



**HAL**  
open science

# Bayesian Game Theorists and non-Bayesian Players

Guilhem Lecouteux

► **To cite this version:**

| Guilhem Lecouteux. Bayesian Game Theorists and non-Bayesian Players. 2018. halshs-01633126v2

**HAL Id: halshs-01633126**

**<https://shs.hal.science/halshs-01633126v2>**

Preprint submitted on 12 Jul 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ  
CÔTE D'AZUR



Université  
Nice  
Sophia Antipolis



# BAYESIAN GAME THEORISTS AND NON-BAYESIAN PLAYERS

***Documents de travail GREDEG***  
***GREDEG Working Papers Series***

**GUILHEM LECOUTEUX**

**GREDEG WP No. 2017-30**

<https://ideas.repec.org/s/gre/wpaper.html>

Les opinions exprimées dans la série des **Documents de travail GREDEG** sont celles des auteurs et ne reflètent pas nécessairement celles de l'institution. Les documents n'ont pas été soumis à un rapport formel et sont donc inclus dans cette série pour obtenir des commentaires et encourager la discussion. Les droits sur les documents appartiennent aux auteurs.

*The views expressed in the GREDEG Working Paper Series are those of the author(s) and do not necessarily reflect those of the institution. The Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate. Copyright belongs to the author(s).*

# Bayesian Game Theorists and Non-Bayesian Players\*

Guilhem Lecouteux

*GREDEG Working Paper No. 2017-30*

Revised version

July 2018

Abstract (100 words): Bayesian game theorists claim to represent players as Bayesian rational agents, maximising their expected utility given their beliefs about the choices of other players. I argue that this narrative is inconsistent with the formal structure of Bayesian game theory. This is because (i) the assumption of common belief in rationality is equivalent to equilibrium play, as in classical game theory, and (ii) the players' prior beliefs are a mere mathematical artefact and not actual beliefs held by the players. Bayesian game theory is thus a Bayesian representation of the choice of players who are committed to play equilibrium strategy profiles.

Keywords: Bayesianism, common belief in rationality, epistemic game theory, interactive epistemology, prior beliefs.

Subject classification codes: B21, C72, D81.

Affiliation: Université Côte d'Azur, CNRS, GREDEG, France. Postal address: CNRS – GREDEG, Campus Azur; 250 rue Albert Einstein, CS 10269, 06905 Sophia Antipolis Cedex, France. Telephone number: +33 (0)4 93 95 43 74. Email: [guilhem.lecouteux@unice.fr](mailto:guilhem.lecouteux@unice.fr) ORCID: 0000-0001-6602-7247

\*Several ideas presented in this paper originated from an ongoing collaboration with Lauren Larrouy on coordination in game theory. I would like to thank her for various discussions on this subject. I am also grateful to participants to the 21<sup>st</sup> ESHET annual conference, and to two anonymous reviewers for their comments and suggestions.

*'There is no way that a book on game theory can be written if Bayesian rationality cannot be assumed' (Binmore 1992, p.119)*

Most of modern game theory assumes that players are Bayes rational agents, maximising their expected utility given their beliefs about the strategies of others. Luce and Raiffa already suggested in the 1950s that in a (two-person non-zero-sum) non-cooperative game:

One *modus operandi* for the decision maker is to generate an *a priori* probability distribution over the [set of pure strategies] of his adversary [...] and to choose an act which is best against this *a priori* distribution (Luce and Raiffa, 1957, p.306)

It is however only in the 1980s that the idea started to burgeon, with the emergence of a literature investigating the epistemic foundations of solution concepts (see Hillas and Kohlberg (2002) for a review) with, among many others, Bernheim (1984, 1986), Pearce (1984), Aumann (1987), Brandenburger and Dekel (1987b), Tan and Werlang (1988). These works attempted to build an alternative to classical game theory [henceforth CGT], which assumes that 'rational' players ought to play Nash strategies. CGT failed to offer a general theory of non-cooperative games because of (i) the problem of multiplicity of equilibria, and (ii) the lack of a convincing justification for equilibrium play. Rather than trying to solve games by identifying equilibrium profiles, an alternative was to consider the problem faced by the players, and to study the beliefs that the players may hold in a strategic context (see e.g. Bernheim 1984, pp.1007-1010).

