'. If the conditions in Proposition 3 are not satisfied for $K = 2$ then for each of the remaining nodes in that information set we can again fine-tune some tremble associated to a choice that is only in the path to that node and none else in h'. The argument can be repeated again for a third zero probability information set h'' but if the conditions in Proposition 3 are not satisfied for $K = 3$ we do not have restrictions in choosing beliefs at h''. Continuing in this fashion we can pick arbitrary beliefs at every zero probability information set.

4. PROOFS

We begin by describing some additional notation.

Given a behavioral strategy profile σ, its *carrier* $\mathscr{C}(\sigma)$ is the union over all information sets h of the choices $a \in A$ that satisfy $\sigma(a \mid h) > 0$. Furthermore, a subset of choices $C \subset A$ is said to be a carrier if for each information set h it contains at least one choice in $A(h)$. The set of all carriers is $\mathscr{C}(\Sigma)$. Every carrier $C \in \mathscr{C}(\Sigma)$ has associated a set of decision nodes that are reached with positive probability when any strategy σ with carrier $\mathscr{C}(\sigma) = C$ is played. We denote that set as $X^+(C)$ and its complement as $X^0(C)$.

Given a system of beliefs μ, its *support* supp(μ) is the union over all information sets h of the decision nodes x for which $\mu(x \mid h) > 0$. We reserve

the term *carrier* to talk about the set of choices that receive strictly positive probability for probabilities defined by strategy profiles, and the term *support* for the analogous concept for probabilities associated to beliefs.

The main tool used in the proofs of Propositions 1, 2 and 3 is Lemma A1 in Kreps and Wilson (1982). Before stating that lemma, some concepts are necessary.

A *labelling* is a function $L : A \to \mathbb{N}$ that maps each choice a into an integer number $L(a)$. For each labelling L there is an associated function $F_L : X \to \mathbb{N}$ defined by

$$F_L(x) = \sum_{a \in \mathscr{P}(x)} L(a).$$

Definition 3. Given a carrier C, the labelling L is said to be a C-labelling if we have $a \in C$ if and only if $L(a) = 0$.

A *basis* (C, Y) is a subset of $A \times X$. We say that the basis (C, Y) is consistent if there exists at least one consistent assessment (σ, μ) such that $\mathscr{C}(\sigma) = C$ and $\mathrm{supp}(\mu) = Y$.

Lemma 1 (Kreps and Wilson (1982, Lemma A1))**.** *The basis (C, Y) is consistent if and only if there is a C-labelling L such that following condition holds:*

$$x \in Y \text{ if and only if } x \text{ minimizes } F_L(\cdot) \text{ on } h(x).$$

For our purposes Lemma 1 implies the following. Take a behavioral strategy profile σ that assigns probability zero to the information set h and suppose that nodes x and y belong to h. If $F_L(x) \le F_L(y)$ *for every* conceivable $\mathscr{C}(\sigma)$-labelling L then a necessary condition for (σ, μ) to be consistent is that $\mu(y \mid h) \ne 1$. In order to prove Proposition 1 we show that, if the extensive-form meets the conditions given in the proposition, we can always find such a strategy profile or, more precisely, such a carrier.

For a behavioral strategy profile σ with carrier C, we are thus interested in inequalities of the form $F_L(x) \le F_L(y)$ that remain true for every C-labelling L. It will be useful to write $F(x, C) \le F(y, C)$ when this is the case. For instance, if node y comes after x we can readily conclude that $F(x, C) \le F(y, C)$. The expression $F(x, C) \le F(y, C)$ can be meaningfully read as "node y cannot be infinitely more likely than node x under any strategy profile with carrier C". If x and y belong to the same zero probability information set this must be respected by consistent beliefs.

To avoid duplication of arguments we provide some results in the next lemma that are used continuously throughout the proofs. Throughout, if E and F are two arbitrary sets, we use $E \subset F$ allowing for equality of the two sets. As usual, the notation $E \setminus F$ represents the set of elements in E that do not belong to F.

Lemma 2. *Consider a carrier C, an information set $h \subset X^0(C)$, and two nodes x and y with a common action $a \in \big(\mathscr{P}(x) \cap \mathscr{P}(y)\big) \setminus C$. Let $C' = C \cup \big(\mathscr{P}(x) \setminus \{a\}\big)$, the following holds:*

(i) $0 < F(x, C') \le F(y, C')$;

(ii) if $h \not\subset X^0(C')$ then there is at least one node $\hat{y} \in h$ that satisfies $\mathscr{P}(y) \subset C \cup \mathscr{P}(x)$; and moreover

(iii) $\big(\mathscr{P}(x) \cap \mathscr{P}(\hat{y})\big) \setminus C \ne \varnothing$, i.e. $0 < F(\hat{y}, C) \le F(x, C)$.

Proof. Take a carrier C that does not contain any action a that satisfies $a \in \mathscr{P}(x) \cap \mathscr{P}(y)$. Suppose that under that carrier $h \in X^0(C)$. (Note that x and y belong to $X^0(C)$ because $a \notin C$ and that we no dot necessarily assume x, $y \in h$.) We obtain a new carrier C' by adding to C all the choices in the path to x except for a, therefore, we also obtain that x and y belong to $X^0(C')$.

To prove part (i) take any C'-labelling L. Action a is the only element in $\mathscr{P}(x)$ that is not in C'. Hence, $F_L(x) = L(a) > 0$ and since a is also in $\mathscr{P}(y)$ we obtain $F_L(y) \ge L(a)$. That is, node y cannot be infinitely more likely than x under a strategy with carrier C'.

Let us turn to part (ii). Suppose now that under C' some node in h, say $\hat{y}$, is reached with positive probability. As the only difference between C and C' is that C' contains actions in the path to x that are not included in C that means that x and $\hat{y}$ have common actions in their respective paths.

Part (iii) follows because every action in the path to $\hat{y}$ that is not in the path to x is in C. Since we also have $\hat{y} \in X^0(C)$ we obtain $0 < F(\hat{y}, C) \le F(x, C)$. That is, node x cannot be infinitely more likely than $\hat{y}$ under a behavioral strategy profile with carrier C'. $\square$

We can prove the first proposition.

