GAMES OF INCOMPLETE INFORMATION WITHOUT COMMON KNOWLEDGE PRIORS

József Sákovics*

Abstract

We relax the assumption that priors are common knowledge, in the standard model of games of incomplete information. We make the realistic assumption that the players are boundedly rational: they base their actions on finite-order belief hierarchies. When the different layers of beliefs are independent of each other, we can retain Harsányi's type-space, and we can define straightforward generalizations of Bayesian Nash Equilibrium (BNE) and Rationalizability in our context. Since neither of these concepts is quite satisfactory, we propose a hybrid concept, Mirage Equilibrium, providing us with a practical tool to work with inconsistent belief hierarchies. When the different layers of beliefs are correlated, we must enlarge the type-space to include the parametric beliefs. This presents us with the difficulty of the inherent openness of finite belief subspaces. Appealing to bounded rationality once more, we posit that the players believe that their opponent holds a belief hierarchy one layer shorter than they do and we provide alternative generalizations of BNE and Rationalizability. Finally, we show that, when beliefs are degenerate point beliefs, the definition of Mirage Equilibrium coincides with that of the generalized BNE.

1 Proem

Consider the following simple situation. Bob is competing in a Cournot duopoly with Ann. The two of them produce perfect substitutes at zero cost and compete in quantities. Unfortunately, the demand for their product is not known. There are two demand forecasting agencies in town, offering two different predictions, type 1: p = q, and type 2: p = 2q. Bob is known to use the type 1 model,[1]

*Thanks to Pierpaolo Battigalli, James Bergin, Roberto Burguet, Françoise Forges, Esther Gal-Or, Joseph Greenberg, Joe Harrington, Alejandro Manelli, Michael Maschler, Debraj Ray, Hyun Shin, Fernando Vega, Asher Wolinsky and attendants of my presentations given at Cergy-Pontoise, Gerzensee, Boston University, Bellaterra, Toronto and Milan. I gratefully acknowledge support from the HCM program of the EC through contract CHRX-CT94-0489 "Games and Markets" and from DGCYT through grant PB93-0679.

[1] We assume this for simplicity.

however, he does not know which agency Ann uses. His prior is that, with probability α she also uses model 1 – and thus she uses model 2 with probability $1-\alpha$. Moreover, he believes that Ann's belief about α is described by the probability density function $\gamma(.)$. How much would/should Bob produce if this were all the information he had available? In this paper, we would like to convince the reader of three points:

1. The above example is representative of interesting economic situations, which hitherto have not been rigorously analyzed by economists.
2. The reason for this apparent neglect is that the available game theoretic models/concepts cannot adequately address these situations.
3. There is a way to deal with these situations, which is consistent in spirit with the standard approach.

2 Introduction

In his seminal paper, János Harsányi (1967-68) invented a way to elude the difficulties associated with the infinite hierarchies of beliefs that arise as a consequence of asymmetric information about the payoffs[2] in a game. His contribution was decisive in opening up the possibilities for the analysis of games of incomplete information in the field of Economics (as it was recognized by the Royal Swedish Academy of Sciences in 1994). However, next to the enormous benefit it created, his theory also caused the negative externality of drawing game theorists' attention away from the analysis of higher-order beliefs.[3] The convenient assumption of priors being common knowledge – and often, even common – is predominant not only in the literature but also in the minds of economists in general. The aim of this paper is to complement the "Harsányi doctrine" (here understood more generally, as the assumption of common knowledge priors) with a less restrictive approach, which allows for the modeling of irreconcilable differences among the players' beliefs, without sacrificing tractability. That is, we propose a level of analysis that is intermediate between Harsányi's and Mertens and Zamir's (1985) treatment. The latter two authors do provide a general (in fact, "universal") model of belief hierarchies, however, their treatment has not proven useful for applied analysis.

In game situations where the assumption of common knowledge priors is inappropriate[4] – that is, there are reasons to believe that the players do not know each other's beliefs – it is necessary to analyze the effects of variations

[2]As Harsányi himself argues, all other types of uncertainty (about strategy sets, number of players etc.) can be reduced to this general case.

[3]At least for a long while. See Monderer and Samet (1989), Morris, Rob and Shin (1995) or Morris, Postlewaite and Shin (1995), for examples of recent contributions.

[4]This, actually, is a very large set, encompassing all areas of economic theory. Of course, there are many variants of the example given in the proem. The uncertainty may be about productivity, external shocks etc. However, Industrial Organization is not the only field for

in the higher-order beliefs to obtain some reliable prediction/prescription about player behavior. Once we desist from imposing common knowledge of priors, we open up a Pandora's box, not only because we have to explicitly consider the maze of higher-order beliefs but mainly since we are no longer obliged to posit that these beliefs are correct. Following the approach laid out in this paper, we are able to model any situation where players hold higher-order beliefs which are inconsistent with the"truth". Such inconsistencies are all the more relevant to investigate, since they are not only present in the static settings analyzed in this paper, but may even persist over time (see the experimental literature on false equilibria, for example, Camerer and Weigelt, 1991).