The objective of *epistemic game theory* [henceforth EGT] is to make Bayesian decision theory the primitive of strategic behaviours, and to model explicitly the knowledge and beliefs of the players regarding each other's rationality and choices:<sup>1</sup>

1 Throughout the rest of this paper, I will use the terms 'Bayesian' and 'epistemic' game theory interchangeably.

Here is a general definition of the field today (for which I am grateful to Sergei Artemov): *EGT makes epistemic states of players an input of a game and devises solution concepts that takes epistemics into account* (Brandenburger, 2010, p.6, emphasis in original)

Game theory is considered by many scholars, such as Harsanyi (1977, p.47), Myerson (1990, p.5), or Aumann, as a direct extension of individual decision theory. In an interview on epistemic logic, Aumann for instance said about a joint work with Drèze (Aumann and Drèze, 2008) on rational expectations in games:

In this paper with Jacques what we're really trying to promote is a revolution. We haven't been too successful yet. We're saying: Listen! Equilibrium is not the way to look at games. Nash equilibrium is *king* in game theory. Absolutely king. We say: No! Nash equilibrium is an interesting concept, it's an important concept, but it's not the most basic concept. The most basic concept should be: *to maximise your utility given your information*, in a game just like in any other situation. (Aumann 2010, p.29, emphasis in original)

Giocoli (2009) argues that the boom of game theory in the late 1970s and 1980s, which contributed to the game-theoretic revolution of neoclassical economics – compared to its limited impact on economic theory in the 50s and 60s (Giocoli 2003) – precisely lies in the extension of non-cooperative game theory to games of incomplete information with the birth of Bayesian game theory.<sup>2</sup> Harsanyi's (1967-68) analysis of games of incomplete information thus played a key role in the success of game theory, by offering a general framework to study games where players have different information, which gave rise to the modern economics of information (Myerson 2004, p.18).

2 According to Giocoli, this evolution results from the combination of Harsanyi's influence on the development of the refinement literature, on models of imperfect competition (in particular as a reaction against the Chicago antitrust arguments in US Courts), and on mechanism design theory (see Schotter and Schwödiauer (1980) for a literature review of the period).

The object of this paper is to highlight that there is actually little Bayesianism in Bayesian game theory, which remains essentially faithful to the tradition of CGT. This is because Bayesian game theory intends to study the strategic interactions of rational players *when their rationality is common belief*. I will argue that the research program in EGT is a program in *interactive epistemology* (i.e. in applied mathematics rather than economics), that focuses on the definition of systems of beliefs consistent with the assumption of common belief in rationality [henceforth CBR]. The players' prior beliefs – the central element of the model – are however a mere mathematical artefact designed by the modeller, and not an actual belief held by the players. This ambiguity can be traced back to Harsanyi's type-based approach, which constitutes the core of his 1967-68 paper: the types of the players are indeed a way to *encode* the infinite hierarchies of beliefs of the players, and to analyse incomplete information games with the standard tools of CGT. Furthermore, there is no explanation about how (even highly rational) players could *form* such consistent beliefs, and how they could discriminate between different consistent beliefs. This means that EGT faces the exact same problems that it was supposed to solve in CGT, namely (i) to provide a rationale for equilibrium play, and (ii) to offer a theory of equilibrium selection.