Proof of Proposition 1. Recall that $h(x)$ represents the information set that contains node x. The conditions listed in the proposition are equivalent to the following set being nonempty.

$$\Phi_1 = \Big\{ (x, y, \bar{a}, C) \in X^2 \times A \times \mathscr{C}(\Sigma) : y \in h(x),\ y \ne x,\ h(x) \subset X^0(C),$$

$$\bar{a} \in \big(\mathscr{P}(x) \cap \mathscr{P}(y)\big) \setminus C \Big\}.$$

Take an arbitrary element $(x, y, \bar{a}, C) \in \Phi_1$ and construct the carrier

$$\hat{C} = C \cup \big(\mathscr{P}(x) \setminus \{\bar{a}\}\big).$$

Using Lemma 2 (i) we obtain that $F(x, \hat{C}) \le F(y, \hat{C})$. If $h(x) \subset X^0(\hat{C})$ then the desired result follows. If otherwise some $\hat{y} \in h(x)$ satisfies $\hat{y} \in X^+(\hat{C})$ then Lemma 2 (iii) implies $F(\hat{y}, C) \le F(x, C)$. Since $h(x) \subset X^0(C)$ this concludes the proof. $\square$

The argument behind the proof of Proposition 2 is similar but slightly more involved because we have to deal with nodes in two different information sets. Given two information sets h_1 and h_2, we want to use the structure of the extensive-form to find a carrier C and four nodes $x_1, y_1 \in h_1$, $x_2, y_2 \in h_2$ such that $F(x_1, C) \le F(y_2, C)$ and $F(x_2, C) \le F(y_1, C)$. If h_1 and h_2 are subsets of $X^0(C)$ then beliefs at those information sets need to be determined in every assessment (σ, μ) with $\mathscr{C}(\sigma) = C$. Consistency implies

that those beliefs cannot be chosen so that $\mu(y_1 \mid h_1) = \mu(y_2 \mid h_2) = 1$ because that would imply that for some C-labelling L it holds $F_L(y_1) < F_L(x_1)$ and $F_L(y_2) < F_L(x_2)$, but necessarily we also have $F_L(x_1) \leq F_L(y_2)$ and $F_L(x_2) \leq F_L(y_1)$. The main difficulty of the argument consists of showing that the information sets h_1 and h_2 are reached with probability zero. Lemma 2 will be of great help at this effect.

Proof of Proposition 2. Assume that $\Phi_1 = \varnothing$. The conditions in the proposition imply that the following set is nonempty:

$$\Phi_2 = \Big\{ (x_1, y_1, x_2, y_2, \bar{a}_1, \bar{a}_2, C) \in X^4 \times A^2 \times \mathscr{C}(\Sigma) : h(x_1) \neq h(x_2),$$

$$y_1 \in h(x_1),\ y_1 \neq x_1,\ y_2 \in h(x_2),\ y_2 \neq x_2,\ h(x_1) \cup h(x_2) \subset X^0(C),$$

$$\bar{a}_1 \in \big(\mathscr{P}(x_1) \cap \mathscr{P}(y_2) \big) \setminus C,\ \bar{a}_2 \in \big(\mathscr{P}(x_2) \cap \mathscr{P}(y_1) \big) \setminus C \Big\}.$$

Let $h_1 = h(x_1)$ and $h_2 = h(x_2)$. Take any $(x_1, y_1, x_2, y_2, \bar{a}_1, \bar{a}_2, C) \in \Phi_2$ and construct the carrier

$$\hat{C} = C \cup \big(\mathscr{P}(x_1) \setminus \{\bar{a}_1\} \big) \cup \big(\mathscr{P}(x_2) \setminus \{\bar{a}_2\} \big).$$

From Lemma 2 (i) we obtain $F(x_1, \hat{C}) \leq F(y_2, \hat{C})$ and $F(x_2, \hat{C}) \leq F(y_1, \hat{C})$. If both h_1 and h_2 are contained in $X^0(\hat{C})$ then no consistent assessment (σ, μ) with $\mathscr{C}(\sigma) = \hat{C}$ satisfies $\mu(y_1 \mid h_1) = \mu(y_2 \mid h_2) = 1$. So we need to prove the result for the cases where either $h_1 \not\subset X^0(\hat{C})$ or $h_2 \not\subset X^0(\hat{C})$.

Thus, let us assume that $h_2 \not\subset X^0(\hat{C})$ (the other case is analogous) and construct the carrier

$$\hat{C}_1 = C \cup \big(\mathscr{P}(x_1) \setminus \{\bar{a}_1\} \big).$$

Now we show that $h_1 \subset X^0(\hat{C}_1)$. Suppose to the contrary that $h_1 \not\subset X^0(\hat{C}_1)$. Lemma 2 implies that some node in h_1 (that is not x_1) and x_1 have a common actions in their respective paths that is not contained in C. But this would imply that the set Φ_1 is not empty as we assumed at the beginning. We can conclude $h_1 \subset X^0(\hat{C}_1)$.

We have $x_1 \in X^0(\hat{C}_1)$ because $\bar{a}_1$ does not belong to $\hat{C}_1$. If some node in h_1 does not belong to $X^0(\hat{C}_1)$ then this node must contain in its path a choice that is also in the path to x_1. This is because $h_1 \subset X^0(C)$ and the only difference between C and $\hat{C}_1$ is that the latter contains actions in the path to x_1 that are not contained in the former. However, we assumed that Φ_1 is empty so that we can conclude $h_1 \subset X^0(\hat{C}_1)$.

We also know that $\bar{a}_2 \notin \hat{C}_1$, for otherwise $\bar{a}_2$ would be a common choice between x_1 and y_1. Furthermore, $\bar{a}_1 \notin \hat{C}_1$ by construction. It follows that $x_2, y_2 \in X^0(\hat{C}_1)$. Since $h_2 \not\subset X^0(\hat{C})$ and $\Phi_1 = \varnothing$ we know that there is a subset of nodes $\hat{h}_2 \subset h_2$ that does not include x_2 nor y_2 and that satisfies $\hat{h}_2 \subset X^+(\hat{C}_1)$. From Lemma 2 (iii) we obtain $F(\hat{y}_2, C) \leq F(x_1, C)$ for each $\hat{y}_2 \in \hat{h}_2$.

Construct the new carrier

$$\hat{C}'_1 = C \cup \big(\mathscr{P}(y_1) \setminus \{\bar{a}_2\}\big).$$

By Lemma 2 (i), $F(y_1, \hat{C}'_1) \leq F(x_2, \hat{C}'_1)$. Using the same arguments as before we can show that $h_1 \subset X^0(\hat{C}'_1)$ and $x_2, y_2 \in X^0(\hat{C}'_1)$. If $h_2 \subset X^0(\hat{C}'_1)$ then the desired result follows because $F(\hat{y}_2, C) \leq F(x_1, C)$ keeps holding when we change C for $\hat{C}'_1$. That is, a consistent assessment does not satisfy $\mu(x_1 \mid h_1) = \mu(x_2 \mid h_2) = 1$.