In a strategic situation, before deciding on what to do, a player has to predict how her opponents will behave. To be able to do that, she has to have a belief about her opponents' view of the game. But they also have to do the same, etc. and therefore, the uncertainty about the game being played generates an infinite belief hierarchy. Unless the game played is common knowledge, these hierarchies make the analysis practically intractable. Harsányi (1967-68) has argued that if the game itself is not common knowledge but the distribution which each possible view(belief hierarchy)-holder – *type* – attaches to the space of these types *is* common knowledge, we can carry out a meaningful analysis in the context of a game of imperfect information, with the above mentioned types as its players. This has brought us to the concept of Bayesian games. These games are versatile, encompassing many different types of asymmetries among the players.

Indeed, we start our study by setting up the benchmark Bayesian game proposed by Harsányi. Having defined the framework, we are then able to give the definitions of the two salient solution concepts in this context: Rationalizability and Harsányi's Bayesian Nash Equilibrium (BNE).

The next question has been, what if the *Bayesian* game played is not common knowledge? By redefining the notion of a type as one which includes the entire belief hierarchy, Mertens and Zamir (1985) showed that, by sufficiently enlarging the type-space – that is, explicitly modeling more and more levels of beliefs – we can replace the assumption of common knowledge of priors (that is, first-order beliefs) over the attributes by common knowledge of beliefs over this "universal" type-space. Thus, if we substitute the attribute-space with the universal type-space, we obtain a game that is qualitatively equivalent to Harsányi's, and in particular is one with common knowledge "priors". Their construction is based on the proof that the description of a universal type contains all its beliefs, including the one over the universal type-space. That is, common knowledge of beliefs can be derived as a result, instead of having to assume it. While the closedness property of the universal belief-spaces does provide the theoretical foundations for Harsányi's model, we claim that for applications it does not

---

which the analysis developed in this paper may be relevant. Consider, for example, the following: do countries know each other's CGE models in a game of international macroeconomic policy coordination?; do we really need noise traders for trade in equilibrium in a financial market?; etc.

serve the purpose. A problem with the Mertens-Zamir types is that – unless some regularity assumptions are made – in general,[5] they are of such complexity that they are not even describable and, therefore, they are not useful for applied analysis. Therefore, if we want to retain the tractability of Harsányi's approach, we need to give a different answer to the question asked above. This is what we do in this paper.

The need for tractability and the absence of common knowledge leads us to model the view of each player as a game where the players hold *finite* belief hierarchies over some primitive space of attributes, which do not include beliefs. The assumption of finite depth reasoning – apart from being very plausible – has been corroborated by both theoretical and experimental evidence. Stahl (1993) showed that in an evolutionary setting, a population initially composed of players of different depth of reasoning does not converge in general towards an infinitely smart population. Stahl and Wilson (1995) found that the hypothesis of the presence of a "rational expectations type" in the pool of their experimental subjects – corresponding to a player who forms unbounded belief hierarchies – can be rejected. Nagel (1995) also demonstrated that her model of finite steps of reasoning is consistent with the observed behavior in her experiment.

We continue our inquiry along two strands, depending on whether the different layers of the belief hierarchy are independent. When they are, it is possible – though not necessary – to maintain the definition of a strategy as a function exclusively of the attributes. This result follows from the reinterpretation of higher order beliefs as beliefs over the attribute space, by "integrating out" the rest of the variables. With this simple type-space, it is straightforward to define the appropriate generalizations of BNE and Rationalizaility (Subjective Bayesian Equilibrium, SBE, and Partially Subjective Rationalizability, PASUR, respectively). We have found neither of these new concepts satisfactory for application. SBE suffers from the caveat that it hypothesizes a strategy for the other player, which is not consistent with the solution itself. On the other hand, PASUR is very weak, since the highest order strategic beliefs are assumed unconstrained. In order to provide an internally consistent *and* powerful solution for this context, we define a hybrid concept between SBE and PASUR: Mirage Equilibrium. In a Mirage equilibrium, each player forms a strategic belief hierarchy (of the same length as the parametric belief hierarchy, which is given as part of the description of the game). These strategic belief hierarchies satisfy two conditions. Every layer is a best response to the next one conditional on the parametric beliefs and, the last two layers form a SBE.

Next, we turn to the case of general (though still finite) belief hierarchies. In this case, it is no longer possible to retain all the relevant information with the strategies depending only on the attributes. On the other hand, if we allowed the entire belief hierarchy to form part of the strategies' arguments, we would run into the problem that the players would not be able to evaluate the

[5]Note that we do not claim that there are no models using the universal types, which can be parsimoniously described. Rather, that the applied analyst is likely to encounter situations, where this is not the case.

expected actions of their opponents because of the lack of an additional layer of belief. This is the openness property of finite-order belief spaces. To resolve this problem, we assume that every player believes that the other player conditions his strategy on one fewer layers of beliefs than herself.[6] This assumption clears the way for the generalization of Rationalizability to this context, however, for a Nash-like concept, we need to make a further choice. Note that, the above assumption implies that when we check that two strategies are best responses to each other, the same strategy would appear with a different number of arguments at different places. In order to restore the equivalence, we augment the domain of the higher-order (and therefore the one of fewer arguments) strategic belief with the corresponding arguments of the lower-order belief. The idea behind this rule is that this way, when the beliefs are correct we recover BNE.