The idea of CBR seems to be implicit in Harsanyi (1967-68), and is a central point of the 1982 discussion between Harsanyi and Kadane and Larkey in *Management Science*. It then constituted the starting point of Bernheim and Pearce's works on rationalisability, and of the subsequent literature in epistemic game theory. The tight connection between CBR and equilibrium behaviour – established by Brandenburger and Dekel (1987b) and Aumann (1987), but whose intuition was already given by Luce and Raiffa (1957, p.306) – is however problematic, since it reduces Bayesianism to a tautological representation of strategic behaviours with *ad hoc* priors. This raises a serious issue for the proponents of EGT: contrary to Brandenburger's definition, the prior beliefs are *outputs* of the reasoning process of the players (they indeed represent the equilibrium that will be implemented), and cannot therefore be used as inputs of this very same reasoning process. This difficulty results from the ambiguous status of 'prior beliefs' in Bayesian game theory, which can either be given a substantive, 'economic', content (e.g.

there *is* an actual prior stage from which the players can update their beliefs) or be considered as a purely formal device (Gul 1998).

The rest of the paper is organised as follows. Section 1 presents the early developments of Bayesian game theory with Harsanyi's initial work on games of incomplete information, and its progressive extension to the case of strategic uncertainty. Section 2 then presents the 1982 debate in *Management Science* between Kadane and Larkey on the one hand and Harsanyi on the other hand about the proper object of game theory, and especially the role of normative game theory. Section 3 discusses the emergence of modern Bayesian game theory in the 1980s and the equivalence between CBR and equilibrium play. Section 4 then highlights the interpretational difficulties of EGT, and the incoherence between the narrative of the theory and its formal structure. Section 5 concludes by discussing the possibility (and probably the necessity) of developing game theoretic models without assuming CBR.

## **1. Early Bayesian game theory**

Games with incomplete information correspond to strategic interactions in which some or all of the players lack information about the game structure, such as the payoff function or the set of strategies of the other players<sup>3</sup> (or even their *own* payoff function or set of strategies). The study of such games has been problematic for many years, because it requires introducing the beliefs of the players about each relevant aspect of the game, with the risk of developing intractable models. Luce and Adams (1956), in their analysis of bargaining games with misperceived payoff functions, for instance confess that they 'do not have the slightest idea how one should deal with such a profusion of [beliefs about the utility of other players]', and 'are led to search for a condition which will collapse them into a single one' (p.168). An additional difficulty is that we should model the

<sup>3</sup> Games of *imperfect* information, on the other hand, correspond to games of complete information in which players are unaware of the actions chosen by others, for instance in sequential games.

beliefs of the players *about each other's beliefs*. This problem was explicitly raised by Harsanyi (1962) in the context of bargaining games with unknown payoff functions:

In bargaining, and more generally in all non-trivial game situations, the behavior of a rational individual will depend on what he expects the *other* party will do. Party 1 will ask for the best terms he expects *party 2* to accept. But party 1 will know that the terms party 2 will accept in turn depend on what terms party 2 expects *party 1* to accept. Thus, party 1's behavior will depend on what may be called his *second-order* expectations, i.e., on party 1's expectations concerning party 2's *expectations* about party 1's behavior. These again will depend on party 1's *third-order* expectations, i.e., on his expectations concerning party 2's *second-order* expectations, etc. (Harsanyi 1962, p.29, emphasis in original)

Harsanyi offered a solution to deal with such infinite regresses in his three-part paper on 'Games with Incomplete Information Played by "Bayesian" Players' (1967-68). Consider a player  $j$  in a game where the sets of available strategies are known, but not the payoff functions.<sup>4</sup> We define a vector  $a_j = \{a_{jk}\}_{k \in N}$ , with  $a_{jk}$  the information that player  $j$  has about any relevant parameter of player  $k$  payoff function. The vector  $a_j$  thus represents everything that player  $j$  knows that could affect the payoff function of any of the  $n$  players. We also define another variable  $b_j$  that denotes all the information that player  $j$  has that could affect her beliefs about what the others know.<sup>5</sup> The vector  $c_j = \{a_j, b_j\}$  denotes all

<sup>4</sup> Harsanyi argues in Part I that an uncertainty about the set of strategies can be reduced to a strategy about the payoff function. If a strategy  $x_i$  is not available to player  $i$ , then it is as if  $i$  always receives an extremely low payoff when she chooses this strategy (we therefore know that  $i$  will never choose this strategy if it were available).