Hence, the next case we need to explore is when a subset of nodes $\hat{h}'_2 \subset h_2$ exists such that $\hat{h}'_2 \subset X^+(\hat{C}'_1)$. We construct a new carrier using the set of choices:

$$\tilde{B} = \left\{ \mathscr{P}(x_1) \setminus \left(\bigcup_{\hat{y}_2 \in \hat{h}_2} \mathscr{P}(\hat{y}_2) \right) \right\} \cup \left\{ \mathscr{P}(y_1) \setminus \left(\bigcup_{\hat{x}_2 \in \hat{h}'_2} \mathscr{P}(\hat{x}_2) \right) \right\}.$$

Let the new carrier be:

$$\tilde{C} = C \cup \big(\tilde{B} \setminus \{\bar{a}_1, \bar{a}_2\}\big).$$

We still obtain $h_1 \subset X^0(\tilde{C})$ from Lemma 2 (ii) and $\Phi_1 = \varnothing$. We obtain $h_2 \subset X^0(\tilde{C})$ by construction. The last carrier that we consider is:

$$C^* = \begin{cases} \tilde{C} \cup \big(\mathscr{P}(x_2) \setminus \{\bar{a}_2\}\big) & \text{if } |\hat{h}_2| \leq |\hat{h}'_2|; \\ \tilde{C} \cup \big(\mathscr{P}(y_2) \setminus \{\bar{a}_1\}\big) & \text{if } |\hat{h}_2| > |\hat{h}'_2|. \end{cases}$$

In either case, the information set h_2 is included in $X^0(C^*)$ because Φ_1 is empty. Regarding h_1, if some node $\tilde{x}_1 \in h_1$ belongs to $X^+(C^*)$ then $F(\tilde{x}_1, \tilde{C}) \leq F(y_2, \tilde{C})$. Since we also have $F(\hat{x}_2, \tilde{C}) \leq F(y_1, \tilde{C})$ for every $\hat{x}_2 \in \hat{h}'_2$ the restriction on the values that consistent beliefs can take for strategy profiles with carrier equal to $\tilde{C}$ results. Therefore, we only need to analyze $h_1 \subset X^0(C^*)$ together with $|\hat{h}_2| \leq |\hat{h}'_2|$ (given that the other case is similar). Let L be an arbitrary C^*-labelling. The following holds:

$$S_L(x_1) = \textstyle\sum_{\hat{y}_2 \in \hat{h}_2} F_L(\hat{y}_2) + L(\bar{a}_1),$$
$$S_L(y_1) = \textstyle\sum_{\hat{x}_2 \in \hat{h}'_2} F_L(\hat{x}_2) + L(\bar{a}_2),$$
$$S_L(y_2) \geq L(\bar{a}_1),$$
$$S_L(x_2) = L(\bar{a}_2).$$

To understand better the first (and the second) equality notice that, by construction, the only actions in the path to x_1 that do not belong to C^* are $\bar{a}_1$ and those that we can also find in the path to some $\hat{y}_2 \in \hat{h}_2$ (although not necessarily all of them).

Take now a consistent assessment (σ, μ) with $\mathscr{C}(\sigma) = C^*$. If $\mu(y_2 \mid h_2) > 0$ then $L(\bar{a}_1) \leq L(\bar{a}_2)$. If moreover $\mu(\hat{y}_2 \mid h_2) > 0$ for all $\hat{y}_2 \in \hat{h}'_2$ then $F_L(x_1) \leq F_L(y_1)$ because $|\hat{h}_2| \leq |\hat{h}'_2|$. This completes the proof as $\mu(y_1 \mid h_1)$ cannot take a strictly positive value. $\qquad\square$

The proof of Proposition 3 is similar. Given K information sets $h_1, \ldots, h_K$ the plan is to find two nodes x_k and y_k for each information set h_k so that for some carrier C we have $F(x_1, C) \leq F(y_2, C), \ldots, F(x_K, C) \leq F(y_1, C)$. If every h_k is included in $X^0(C)$ then we cannot have $F_L(y_k) < F_L(x_k)$ for every k and some C-labelling L. This implies that if (σ, μ) is consistent and $\mathscr{C}(\sigma) = C$ then $\mu(x_k \mid h_k) \neq 1$ for at least one k.

Proof of Proposition 3. First, we assume that $\Phi_1 = \varnothing$. The conditions in the proposition imply that for some integer $K \geq 2$ the following set is nonempty:[10]

$$\Phi_K = \Big\{ (x_1, y_1, \ldots, x_K, y_K, \bar{a}_1, \ldots, \bar{a}_k, C) \in X^{2K} \times A^K \times \mathscr{C}(\Sigma) :$$
$$h(x_i) \neq h(x_j) \text{ for all } i \neq j, \text{ and for all } i = 1, \ldots K,$$
$$h(x_i) \subset X^0(C), \ y_i \in h(x_i), \ y_i \neq x_i, \ \bar{a}_i \in \big(\mathscr{P}(x_i) \cap \mathscr{P}(y_{i+1}) \big) \setminus C \Big\}.$$

We proceed by induction in K assuming that for any other integer $K' < K$ the set $\Phi_{K'}$ is empty. The case $K = 2$ was considered in Proposition 2.

Let $h_i = h(x_i)$. Take an arbitrary element of Φ_K and construct the carrier:

$$\hat{C} = C \cup \left(\bigcup_{i=1}^{K} \big(\mathscr{P}(x_i) \setminus \{\bar{a}_i\} \big) \right).$$

From Lemma 2 (i) we obtain $F(x_i, \hat{C}) \leq F(y_{i+1}, \hat{C})$ for every $i = 1, \ldots K$. If for every i we also obtain $h_i \subset X^0(\hat{C})$ then a consistent assessment (σ, μ) with $\mathscr{C}(\sigma) = \hat{C}$ cannot satisfy $\prod_{i=1}^{K} \mu(y_i \mid h_i) = 1$. Thus, we need to prove the result when, for some i, the set h_i is not included in $X^0(\hat{C})$. Let I represent the collection of indexes i such that $h_i \not\subset X^0(\hat{C})$.