Finally, we carry out a consistency check between our solution concepts. To this end, we assume that, apart from the beliefs about the underlying attribute-space (that is, the first-order beliefs), the rest of the beliefs are degenerate point-beliefs. Calculating the solution concepts defined for the general case under this assumption, we show that the generalized BNE collapses into Mirage Equilibrium, while the generalized Rationalizability into PASUR. This result gives further credibility to the concept of Mirage Equilibrium.

We close the paper with some remarks on the difficulties in extending our results to dynamic games.

3 The Standard Model

In order to fix ideas, in this section we discuss Harsányi's theory. Let us begin with some preliminaries.

For a finite set of N players indexed from 1 to N, we define an *N-person game of incomplete information* as the collection of

i) an underlying space of attributes (types), $K = \times_{i=1}^{N} K_i$,

ii) a set of feasible mixed actions, $S = \times_{i=1}^{N} S_i$,

iii) parametric utility functions for each player, $V_i = K_i \times S \to \Re$,

iv) beliefs over the attribute space (priors) for each player, $R_i \in \Delta(K)$.

The cornerstone assumption of Harsányi's theory is that the entire game, including the priors, is common knowledge among the players. Consequently, the players' strategies are functions of (only) the attributes, $\sigma_i : K_i \to S_i$. Given

[6] While we have found no direct empirical evidence for this phenomenon, it is well established that people tend to be overconfident in their judgements, overrating their information and/or abilities (see, for example, Schmalensee, 1976, Alpert and Raiffa, 1982, or Offerman, Sonnemans and Schram, 1996).

a strategy profile, the payoff function of Player i – that is, her objective function – is

$$U_i\left(k_i, \sigma_i(k_i), \sigma_{-i}(.)\right) = \sum_{\mathbf{k}_{-i} \in \mathbf{K}_{-i}} R_i(k_i, \mathbf{k}_{-i}) V_i\left(k_i, \sigma_i(k_i), \sigma_{-i}(\mathbf{k}_{-i})\right).$$

The fundamental building block for the analysis of the behavior of – even boundedly – rational agents is their best response mapping, which we define next.

Definition 1. *The strategy of Player i, $\sigma_i(.)$, is a* best response *to a profile of strategies of the rest of the players,$\sigma_{-i}(.)$ – denoted $\sigma_i \in BR(\sigma_{-i}(.), R_i)$ – if and only if for all $k_i \in K_i$, $\sigma_i(k_i)$ maximizes $U_i\left(k_i, \sigma_i(k_i), \sigma_{-i}(.)\right)$.*

Remark. For the purpose of the best response mapping, $\sigma_{-i}(.)$ can also be interpreted as a conjecture over the strategies of the players other than i, where the probability attached to each pure strategy is the result of taking expected values according to the original probability function, mapping mixed strategies to probabilities.[7] That is, if Player i believes, for example, that with probability 0.4 Player j is mixing fifty-fifty between pure actions α and β and with probability 0.6 she is taking pure action α, then this is equivalent to i believing almost surely that j is mixing 0.8-0.2 between α and β.

Armed with the best response mapping, we are now ready to give the definitions of what we think of as the possible "solutions" to our Bayesian game. In honor of its inventor, we start with Harsányi's equilibrium.

Definition 2. *The profile of strategies, σ, constitutes a* Bayesian Nash Equilibrium *(*BNE*) if and only if $\sigma_i \in BR(\sigma_{-i}(.), R_i)$, $i = 1, 2, ..., N$.*

While Nash Equilibria are self-enforcing, and they are a likely limit of the converging play in repeated games (see, for example, Kalai and Lehrer, 1993a), they are not necessarily the best predictors of strategic interaction, especially in a one-shot, static setting. On the other hand, if we only assume common knowledge of rationality (conditional utility maximization) on the part of the players, we arrive at a more natural (though less exclusive) concept of solution: Rationalizability. The extension of Bernheim's (1984) and Pearce's (1984) original concept to our setting is straightforward. Let us denote by $i\#j(k)$ a sequence of k player indices whose first element is i, last element is j and where no two successive elements are identical.

Definition 3. *The strategy profile $\sigma(.)$ is* Bayes Rationalizable *if and only if there exist strategy profiles $\sigma^{i\#j(k)}$, such that, for all i*
i) $\sigma_i \in BR(\sigma^i_{-i}(.), R_i)$,
ii) $\sigma^j_i \in BR(\sigma^{i,j}_{-j}(.), R_j)$ for $j \neq i$,
iii) $\sigma^{i\#m(n)}_k \in BR(\sigma^{i\#m(n),k}_{-k}(.), R_k)$ for $k \neq m$, and $n = 2, 3, ...$

[7]This is a standard result, see, for example, Pearce (1984) for a proof.