<sup>5</sup> Note that we assume that *all* the relevant parameters of the model can be summarised in a single variable, without explicitly defining those parameters. Mertens and Zamir (1985) will give a precise mathematical definition of this idea. They offer in particular a proof of the existence of a space of infinite hierarchies of beliefs.



the private information that player  $j$  has about the game, and is called player  $j$ 's *type*.<sup>6</sup> We can notice that a type  $c_j$  specifies a belief about the other's payoff function, but also about other higher-order beliefs, and hence specifies a belief about the other players' types. Harsanyi's idea is therefore to specify a number of types for each player, and to specify for each type a payoff function and a probabilistic belief about the others' types. With this approach, we can *derive* from each type the infinite belief hierarchy it induces, and thus study incomplete information games without *explicitly* writing the infinite beliefs hierarchies of the players.

Harsanyi's *type-based approach* enabled the development of a large literature investigating (repeated) games of incomplete information in the 1970s (e.g. Mertens and Zamir 1971; Zamir 1971; Mertens 1971; Harsanyi and Selten 1972; Ponsard and Zamir 1973; Kohlberg and Zamir 1974; Ponsard 1975a,b; Kohlberg 1975a,b; Sorin 1979). It is however worth noticing that those works – just as Harsanyi's 1967-68 paper – are not 'Bayesian' in the sense of modern Bayesian game theory, since the uncertainty is only about players' payoffs or set of strategies, and never about *the choices themselves*. Harsanyi indeed developed a solution concept for Bayesian games (Part II of the 1967-68 paper), which is formally equivalent to the Nash equilibrium of an auxiliary game of complete information. Contrary to the objective of EGT to challenge the classical approach of studying solution concepts, and to treat explicitly the players' ability to reason about the other players (Perea 2012, p.2), Harsanyi did not explain *why* it is rational for his 'Bayesian' players to play a Bayesian equilibrium. The explanation is essentially the same than in CGT, i.e. rational players *ought* to play a Nash equilibrium.

Harsanyi (1975) extended his framework to allow for an uncertainty over the strategy choices of the other players: he defined how players should adapt their beliefs, until they reach an equilibrium point of the game (see Skyrms 1990, for a detailed description of this process). The question of where the priors come from was then treated

<sup>6</sup> Harsanyi originally used the terms 'information vector', 'attribute vector' and 'type'. Following Aumann and Maschler's (1967, 1968 [reprinted in Aumann and Maschler 1995]) suggestion, Harsanyi and later authors eventually adopted the term 'type' (see Myerson, 2004, p.1820).

with Selten in their *General Theory of Equilibrium Selection in Games* (1988). Harsanyi recognised that the

resolution of this problem will be a very essential part of our whole theory of solutions for  $n$ -person noncooperative games, because in general, the tracing procedure to be described will select different equilibrium points  $s^*$  as solutions for any given game  $\Gamma$ , depending on the  $n$ -tuple  $p$  of prior probability distributions used as a starting point (Harsanyi 1975, p.63)

Priors in Harsanyi's model are thus built such that the outcome of the tracing procedure (i.e. the revisions of players' priors) is a solution of the game. The 'Bayesian' dimension of the theory is therefore largely artificial, since what makes a strategy rational is not that it maximises one's subjective expected utility, but that it belongs to an equilibrium profile. This is indeed precisely the argument Harsanyi and Selten use against a 'naive Bayesian approach' to game theory, according to which each player should simply choose the strategy that maximises her expected payoff, given her beliefs about the choice of others:

Unfortunately this simple theory will not work because this best reply combination  $q^0$  will generally not be an equilibrium point of the game, and therefore it cannot be the outcome chosen by a rational outcome-selection theory (Harsanyi and Selten 1988, p.140)

According to Harsanyi and Selten, subjective expected utility theory (henceforth 'SEUT') is a rational decision theory only if the resulting choice is an equilibrium of the game. The Bayesian dimension of the game is therefore trivialised: a major critique against CGT is indeed that there is no explanation of why a player should play an equilibrium strategy. Harsanyi and Selten give an explanation, but leave unsolved the complementary issue of why a player's prior should correspond to the equilibrium profile. Early Bayesian game theory (directly derived from Harsanyi's initial work on games of incomplete information) seems therefore fundamentally classical. The main inquiry of the theoretician is to find solutions of the game, and then to rationalise them *a posteriori* by assuming that 'rational' priors should correspond to the equilibrium.

## 2. Kadane and Larkey's critique

Kadane and Larkey (1982a, 1983; henceforth 'KL') are among the first authors to openly argue for a subjective view of probability in game theory. They suggest that Harsanyi's approach adopts a necessitarian rather than subjectivist view of probability (1982a, p.115), and are sceptical with the project of treating solution concepts as the basis of prior beliefs in non-zero-sum games<sup>7</sup> (1982a, p.116). KL address themselves to game theorists as 'individuals interested in the development of useful prescriptive and predictive theories of individual and collective decision-making behavior' (Kadane and Larkey 1982a, pp.114-115). They are 'perplexed by the emphases in game theoretic researches' and 'do not understand the search for solution concepts that do not depend on the beliefs of each player about the others' likely actions' (p.115). They emphasise that no principle of rationality mandates a Bayes rational player to assume that the other players are rational: it is therefore not necessarily rational for her to play an equilibrium strategy (such as a minimax strategy in a zero-sum game (pp.115-116)). They therefore argue that – rather than focusing on solution concepts – game theorists should 'look to other disciplines such as cognitive psychology for predictive theories of decisional behavior' (p.118) and give an *empirical* justification to the priors held by the players.

Harsanyi (1982a) vigorously criticises KL position, on the ground that they missed the whole point of game theory. The central question is indeed

how [the probability distribution over the strategy combinations available to the other players] is to be chosen by a rational player who expects the *other players to act rationally*, and also expects these other players to entertain similar expectations about him and about each other (Harsanyi 1982a, p.121, emphasis in original)

<sup>7</sup> Spohn (1982) raises a similar point, and argues that Harsanyi's idea of 'Bayesian' game theory is 'still more game theoretic than decision theoretic in spirit and hence criticisable on similar grounds as the standard accounts' (Spohn 1982, p.240).

Even though the idea is only implicit, Harsanyi seems to have in mind a condition of *common belief in rationality*: all the players are rational, they all believe that the others are rational, they all believe that they all believe that the others are rational, and so on.<sup>8</sup> KL approach would on the other hand ‘trivialise game theory by depriving it of its most interesting problem’ (p.121) and only offer ‘the highly uninformative statement that [...] each player should try to maximize his expected payoffs in terms of his subjective probabilities’ (p.122).

A central point of their disagreement is the role of normative game theory in a theory of games: KL agree that such a theory would be useful if we have good reasons to believe that the other players are rational. However:

in deciding on the best strategy against an actually or potentially irrational opponent or opponents, normative game theory can provide only indirect help. Rather, what we need is an empirically supported *psychological* theory making at least probabilistic predictions about the strategies people are likely to use, ... given the nature of the game and their own psychological makeup. If we had such a theory, deciding on their best strategy against such an opponent ... would ... involve ... a solution of a simple maximization ... problem (Kadane and Larkey 1982b, p.124, quoting Harsanyi 1982a, p.122)

Harsanyi recognises that ‘*normative* game theory and *psychological* – explanatory and predictive – theories of actual game-playing behavior are very different intellectual enterprises, using very different methodologies as a matter of logical necessity’ (Harsanyi

<sup>8</sup> The condition stated in this extract only implies a 2<sup>nd</sup>-order belief in rationality (we all believe that we all believe that we are rational). Harsanyi seems however to interpret the condition of mutually expected rationality as common belief in rationality: this is clearly the case in his defence of the rationality postulates of game theory (Harsanyi 1961, p.180), where he argues that the postulates of ‘utility-maximisation’ and ‘mutually expected rationality’ implies the iterated elimination of dominated strategies (which however requires CBR).