Take an $i \in I$. Let $\hat{h}_i$ represent the subset of those $\hat{y}_i \in h_i$ that belong to $X^+(\hat{C})$. Nodes x_i and y_i do not belong to $\hat{h}_i$ because $\bar{a}_i$ and $\bar{a}_{i-1}$ are not in $\hat{C}$. Moreover, every action $a \in \mathscr{P}(\hat{y}_i) \setminus C$ is contained in $\mathscr{P}(x_{i-1})$ and, therefore, $F(\hat{y}_i, C) \leq F(x_{i-1}, C)$ for all $\hat{y} \in \hat{h}_i$. In particular no action $a \in \mathscr{P}(\hat{y}_i) \setminus C$ can be contained in some $\mathscr{P}(x_k)$ with $k \neq i - 1$. This would imply that $\Phi_{K'} \neq \varnothing$ for some integer $K' < K$.[11]

Construct the carrier:

$$\hat{C}' = C \cup \left(\bigcup_{k \notin I} \big(\mathscr{P}(y_k) \setminus \{\bar{a}_{k-1}\} \big) \right).$$

Lemma 2 (i) implies $F(y_k, \hat{C}') \leq F(x_{k-1}, \hat{C}')$ for every $k \notin I$. Furthermore, it also implies that we still have $F(\hat{y}_i, \hat{C}') \leq F(x_{i-1}, \hat{C}')$ for every $i \in I$ and

---

[10] Throughout the proof, when the index i equals K, the index $i+1$ refers to 1. Likewise, if $i = 1$, the index $i - 1$ refers to K.

[11] For instance, if action $a \in \mathscr{P}(\hat{y}_i) \setminus C$ belongs to $\mathscr{P}(x_{i-2})$ then Φ_{K-1} would be nonempty. One element of this set can be obtained from our initial choice from Φ_K by dropping the entries corresponding to the nodes at information set h_{i-1} and action a_{i-1}, and substituting node y_i by node $\hat{y}_i$ and action a_{i-2} by action a.

every $\hat{y}_i \in \hat{h}_i$. If all the information sets $h_1, \ldots, h_K$ are contained in $X^0(\hat{C}')$ then we obtain that no consistent assessment whose strategy has carrier $\hat{C}'$ can assign a belief equal to one to every decision node x_k. Therefore, we need to analyze what would happen otherwise. Let J represent the collection of indexes j such that $h_j \not\subset X^0(\hat{C}')$, furthermore, for each $j \in J$, let $\hat{h}'_j \subset h_j$ be the subset of nodes in h_j that belong to $X^+(\hat{C}')$. The assumption that $\Phi_{K'} = \varnothing$ whenever $K' < K$ indicates that if $j \in J$ then for every $\hat{x}_j \in \hat{h}'_j$ we obtain $\left(\mathscr{P}(\hat{x}_j) \setminus C\right) \subset \mathscr{P}(y_{j+1})$. With this in mind we use the set of choices

$$
\tilde{B} = \left\{ \bigcup_{i \in I} \left[\mathscr{P}(x_{i-1}) \setminus \left(\bigcup_{\hat{y}_i \in \hat{h}_i} \mathscr{P}(\hat{y}_i) \right) \right] \right\} \cup \left\{ \bigcup_{i \in I} \mathscr{P}(y_i) \right\} \cup
$$

$$
\left\{ \bigcup_{j \in J} \left[\mathscr{P}(y_{j+1}) \setminus \left(\bigcup_{\hat{x}_j \in \hat{h}'_j} \mathscr{P}(\hat{x}_j) \right) \right] \right\} \cup \left\{ \bigcup_{j \in J} \mathscr{P}(x_j) \right\}
$$

to construct the new carrier

(4.1) $$\tilde{C} = C \cup \left(\tilde{B} \setminus \{\bar{a}_1, \ldots, \bar{a}_K\} \right).$$

We can assume that every information set $h_1, \ldots, h_K$ is contained in $X^0(\tilde{C})$. To see why note that if $h_k \not\subset X^0(\tilde{C})$ then there must be an action $\tilde{a}$ in the second or in the fourth component of $\tilde{C}$ that is in the path of some node $\tilde{x}_k$ of h_k. Moreover, this node cannot be either x_k or y_k. Since $\Phi_{K'} = \varnothing$ for every $K' < K$ the action $\tilde{a}$ must be contained in either $\mathscr{P}(y_{k+1})$ or $\mathscr{P}(x_{k-1})$. Consider that $\tilde{a} \in \mathscr{P}(y_{k+1})$ then by Lemma 2 (iii) we have that $F(\tilde{x}_k, \hat{C}) \leq F(y_{k+1}, \hat{C})$ and we only need to replace x_k by $\tilde{x}_k$ and redefine I so that it does not include k. Analogously, suppose now that $\tilde{a} \in \mathscr{P}(x_{k-1})$. Lemma 2 (iii) implies $F(\tilde{x}_k, \hat{C}) \leq F(x_{k-1}, \hat{C})$. We now need to replace y_k by $\tilde{x}_k$ and remove the index k from J.

The last carrier that we consider is:[12]

$$
C^* = \begin{cases}
\tilde{C} \cup \left(\displaystyle\bigcup_{k \notin (I-1) \cup (J+1)} \left(\mathscr{P}(x_k) \setminus \{\bar{a}_k\} \right) \right) & \text{if } \dfrac{\prod_{i \in I}(|\hat{h}_i| + 1)}{\prod_{j \in J}(|\hat{h}'_j| + 1)} \geq 1, \\[3em]
\tilde{C} \cup \left(\displaystyle\bigcup_{k \notin (I-1) \cup (J+1)} \left(\mathscr{P}(y_k) \setminus \{\bar{a}_{k-1}\} \right) \right) & \text{otherwise.}
\end{cases}
$$

Suppose that $\prod_{i \in I}(|\hat{h}_i| + 1) \geq \prod_{j \in J}(|\hat{h}'_j| + 1)$. We now show that a consistent assessment (σ, μ) with $\mathscr{C}(\sigma) = C^*$ cannot satisfy at the same time all of the following:

   (i) for every $i \notin I$, $\mu(y_i \mid h_i) = 1$,

---

[12] Note the sets $(I-1) = \{i : i+1 \in I\}$ and $(J+1) = \{j : j-1 \in J\}$.

(ii) for every $i \in I$, $\mu(y_i \mid h_i) > 0$ and $\mu(\hat{y}_i \mid h_i) > 0$ for all $\hat{y}_i \in \hat{h}_i$,

(iii) for every $i \in I$, $\mu(y_i \mid h_i) + \sum_{\hat{y}_i \in \hat{h}_i} \mu(\hat{y}_i \mid h_i) = 1$.