Here, a strategy like $\sigma_j^{i,k,i}$ is the belief i has about k's belief about i's belief about j's strategy. Given the equivalence between point beliefs and conjectures (c.f. the remark following the definition of the best response mapping), we suggest, however, that it should be thought of as the point-belief i holds about j's strategy.

In order to assist the reader in grasping the different information structures and the corresponding solution concepts appearing in this paper, we will work out as an example for each case the situation described in the Proem. For this section, we assume that the above setup – and especially Bob's prior – is common knowledge (and thus γ puts probability one on α). Then the BNE strategies of the three types are given by the solution to the following system:

$$\begin{aligned} b &= \arg\max_q q\,(1 - q - \alpha a_1 - (1-\alpha)a_2) \\ a_1 &= \arg\max_q q(1 - q - b) \\ a_2 &= \arg\max_q q(1 - 2(q + b)) \end{aligned} \tag{I}$$

Resolving (I) we obtain

$$b = \frac{3-\alpha}{6}, \quad a_1 = \frac{3+\alpha}{12}, \quad a_2 = \frac{\alpha}{12}.$$

Because of the strategic complementarities, it is straightforward to show that Bob has a unique rationalizable strategy and thus the set of rationalizable strategies is a singleton for each type, coinciding with the BNE.

4 Private Beliefs

In this section, we relax the cornerstone assumption of the Harsányi theory about the prior beliefs over the attribute space being common knowledge. As we discussed it in the Introduction, Mertens and Zamir (1985) have generalized Harsányi's result to this case, by enlarging the type-space. In order to complement the Mertens-Zamir construction in a way, which is more tractable, and also captures the bounded rationality of the players, we take an alternative route. Our main assumption is that the players only construct finite-depth[8] belief hierarchies over the attribute space. Given this restriction on the type-space, the Mertens-Zamir approach is no longer valid. Our new approach has two strands. Like in Auction Theory, we distinguish between the cases when information/beliefs are independent and when they are not. For simplicity, we restrict our attention to two-player games.

The lack of common knowledge about the game being played has some direct implications on what we require of "reasonable" solution concepts. For example, the self-enforcing nature of equilibrium is no longer the right characteristic to

[8]For notational simplicity, we assume that the belief hierarchies are composed of three layers. The generalization to more layers is straightforward.

look for. To see why, consider a candidate equilibrium profile which is such that all the players would be willing to follow their part if they expected the rest of the players to play according to the candidate profile (self-enforcingness); however, in some player's view, the strategy corresponding to some other player is not a best response for that player to the candidate profile. Clearly, equilibrium, rationality, and the players' beliefs are not reconcilable in this case.[9] We could try to resolve this problem by throwing away all the candidate equilibria that carry an inner contradiction as above. However, on the one hand, this may well leave us without an equilibrium. On the other hand, in a one-shot interaction, even if the social norm is to play Nash, there is no reason to expect in our setup any correlation among the players' choice of strategies. Therefore, we should consider each player as a separate Bayesian decision-maker (c.f. Tan and Werlang, 1998), who chooses her action according to her own view of the world, whether this view is correct or not. Thus, we suggest that each player should be able to derive her "optimal" strategy via eductive reasoning. Those who expect more collective rationality of equilibrium play can always impose a refinement, which forces strategies to satisfy the individual constraints simultaneously (c.f. Greenberg, 1996).

4.1 Independent Beliefs

In this subsection, we assume that the different layers of any belief hierarchy are uncorrelated. Note that we do not rule out correlation within the same layer: thus, a (subjective) prior distribution of the vector of attributes, or the distribution of the vector of beliefs about the attributes need not be independent. The most important consequence of this independence assumption is that the conditional expectations with respect to the different layers of beliefs can be evaluated separately. As we will see, under this scenario we can[10] maintain Harsányi's definition that types and attributes are equivalent, and thus players' strategies continue to be a function of only their attributes.

More formally, let us define Player i's beliefs as follows

$$R_i^1 \in \Delta(K), \quad R_i^2 \in \Delta\left(\Delta(K)\right), \quad R_i^3 \in \Delta\left(\Delta\left(\Delta(K)\right)\right).$$

That is, Player i's first-order belief is a probability distribution over the attribute space. Her second-order belief is a probability distribution over the first-order beliefs of Player j, etc.

Next, note that – for the purposes of (conditional) expected utility calculations, with strategies that only depend on the attributes – the beliefs described above can be integrated out[11] to yield

$$R_i \in \Delta(K), \quad R_i^j \in \Delta(K), \quad R_i^{ji} \in \Delta(K).$$

[9]Note that if the game is common knowledge, this problem does not arise in equilibrium, since whenever a player believes that another player should not follow the recommendation, that player in fact does not.