If the consistent assessment (σ, μ) satisfies (i), (ii) and (iii) then we should be able to find a C^*-labelling, say L, as in Lemma 1. Given that the set $\tilde{C}$ defined in (4.1) is contained in C^* we can write $F_L(x_{i-1}) = \sum_{\hat{y}_i \in \hat{h}_i} F_L(\hat{y}_i) + F_L(y_i)$ for every $i \in I$. Additionally, (ii) and (iii) above imply that $S_L(x_{i-1}) = (|\hat{h}_i| + 1)S_L(y_i)$ for every $i \in I$. A similar argument shows that for every $j \in J$ the equality $S_L(x_j) = (|\hat{h}'_j| + 1)^{-1}S_L(y_{j+1})$ also holds. The definition of C^* for the case that we are considering entails $S_L(x_k) \leq S_L(y_{k+1})$ whenever $k \notin (I-1) \cup (J+1)$. Finally, $S_L(y_k) < S_L(x_k)$ for every $k = 1, \ldots, K$ given that we always have $\mu(x_k \mid h_k) = 0$ and $\mu(y_k \mid h_k) > 0$. We only have to put all these inequalities together to obtain:

$$\frac{\prod_{i \in I} (|\hat{h}_i| + 1)}{\prod_{j \in J} (|\hat{h}'_j| + 1)} < 1.$$

Which provides a contradiction and concludes the proof since the case $\prod_{i \in I} (|\hat{h}_i| + 1) < \prod_{j \in J} (|\hat{h}'_j| + 1)$ is analogous. □

The next step is to prove that the conditions given Propositions 1 and 3 are not only sufficient but also necessary. In order to prove this we need a characterization of consistent assessments.

Lemma 3 (Kreps and Wilson (1982, Lemma A2)). *Let (C, Y) be a consistent basis and let (σ, μ) satisfy $\mathscr{C}(\sigma) = C$ and $\mathrm{supp}(\mu) = Y$. The assessment (σ, μ) is consistent if and only if there exists a function $\pi : A \to (0, 1)$ such that $\pi(a) = \sigma(a \mid h)$ whenever $\sigma(a \mid h) > 0$ and, moreover, for every $x \in X$ with $\mu(x \mid h) > 0$:*

$$(4.2) \qquad \mu(x \mid h) = \frac{\displaystyle\prod_{a \in \mathscr{P}(x)} \pi(a)}{\displaystyle\sum_{\{x' \in h : \mu(x' \mid h) > 0\}} \left(\prod_{a \in \mathscr{P}(x')} \pi(a) \right)}.$$

Now we can turn to prove Theorem 1.

Proof of Theorem 1. Fix an extensive-form that does not satisfy neither the conditions of Proposition 1 nor the conditions of Proposition 3. Given any carrier C a consistent basis (C, Y) always exists (Lemma 1 gives a way of seeing this). Take a consistent assessment (σ, μ) with $\mathscr{C}(\sigma) = C$ and $\mathrm{supp}(\mu) = Y$. Let L be the associated labelling and let π be a function such as the one in equation (4.2).

The collection of non-singleton information sets h that satisfy $h \subset X^0(C)$ is denoted H^0. Take any information set $h \in H^0$. Is is enough to prove that for every μ' that only differs from μ at information set h, i.e. satisfies $\mu'(\cdot \mid h') = \mu(\cdot \mid h')$ for every $h' \neq h$, the assessment (σ, μ') is consistent.

First we show that if $\text{supp}(\mu') = \text{supp}(\mu)$ then (σ, μ') is consistent. Given the system of beliefs μ' we are going to construct a function π' such as the one in Lemma 3 that justifies it. Fix an arbitrary node x^* that belongs to h and is y, the support of μ and μ'. Let $\pi'(a) = \pi(a)$ for every $a \in \mathscr{P}(x^*)$. For the rest of the nodes in h and Y we only modify the value taken by π' with respect to π for just one choice in its path. In symbols, for each $x \in (h \cap Y) \setminus \{x^*\}$ choose any action $a_x \in \mathscr{P}(x) \setminus C$ and let $\pi'(a) = \pi(a)$ for every other $a \in \mathscr{P}(x) \setminus \{a_x\}$. The value of $\pi'(a_x)$ is given so that we adjust the relative values of μ' with respect to μ appropriately:

$$(4.3) \qquad \pi'(a_x) = \frac{\mu'(x \mid h)}{\mu'(x^* \mid h)} \frac{\mu(x^* \mid h)}{\mu(x \mid h)} \pi(a_x).$$

Equation 4.3 may have modified the value taken by one or several choices that are in the path to a node that is not in h. To keep track of those changes we let the set A_h consists of those actions whose value under π' has been assigned by (4.3). And the set Y_h consists of those nodes y that belong to some information set in $H^0 \setminus \{h\}$ and that have an action in their paths that belong to A_h. By assumption, a node in Y_h may contain in its path more than one choice in A_h but Y_h cannot contain two nodes that belong to the same information set.

For each $y^* \in Y_h$ we maintain $\pi'(a) = \pi(a)$ for every $a \in \mathscr{P}(y^*) \setminus A_h$. For the rest of the nodes $y \in h(y^*) \setminus \{y^*\}$ that belong to Y, the support of μ, we choose any action $a_y \in \mathscr{P}(y) \setminus C$ and let $\pi'(a) = \pi(a)$ for every other action $a \in \mathscr{P}(y) \setminus \{a_y\}$. We have to adjust the value of π' to maintain in the information set $h(y^*)$ the same beliefs as in μ. To do that we offset the changes made in 4.3 so that

$$(4.4) \qquad \pi'(a_y) = \pi(a_y) \prod_{a \in \mathscr{P}(y^*) \cap A_h} \left(\frac{\pi(a)}{\pi'(a)} \right).$$

Again we can define the set of actions $A_{h(y^*)}$ whose value under π' has been defined by (4.4) and the set $Y_{h(y^*)}$ of nodes that belong to some information set in $H^0 \setminus h(y^*)$ and that satisfy $\mathscr{P}(y) \cap A_{h(y^*)} \neq \varnothing$. The set $A_{h(y^*)}$ does not contain two nodes from the same information set and, since the conditions given in Proposition 3 are not met, it does not contain nodes in h or $h(y'^*)$ either, for any $y'^* \in A_h$.

Since the set H^0 is finite, we can continue in the same fashion until all the actions in the paths to nodes in information sets that belong to H^0 are exhausted without redefining any value of π'. Finally, we have to set $\pi'(a) = \pi(a)$ for every unassigned a. One can check that the resulting π' satisfies equation (4.2) for the system of beliefs μ'.

Now we prove that for any $x^* \in h$ the basses $(C, Y \cup \{x^*\})$ and $(C, Y \setminus \{x^*\})$ are also consistent. We show it first for the basis $(C, Y \cup \{x^*\})$.

Let $Y' = Y \cup \{x^*\}$. As estated in Lemma 1 we are going to construct a C-labelling L' such that $x \in Y'$ if and only if x minimizes $S_{L'}(\cdot)$ over $h(x)$. Set $L'(a) = L(a)$ for every $a \in \mathscr{P}(x^*)$ and for the rest of the nodes $x \neq x^*$ in

h take an arbitrary $a_x \in \mathscr{P}(x) \setminus C$ and let

(4.5) $$L'(a_x) = L(a_x) + F_L(x^*) - F_L(x).$$

We fix $L'(a) = L(a)$ for every other action $a \in \mathscr{P}(x) \setminus \{a_x\}$. That is, we are adjusting L' so that $S_{L'}(x) = S_{L'}(x^*)$ for every $x \in Y$.

We will assign the remaining values of L' recursively. For the same reasons as before, we know that no value is going to be redefined. Let A_h be the set of those actions whose value under L' has been assigned in (4.5) and let Y_h be the set of those nodes y that belong to some information set in $H^0 \setminus \{h\}$ and whose paths have an action in A_h. For each $y^* \in Y_h$ we fix $L'(a) = L(a)$ for every action $a \in \mathscr{P}(y^*)$ and for each $y \in h(y^*) \setminus \{y^*\}$ select an arbitrary $a_y \in \mathscr{P}(y) \setminus C$. Let $L'(a) = L(a)$ for every $a \in \mathscr{P}(y) \setminus \{a_y\}$ and

$$L'(a_y) = L(a_y) + \sum_{a \in \mathscr{P}(y^*) \cap A_h} \big(L'(a) - L(a) \big).$$

We can continue in the same fashion until we have exhausted all the actions in the paths to the nodes that belong to some information set in H^0. In order to make L' completely defined let $L'(a) = L(a)$ for every action that remains unassigned. It is easy to check that the labelling L' satisfies the condition given in Lemma 1 for the basis (C, Y').

To conclude it remains to show that for any $x^* \in h$ the basis $(C, Y \setminus \{x^*\})$ is also consistent. Take an arbitrary $a_{x^*} \in \mathscr{P}(x^*) \setminus C$ and let $L'(a_{x^*}) = L(a_{x^*}) + 1$. We fix $L'(a) = L(a)$ for every other action $a \in \mathscr{P}(x^*) \setminus \{a_{x^*}\}$ in the path to x^* and also for every action $a \in \mathscr{P}(x)$ in the path to any other node $x \in h$ different form x^*. The next step is to assign the values of L' for those actions leading to nodes contained in each information set $h(y^*) \in H^0$ that satisfies $a_{x^*} \in \mathscr{P}(y^*)$. Since hereafter everything is analogous to the previous case we can conclude the proof. $\square$

5. SEQUENTIALLY RATIONAL BAYESIAN ASSESSMENTS

In this section we consider extensive-form games and sequentially rational Bayesian assessments. Obviously, if for an extensive-form every Bayesian assessment is consistent then, for every payoff vector, every sequentially rational Bayesian assessment is a sequential equilibrium.[13] Suppose that we are given an extensive-form where some Bayesian assessment is not consistent. We want to address whether we can always find payoffs so that in the resulting extensive-form game sequential equilibrium refines the set of sequentially rational Bayesian assessments.

We first introduce some additional notation needed to define sequential rationality. Denote as σ_n the restriction of σ to those information sets owned by Player n and denote as σ_{-n} the restriction of σ to the remaining information sets. A behavioral strategy profile σ induces a probability distribution P^σ on the set of final nodes Z. The expected utility to player n is then given

---

[13] A sequentially rational Bayesian assessment is a *weak perfect Bayesian equilibrium* as defined by Mas-Colell, Whinston, and Green (1995, p. 285).

by the expression $U_n(\sigma) = \sum_{z \in Z} P^\sigma(z) u_n(z)$. Let P_x^σ be the probability distribution on Z if players use the strategy profile σ and the game starts at the decision node x. (Note that P_x^σ is always well defined.) The expected utility to player n from the strategy profile σ at the information set h given the system of beliefs μ is equal to $U_n(\sigma \mid h, \mu) = \sum_{x \in h} \mu(x \mid h) \sum_{z \in Z} P_x^\sigma(z) u_n(z)$.

Definition 4. The assessment (σ, μ) is sequentially rational if at every information set h, the strategy of the player moving at information set h, say Player n, satisfies

$$U_n(\sigma \mid h, \mu) \geq U_n(\sigma_{-n}, \sigma_n' \mid h, \mu) \text{ for every } \sigma_n' \in \Sigma_n.$$

The next lemma asserts that if we can find Bayesian assessments that are not consistent then, for some payoffs, there are behavioral strategies that are part of sequentially rational Bayesian assessments that are not sequential equilibrium strategies. The proof of the theorem consists of constructing such a payoff vector.

Lemma 4. *Consider an extensive-form where the set of set consistent assessments is strictly contained in the set of Bayesian assessments. Then we can find a game with that extensive-form such that set of sequential equilibrium strategies is a strict subset of the projection on Σ from the set of sequentially rational Bayesian assessments.*

Proof. Let K be such that $\Phi_K \neq \varnothing$ and either $\Phi_{K-1} = \varnothing$ or $K = 1$. Propositions 1 and 3 imply that we can find a carrier C and K information sets $h_1, \ldots, h_K$ that belong to $X^0(C)$ such that, for every consistent assessment (σ, μ) with $\mathscr{C}(\sigma) = C$, each h_i strictly contains a subset $\hat{h}_i$ with $\prod_{i=1}^{K} \left(\sum_{y \in \hat{h}_i} \mu(y \mid h) \right) < 1$. That is, if (σ, μ) is a consistent assessment there must be at least one information set $h_i \in \{h_1, \ldots, h_K\}$ with at least one node $x \in h_i \setminus \hat{h}_i$ that satisfies $\mu(x \mid h_i) > 0$.

For each $i = 1, \ldots, K$ let c_i be an action available at h_i such that $\sigma(c_i \mid h_i) = 0$. (If at least one does not exist we only need to modify the carrier C appropriately.) Assign a payoff equal to zero to the player who moves at h_i at every ending node that follows some action in $A(h_i) \setminus \{c_i\}$. Also assign a payoff equal to zero to ending nodes that follow action c_i when taken at any node in $\hat{h}_i$. Assign a payoff equal to 1 to every player elsewhere. A Bayesian assessment (σ, μ') such that $\prod_{i=1}^{K} \left(\sum_{y \in \hat{h}_i} \mu'(y \mid h_i) \right) = 1$ is sequentially rational but not consistent. $\square$

A possible criticism to the relevance of Lemma 4 is that (as the proof takes advantage of) differences in strategies may only occur in those parts of the tree reached with probability zero. In principle, we would like to show that if some Bayesian assessment is not consistent then, for some payoffs, sequential equilibrium selects only a strict subset from the set outcomes generated by sequentially rational Bayesian assessments. However this may not be possible. We illustrate why by means of three examples.
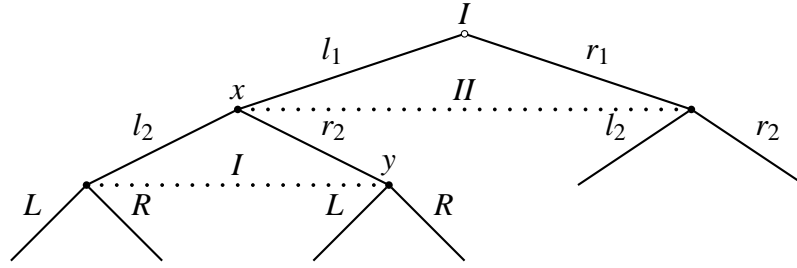
FIGURE 9.