[10]Though we need not, see the next subsection.

[11]C.f. the remark following Definition 1.

That is, all the beliefs can be interpreted as beliefs over some player's attribute space. In particular, Player i can "use" her second-order belief just as if she knew Player j's belief about her attribute. Similarly, Player i's third-order belief represents the belief she thinks Player j thinks she has about his attribute. In other words, it is *as if* the higher-order beliefs were all degenerate point beliefs.

We are now ready to define the corresponding generalizations of Bayesian Nash Equilibrium and Bayes Rationalizability.

Definition 4. *Conditional on his beliefs, a strategy of Player i, $\sigma_i(.)$, forms part of a* Subjective Bayesian Equilibrium *(SBE) profile if and only if there exists a strategy $\sigma_j^i(.)$, $j \neq i$, such that*

$$\sigma_i \in BR(\sigma_j^i, R_i) \text{ and } \sigma_j^i \in BR(\sigma_i, R_i^j).$$

Note that, when Player i's second-order beliefs are correct, the sets of SBE and BNE strategies (not the profiles!) coincide. On the other hand, SBE suffers from the drawback that it does not use all the information available: it is independent of R_i^{ji}. This results in an internal inconsistency, since in Player i's mind, Player j does not behave according to SBE.

Definition 5. *Conditional on his beliefs, a strategy of Player i,$\sigma_i(.)$, is* Partially Subjective Rationalizable *(PASUR) if there exist strategies $\sigma_j^i(.)$, $\sigma_i^{ij}(.)$ and $\sigma_j^{iji}(.)$, $j \neq i$, such that*

$$\sigma_i \in BR(\sigma_j^i, R_i),\ \sigma_j^i \in BR(\sigma_i^{ij}, R_i^j) \text{ and } \sigma_i^{ij} \in BR(\sigma_j^{iji}, R_i^{ji}).$$

PASUR does use all the information, however, it is a rather weak concept, even when the beliefs consistent. In order to combine the strengths of SBE and PASUR, we propose the following hybrid concept instead:[12]

Definition 6. *Conditional on his beliefs, a strategy of Player i, $\sigma_i(.)$, forms part of a* Mirage Equilibrium *profile if and only if there exist strategies $\sigma_j^i(.)$ and $\sigma_i^{ij}(.)$, $j \neq i$, such that*

$$\sigma_i \in BR(\sigma_j^i, R_i),\ \sigma_j^i \in BR(\sigma_i^{ij}, R_i^j) \text{ and } \sigma_i^{ij} \in BR(\sigma_j^i, R_i^{ji}).$$

In a Mirage Equilibrium, the highest levels of the strategic belief hierarchy constitute a SBE, while the lower levels are calculated via the backward induction logic of PASUR. This assumption has both a positive and a normative reading. The descriptive interpretation is based on the fact that the SBE assumption is equivalent to positing that the players substitute their third-order strategic belief about their opponent's strategy σ_j^{iji} , with their first-order one, σ_j^i . We claim that such a substitution is plausible. Note that what our boundedly rational agent is trying to do is to check the "consistency" of her beliefs.

[12] It is immediate that Mirage Equilibrium is a refinement of PASUR.

She does that by ensuring that at every level, her strategic beliefs are a best response to the following level, conditional on her corresponding parametric beliefs. When she attempts the verification of her third-order strategic beliefs, she is at a loss since there are no parametric beliefs available. In the absence of verifiable third-order strategic beliefs, she could leave her second-order strategic beliefs unrestricted as well. However, a more sensible solution is to substitute the "missing" beliefs by the "closest" proxy available: the previous level of beliefs over the same object. This rule of thumb is consistent with the so-called "false consensus" effect found by Ross, Greene and House (1977). According to this phenomenon, experimental subjects tend to believe that others share their beliefs to a much larger extent than should follow from Bayesian updating.

A normative justification of the Mirage Equilibrium concept is based on the assumption that Nash behavior is to be desired from – even boundedly – rational agents. Consequently, they are supposed to conjecture Nash-like strategies whenever it does not contradict their information. In our case, in the lower levels of their (subjective) belief hierarchy the Nash conjecture would be inconsistent with their parametric beliefs together with their opponents' rationality. At the top of their belief hierarchy, however, this restriction disappears and they are free to posit Nash behavior. We offer a further justification in Subsection 4.3.

Returning to our Cournot example, it is easy to see that SBE coincides with BNE, since SBE does not use the information contained in $\gamma(.)$ and, therefore, even if it is inconsistent it does not show in the solution.