Consider Figure 9. Proposition 1 together with Lemma 4 dictates that there exists a payoff assignment such that the strategy of some sequentially rational Bayesian assessment is not a sequential equilibrium strategy. If Player *I* moves r_1 at the root of the game then action l_1 is taken with probability zero and it is common to the paths to the two nodes in the second information set of Player *I*. Knowing this, we can assign payoffs to ending nodes as described in the proof of Lemma 4. However, we cannot find payoffs such that the set of outcomes differ. The reason is that sequential rationality does not let Player *I* deceive himself. (If we replace Player *I* by a third player in her second information set then there exists a game with that extensive-form where the sets of outcomes generated by sequentially rational Bayesian assessments and sequential equilibria are different.[14]) Suppose that Player *I* takes action r_1. It is unimportant that she does not use Player *II*'s strategy to construct her beliefs at her second information set as consistency requires. Hence, if moving r_1 is optimal then behavior at her second information set is irrelevant, even for Player *II*, who will always play a best response to Player *I* behavior because her information set is reached with positive probability.[15]

A second example is Figure 5 after substituting Player *III* by Player *II* so that she has two consecutive information sets. Again, for every assignment of payoffs to ending nodes, every outcome generated by a sequentially rational Bayesian assessment is a sequential equilibrium outcome. Note that sequential rationality by itself may impose restrictions on beliefs at unreached information sets. That is, suppose that Player *I* moves *Out* and that for some payoffs and beliefs at the first information set of Player *II* this player prefers (r_2, L) to both (r_2, R) and l_2. Then beliefs at her second

---

[14] From left to right assign the following payoffs to ending nodes: $(0,0,1)$, $(0,0,0)$, $(0,1,0)$, $(2,1,1)$, $(0,0,0)$ and $(1,1,0)$. The outcome $(1,1,0)$ is generated by a sequentially rational Bayesian assessment. The unique sequential equilibrium outcome is $(2,1,1)$.

[15] Suppose that we add a move *Out* at the root of the extensive-form that leads to an ending node so that if player *I* moves *Out* Player *II*'s information set is reached with probability zero. Give payoff $(1,0)$ to that ending node. To the rest of ending nodes, and from left to right, assign payoffs $(0,0)$, $(.5,0)$, $(2,0)$, $(2,1)$, $(0,0)$ and $(2,1)$. The sequentially rational Bayesian assessment $(Out, l_2, L, \mu(x) = 1, \mu(y) = 1)$ generates outcome $(1,0)$. But this is not a sequential equilibrium outcome. Notice, in particular, that that assessment is not consistent.

information set must be such that she, in fact, prefers *L* to *R*. If otherwise Player *II* prefers l_2 to both (r_2, L) and (r_2, R) then sequential rationality does not impose restriction on beliefs. On the other hand, consistency does not give any degree of freedom when choosing beliefs at Player *II*'s second information set. This explains why in this example we can obtain differences in behavior as implied by sequentially rational Bayesian assessments and sequential equilibrium but only at unreached parts of the game tree. For somewhat similar reasons, the same is also true when Player *I*'s action *Out* is removed from the extensive-form.

The example in Figure 10 is quite different in nature. The five open circles are the initial nodes, each of them is selected with equal probability by Nature. Proposition 3 implies that in this extensive-form some Bayesian assessment is not consistent. Those assessments must attach probability zero to the two information sets of Player *IV*. That means that the actions l_1, r_2, l_3 and r_4 have to be taken with probability one which leaves, for instance, actions r_1 and l_2 as the two actions that Proposition 3 requires for $K = 2$. (This corresponds to the carrier C^* constructed in the proofs of propositions 2 and 3.) In this example, consistent Beliefs can be arbitrary at the bottom information set of Player *IV* but they imposes restrictions on the set of consistent beliefs at her top information set. Bayesian beliefs can take, by definition, be arbitrary values at both information sets. Consider now any game with that extensive-form. Whether or not actions actions l_1, r_2, l_3 and r_4 are sequentially rational does not depend on what is the behavior at the top information set of Player *IV*. The reason is that that information set can only be reached from zero probability nodes at positive probability information sets. This implies that if both information sets of Player *IV* are reached with probability zero the strategy part of a sequentially rational Bayesian assessment and a sequential equilibrium strategy may only difer in behavior at Player *IV*'s top information set. However, behavior at that information set cannot affect the sequential equilibrium path.

It seems, however, that these extensive-forms are rather particular and that, although this is difficult to characterize in any precise manner, if consistent assessments strictly refine Bayesian assessments then, typically, a payoff vector exists such that sequential equilibrium generates a smaller set of outcomes as compared to that generated by sequentially rational Bayesian assessments. See footnotes 14 and 15 for some examples. For another example, assign payoff $(1, 1, 1,)$ after *Out* in the extensive-form of Figure 3. Then from left to right $(2, 0, 1)$, $(0, 1, 0)$, $(2, 1, 0)$, $(0, 0, 1)$, $(0, 0, 1)$, and $(0, 0, 0)$. The outcome $(1, 1, 1)$ can be supported a the Bayesian assessment where players *I* and *II* play (Out, l_2) and Player *III* puts a belief equal to one to the center node of her information set. Similar examples can be constructed for the extensive-forms in figures 6, 7 and 8.
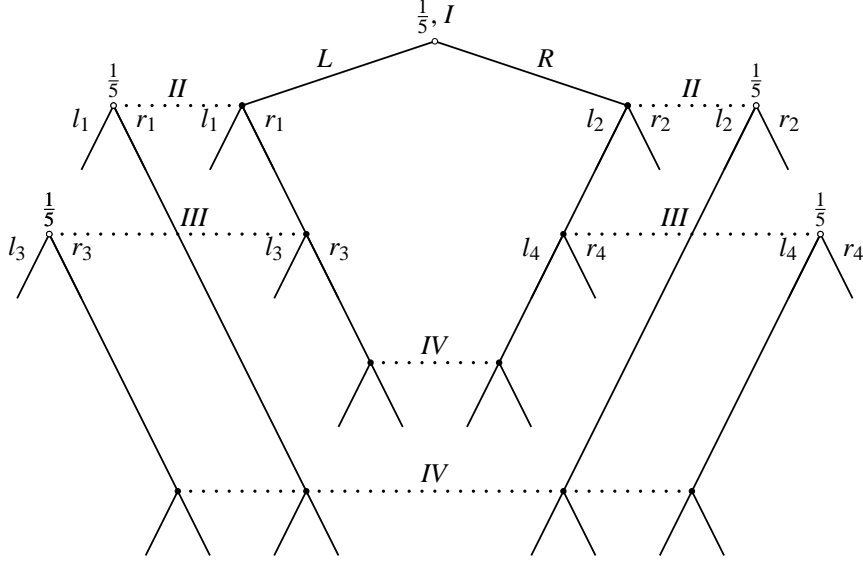
FIGURE 10.