Bob's PASUR strategies can be deduced from the solutions of the following system:

$$\begin{aligned}
b &= \arg\max_q q\left(1-q-\alpha a_1^i-(1-\alpha)a_2^i\right),\\
a_1^i &= \arg\max_q q(1-q-b^{ij}),\\
a_2^i &= \arg\max_q q(1-2(q+b^{ij})),\\
b^{ij} &= \arg\max_q \int q\left(1-q-ya_1^{iji}-(1-y)a_2^{iji}\right)\gamma(y)dy\\
&= \arg\max_q q\left(1-q-\overline{\gamma}a_1^{iji}-(1-\overline{\gamma})a_2^{iji}\right),
\end{aligned} \tag{II}$$

where $\overline{\gamma}$ is the mean of $\gamma(.)$, and $a_1^{iji}\in[0,1/2]$ and $a_2^{iji}\in[0,1/4]$ are (strategically) unconstrained. Resolving (II) for b, with the extreme values of the unconstrained strategies, we obtain:

$$\frac{15-4\alpha-\overline{\gamma}}{32}\leq b\leq\frac{16-4\alpha}{32}.$$

Note that, when α is close to 1 and $\overline{\gamma}$ is close to zero, the SBE strategy is not PASUR! This is because SBE does not incorporate the highly inconsistent third-order belief.

Bob's Mirage strategy can be deduced from the solution of the following system:

$$\begin{aligned} b &= \arg\max_q q\left(1 - q - \alpha a_1^i - (1-\alpha)a_2^i\right), \\ a_1^i &= \arg\max_q q(1 - q - b^{ij}), \\ a_2^i &= \arg\max_q q(1 - 2(q + b^{ij})), \\ b^{ij} &= \arg\max_q q\left(1 - q - \overline{\gamma}a_1^i - (1-\overline{\gamma})a_2^i\right). \end{aligned} \tag{III}$$

Resolving (III) for b, we obtain:

$$b = \frac{16 - 4\alpha - 4/3\overline{\gamma}}{32}.$$

4.2 Correlated Beliefs

If we allow the different layers of beliefs to be correlated, we no longer can reduce them to distributions over the other player's attribute. So we have to work with

$$R_i^1 \in \Delta(K), \quad R_i^2 \in \Delta(K \times \Delta(K)),$$
$$R_i^3 \in \Delta(K \times \Delta(K) \times \Delta(K \times \Delta(K))).$$

Note that, given a higher-order belief, the lower-order ones are redundant, since they are given by the appropriate marginals of the higher-order belief.[13]

Given this more complex belief hierarchy, we need to reconsider the proper definition of a strategy in this context. The standard approach is that the strategy mapping has as its argument all the relevant parameters of a player that are not common knowledge, that is, the "types". When common knowledge is lacking, however, this rule is inappropriate. Since it is within the eductive reasoning of a player where the others' strategies appear, the right definition is that a player's strategy depends on all of the relevant parameters he has a belief about *and* are not common knowledge. In other words, since our theory is Bayesian, and thus it is based on the Savage axioms, everything whose existence a player knows about he has a belief about and, consequently, anything a player does not have a belief about does not exist in that player's world. Then, obviously, he should not expect his opponents' strategy to depend on such objects. This definition, results in every player believing that the other player conditions his strategy on one layer less of his belief hierarchy than herself.

Based on the above discussion, we define a strategy for Player i as $\sigma_i : K_i \times \Delta(K \times \Delta(K) \times \Delta(K \times \Delta(K))) \to S_i$, and for Player j as $\sigma_j^i : K_j \times \Delta(K \times \Delta(K)) \to S_j$. To follow in the tradition of Nash, we can then require that players should use strategies, which form part of a profile that is autoconfirming given their belief hierarchy. To choose among the many ways the model can be closed,

[13] Consequently, to be precise, we need to require that the hierarchy be "coherent", in the sense that it does not contain inner contradictions. For more on this aspect, see Brandenburger and Dekel (1993).

we impose continuity, in the sense that, when beliefs are correct, our solution should coincide with that of a model with common knowledge. Let us denote an arbitrary belief of Player i by r_i^m, $m \leq 3$.

Definition 7. *The strategy of Player i, $\sigma_i(k_i, r_i^3)$, forms part of an* Third-Order Subjective Bayesian Equilibrium *(TOSUBE) profile if and only if there exists a strategy profile $\sigma_j^i(k_j, r_j^2)$, such that*
i) $\sigma_i(k_i, r_i^3) \in BR(\sigma_j^i(k_j, r_j^2), r_i^3)$, *and*
ii) $\sigma_j^i(k_j, r_j^2) \in BR(\sigma_i(k_i, r_i^1 \times m \arg_{\Delta(K)\times\Delta(K\times\Delta(K))}\{r_i^3\}), r_j^2)$ *for* $j \neq i$.

The generalization of Rationalizability is more straightforward:

Definition 8. *The strategy of Player i, $\sigma_i(k_i, r_i^3)$, is* Third-Order Subjective Rationalizable *(TOSUR) if and only if there exist strategy profiles $\sigma_j^i(k_j, r_j^2), \sigma_i^{ij}(k_i, r_i^1)$ and $\sigma_j^{iji}(k_j)$, $j \neq i$, such that*
i) $\sigma_i(k_i, r_i^3) \in BR(\sigma_j^i(k_j, r_j^2), r_i^3)$,
ii) $\sigma_j^i(k_j, r_j^2) \in BR(\sigma_i^{ij}(k_i, r_i^1), r_j^2)$, *and*
iii) $\sigma_i^{ij}(k_i, r_i^1) \in BR(\sigma_i^{iji}(k_j), r_i^1)$.