6. APPLICATIONS

The proofs of the propositions and of Theorem 1 are valid for slightly stronger results that have the potential of being more useful in applications. If we have some information about a sequential equilibrium candidate, such as its carrier, the outcome that it induces, or just some set of choices that are to be taken with positive probability, then we can apply the results taking advantage of that information. That is, in applications, no longer do we need to find one behavioral strategy profile to check whether or not every Bayesian assessment is consistent. Instead we can simply focus on a particular family of profiles of interest. Consider the following results.

Proposition 1'. *Take an extensive-form without proper subforms and a set of actions $B \subset A$. If we can find a carrier C that contains the set B and an information set h with two nodes x and y such that*

  (i) the information set h is included in $X^0(C)$; and
  (ii) there is an action $\bar{a} \in \mathscr{P}(x) \cap \mathscr{P}(y)$ with $\bar{a} \notin C$.

Then there exists a behavioral strategy profile σ with $C \subset \mathscr{C}(\sigma)$ such that some Bayesian assessment (σ, μ) is not consistent.

Proposition 3'. *Take an extensive-form without proper subforms and a set of actions $B \subset A$. If we can find a carrier C that contains the set B and K information sets $h_1, \ldots, h_K$, where each information set h_i contains two nodes x_i and y_i, such that*

  (i) the information set h_i is included in $X^0(C)$; and
  (ii) there is an action $\bar{a}_i \in \mathscr{P}(x_i) \cap \mathscr{P}(x_{i+1})$ with $\bar{a}_i \notin C$.

Then there exists a behavioral strategy profile σ with $C \subset \mathscr{C}(\sigma)$ such that some Bayesian assessment (σ, μ) is not consistent.

To see why these results hold, notice that the proofs of the propositions start by selecting one arbitrary element out of a set where we allowed any carrier satisfying the conditions in the proposition. If we restrict that original set (but always allowing it to contain "bigger" carriers) then the proof goes through without modifications.

The proof of Theorem 1 fixes an arbitrary carrier and then asks if every strategy with that carrier is consistent. Hence, the following also holds:

Theorem 1'. *Take an extensive-form without proper subforms and a set of actions $B \subset A$ for which the conditions in propositions 1' and 3' are not satisfied. For every strategy σ with $B \subset \mathscr{C}(\sigma)$ every Bayesian assessment (σ, μ) is consistent.*

In our examples, these results imply that in figures 3, 5, 7 and 8, if Player *I* does not move *Out* in the strategy of the Bayesian assessment (σ, μ) then (σ, μ) is consistent. Likewise, if (σ, μ) is a Bayesian assessment in Figure 6 such that Player *I* does not move (F_1, F_2) then (σ, μ) is consistent.

To conclude, we have disentangled the different restrictions on beliefs imposed by consistency at zero probability information sets. Some restrictions apply to a single information set in isolation (Proposition 1) and some other restrictions apply to K different information sets when considered all together (Proposition 3). That is, if the conditions of Proposition 3 are satisfied for K information sets and not for any $K' < K$ then we can choose arbitrary beliefs at $K - 1$ of those information sets but this would, for some behavioral strategy profile, restrict consistent beliefs at the remainder information set. Moreover, these propositions are necessary and sufficient. This can help establish similar results between sequentiality and stronger restrictions on assessments that ask for more than simply Bayesian updating.

REFERENCES

BATTIGALLI, P. (1996): "Strategic independence and perfect Bayesian equilibria," *J. Econ. Theory*, 70(1), 201–234.

BLUME, L., A. BRANDENBURGER, AND E. DEKEL (1991): "Lexicographic probabilities and choice under uncertainty," *Econometrica*, 59(1), 61–79.

FUDENBERG, D., AND J. TIROLE (1991): "Perfect Bayesian Equilibrium and Sequential Equilibrium," *J. Econ. Theory*, 53, 236–260.

HAMMOND, P. (1994): "Elementary non-Archimedean representations of probability for decision theory and games," in *Patrick Suppes: Scientific Philosopher*, vol. 1, Probability and Probabilistic causality, pp. 25–25. Kluwer Academic.

KOHLBERG, E. (1990): "Refinement of Nash Equilibrium: The Main Ideas," in *Game Theory and Applications*, ed. by T. Ichiishi, A. Neyman, and Y. Tauman, pp. 3–45. Academic Press, San Diego.

KOHLBERG, E., AND P. RENY (1997): "Independence on Relative Probability Spaces and Consistent Assessments in Game Trees," *J. Econ. Theory*, 75(2), 280–313.

KREPS, D., AND R. WILSON (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

LITAN, C., AND C. PIMIENTA (2008): "Conditions for Equivalence Between Sequentiality and Subgame Perfection," *Econ. Theory*, 35(3), 539–553.

MAS-COLELL, A., M. WHINSTON, AND J. GREEN (1995): *Microeconomic theory*. Oxford.

MCLENNAN, A. (1989a): "Consistent conditional systems in noncooperative game theory," *Int. J. Game Theory*, 18(2), 141–174.

MCLENNAN, A. (1989b): "The space of conditional systems is a ball," *Int. J. Game Theory*, 18, 125–139.

MYERSON, R. (1986): "Multistage games with communication," *Econometrica*, pp. 323–358.

OSBORNE, M., AND A. RUBINSTEIN (1994): *A Course in Game Theory*. The MIT Press, Cambridge.

PEREA Y MONSUWÉ, A., M. JANSEN, AND H. PETERS (1997): "Characterization of consistent assessments in extensive form games," *Games Econ. Behav.*, 21(1-2), 238–252.

SWINKELS, J. M. (1993): "Independence for Conditional Probability Systems," Discussion Papers 1076, Northwestern University, Center for Mathematical Studies in Economics and Management Science.