Let is return to our example. Bob's TOSUBE strategy can be deduced from the solution of the following system:[14]

$$\begin{aligned} b(1, y, 1, 1_\gamma) &= \arg\max_q q\left(1 - q - ya(1, 1, \gamma) - (1-y)a(2, 1, \gamma)\right), \\ a(1, 1, \gamma) &= \arg\max_q \int q(1 - q - b(1, y, 1, 1_\gamma))\gamma(y)dy, \\ a(2, 1, \gamma) &= \arg\max_q \int q(1 - 2(q + b(1, y, 1, 1_\gamma)))\gamma(y)dy. \end{aligned} \tag{IV}$$

Resolving (IV) for $b(1, \alpha, 1, 1_\gamma)$, we obtain:

$$b(1, \alpha, 1, 1_\gamma) = \frac{12 - 3\alpha - \overline{\gamma}}{24},$$

which clearly coincides with his Mirage strategy, if only because the payoffs depend linearly on α. As we show in the next subsection, there exists a general equivalence result, when the belief space is further restricted.

4.3 Degenerate Higher-Order Beliefs: A Consistency Result

It is interesting to compare the solution concepts that we defined for independent and correlated beliefs. Needless to say, the latter ones can be directly applied even if beliefs are independent. The more relevant difference is what are the arguments of the strategies. Consequently, in order to reach a common ground

[14]The arguments of Bob's (b) and Ann's (a) strategies are the attribute and the layers of beliefs, while y is a dummy variable.

between the two types of solution, in this subsection, we restrict attention to the case when higher-order beliefs put all the mass on a single point in their support: they are point beliefs.

Proposition 1. *When higher-order beliefs are point beliefs, conditional on her beliefs, Player i's set of* TOSUBE *strategies coincides with her set of Mirage Equilibrium strategies, while her set of* TOSUR *strategies coincides with her set of* PASUR *strategies.*

Proof: We only prove the first statement, the proof of the second one is analogous. Let us write down the system of equations,[15] generated by the iterated use of the two rules in Definition 7, which needs to be solved to calculate Player i's FOSBE strategy, given her beliefs, $R_i^3 = k_i, P_i^1, P_i^2, P_i^3$:

$$\sigma_i\left(k_i, P_i^1, P_i^2, P_i^3\right) = \arg\max_\sigma \sum_{x\in K_j} P_i^1(x)V_i(k_i,\sigma,\sigma_j\left(x,P_i^2,P_i^3\right))$$

$$\sigma_j\left(x, P_i^2, P_i^3\right) = \arg\max_\sigma \sum_{y\in K_i} P_i^2(y)V_j(x,\sigma,\sigma_i\left(y,P_i^3,P_i^2,P_i^3\right))$$

$$\sigma_i\left(y, P_i^3, P_i^2, P_i^3\right) = \arg\max_\sigma \sum_{r\in K_j} P_i^3(r)V_i(y,\sigma,\sigma_j\left(r,P_i^2,P_i^3\right)).$$

Here we have used the fact that the beliefs $\left(P_i^2, P_i^3\right)$ are point beliefs, twice. Once when we substituted them in directly when taking expectations, and once when we interpreted them as beliefs over the parameter space. Next, note that we can rewrite the above "equations" taking into account that the P's are given, and therefore rather than variables they are parameters. Consequently, we can take them out of the arguments of the strategies and replace them, for example, by sequences of player indices:

$$\sigma_i\left(k_i\right) = \arg\max_\sigma \sum_{x\in K_j} P_i^1(x)V_i(k_i,\sigma,\sigma_j^i\left(x\right))$$

$$\sigma_j^i\left(x\right) = \arg\max_\sigma \sum_{y\in K_i} P_i^2(y)V_j(x,\sigma,\sigma_i^{ij}\left(y\right))$$

$$\sigma_i^{ij}\left(y\right) = \arg\max_\sigma \sum_{r\in K_j} P_i^3(r)V_i(y,\sigma,\sigma_j^i\left(r\right)).$$

But this is just the definition of Mirage Equilibrium. ■

[15] In fact, they are not equations but best response relations, unless the latter are single-valued. For our argument this distinction is immaterial.

5 Final Remarks

In this paper, we have sketched the "gains and pains" associated with leaving the standard paradigm of common knowledge priors. Rather than setting standards for this uncharted territory, our aim was to generate discussion, and to encourage other researchers to investigate the issues involved. We have given a first approximation, the rest should be filled in by people more able than us.

An obvious extension would be to go beyond the scenario of static, normal form games. Our analysis could be seen as part of a dynamic scenario, where the players in each period – or at each information set – update their subjective beliefs and play, for example, a Mirage Equilibrium strategy according to their posteriors. The real obstacle in extending our analysis to consider dynamic phenomena derives from the difficulty in characterizing the appropriate process for updating beliefs.

Battigalli and Guaitoli (1988), Fudenberg and Levine (1993), Kalai and Lehrer (1993b) and Rubinstein and Wolinsky (1994) introduce different versions of Conjectural Equilibrium. This is a concept based on von Hayek's observation that a certain pattern of play may be stable (that is, a steady state of a learning process) just because the players lack the necessary information to realize that they are playing sub-optimally. Thus, in a Conjectural Equilibrium, subjective beliefs[16] are restricted endogenously: only those beliefs are allowed which are not contradicted by the resulting play. Obviously, in the class of one-shot simultaneous move games, which are the object of this study, these endogenous constraints are not binding.[17] However, in most applications it is likely that the players eventually get confronted with some evidence that is inconsistent with their subjective beliefs. Of course, one should expect them to change (update) their beliefs upon such an event. Since we were already taking into account all the consistent strategic belief systems to derive the set of Mirage equilibria, it follows that the exogenous beliefs will (also) have to be changed. But at which level? Should the first-order beliefs be modified? Or the last-order ones? Alternatively, should the magnitude of change be allocated among different levels of beliefs according to a more complex rule? Without a specific model at hand, we are unable to answer these questions. In fact, even in specific examples one is most likely forced to adopt some *ad hoc* rule.

References

Alpert, M. and Raiffa, H. (1982), A progress report on the training of probability assessors, in Kahnemann et al. (eds.) *Judgment under uncertainty: Heuristics and biases* Cambridge University Press.

[16]They only consider strategic beliefs.

[17]If one treats the case of repeated interaction instead (as in Rubinstein and Wolinsky (1994)), then one may assume that some aspects of the current play are observable.

Battigalli, P. and Guaitoli, D., (1988), Conjectural equilibria and rationalizability in a macroeconomic game with incomplete information, W.P. 1988-6, Bocconi University.

Bernheim, D. (1984), Rationalizable strategic behavior, *Econometrica* 52: 1007-1028.

Brandenburger, A. and Dekel, E. (1993), Hierarchies of beliefs and common knowledge, *Journal of Economic Theory*, 59: 189-198.

Camerer, C. and Weigelt, K. (1991), Information mirages in experimental asset markets, *Journal of Business* 64: 463-493.

Fudenberg, D. and Levine, D. (1993), Self-confirming equilibrium, *Econometrica* 61: 523-546.

Greenberg, J. (1996), Towering over babel: worlds apart – but acting together, mimeo, McGill University.

Harsányi, J. (1967-68), Games with incomplete information played by 'Bayesian' Players, Parts I-III, *Management Science* 14: 159-182, 320-334, 486-502.

Kalai, E. and Lehrer, E. (1993a), Rational learning leads to Nash Equilibrium, *Econometrica* 61: 1019-1045.

Kalai, E. and Lehrer, E. (1993b), Subjective equilibrium in repeated games, *Econometrica* 61: 1231-1240.

Merten, J.-F. and Zamir, S. (1985), Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory* 14: 1-29.

Monderer, D. and Samet, D. (1989), Approximating common knowledge with common beliefs, *Games and Economic Behavior* 1: 170-190.

Morris, S., Postlewaite, A. and Shin, H. (1995), Depth of knowledge and the effect of higher-order uncertainty, *Economic Theory*, 453-467.

Morris, S., Rob, R. and Shin, H. (1995), p-Dominance and belief potential, *Econometrica*, 63: 145-157.

Nagel, R. (1995), Experimental results on interactive competitive guessing, *American Economic Review* 85: 1313-1326.

Offerman, T., Sonnemans, J. and Schram, A. (1996), Value orientations, expectations and voluntary contributions in public goods, *Economic Journal* 106: 817-845.

Pearce, D. (1984), Rationalizable strategic behavior and the problem of perfection, *Econometrica* 52: 1029-1050.

Ross, L., Greene, D. and House, P. (1977), The 'false consensus' effect: An egocentric bias on social perception and attribution, *Journal of Experimental Social Psychology* 13: 279-301.

Rubinstein, A. (1989), The electronic mail game: strategic behavior under 'Almost Common Knowledge', *American Economic Review* 79: 385-391.

Rubinstein, A. and Wolinsky, A. (1994), Rationalizable conjectural equilibrium: Between Nash and rationalizability, *Games and Economic Behavior* 6: 299-311.

Schmalensee, R. (1976), An experimental study of expectation formation, *Econometrica* 44: 17-41.

Stahl, D. (1993), Evolution of Smart$_n$ players, *Games and Economic Behavior* 5: 604-617.

Stahl, D. and Wilson, P. (1995), On players' models of other players: Theory and experimental evidence, *Games and Economic Behavior* 10: 218-254.

Tan, T. and Werlang, S. (1998), The Bayesian foundations of solution concepts of games, *Journal of Economic Theory* 45: 370-391.