1982a, p.122, emphasis in original). This is why KL, in their reply to Harsanyi's comment, suggest that 'our differences with Professor Harsanyi are not as profound as might appear' (Kadane and Larkey 1982b, p.124). Harsanyi however resists reconciliation, and suggests that normative game theory is not only useful for prescribing choices against rational agents, but also offers a *benchmark* for actual behaviours. This point is central in Harsanyi's (1977) defence of rational choice models:

the heuristic importance of rational behavior theories is based on the fact that in most cases we cannot really understand and explain a person's behavior [...] unless we can interpret it *either as rational behavior* in this particular situation *or as an understandable deviation* from rational behavior. For example, we may be able to interpret it as an understandable mistake, as an understandable emotional reaction, or as an understandable intentionally suboptimal response (Harsanyi 1977, p.18)

The disagreement between Harsanyi and KL is about the *empirical* validity of normative game theory. Even if players do not behave as prescribed by CGT, the question is whether their behaviour can be analysed as a psychological 'deviation' from a latent rational 'benchmark' – in which case normative game theory could offer the basis of a descriptive theory of games. It is worth noting that Harsanyi's interpretation of rational choice has been quite pervasive in economics, and constitutes nowadays the mainstream approach to *new* behavioural economics (Sent 2004). An economic agent is represented as an 'inner rational agent [...] trapped inside a psychological shell' (Infante *et al.* 2016, p.1), and its psychology is being treated as a set of forces that *interferes* with its latent rationality. However, the latent rationality of the inner agent is not given any psychological foundations (Sugden 2015), which jeopardises the project of using rational choice as a benchmark for an empirically relevant theory of games. In the absence of a psychologically valid interpretation of this model, KL are right in questioning the usefulness of normative game theory for practical purposes. They later argue (Kadane and Larkey 1983) that the absence of a clear validation method for game-theoretic results contributes to a general confusion about the purposes of game theory. Harsanyi's empirical defence of normative game theory would for instance require testing

experimentally the theory, but he merely advances a pragmatic argument, i.e. the theory ‘*does the job*’ (Harsanyi 1982a, p.121, emphasis in original). This is because the research standards in game theory are those of mathematics research (Kadane and Larkey 1983, p.1377), i.e. game theorists are looking for new and logically correct results, without bothering about possible applications.

Harsanyi’s main argument against KL is thus that game theory is interested in the behaviours of ‘highly rational opponents’. This may however not be the initial intention of Von Neumann and Morgenstern. Harsanyi indeed suggests that ‘Von Neumann’s [maximin] argument is restricted to the case where the two players *expect each other to act rationally*’ (p.122, emphasis in original), while Giocoli (2003) argues that Von Neumann and Morgenstern objective was:

to write down the “handbook for the good player”, rather than to analyse reality under the assumption that the players knew and abided by the handbook’s rules (Giocoli 2003, p.211)

This is the reason why:

what is looked for in the TGEB is a characterization of rationality capable of giving the player the *full control* over the outcome of the strategic situation. Hence, to be rational, that is, to “play well”, in vN/M’s sense means precisely to exploit the payoff structure of the game to devise a strategy that allows the player to get the maximum among the returns that can be obtained *independently of the rival’s behavior* (Giocoli 2003, p.265, emphasis in original)

If Giocoli’s interpretation is right, then – although Von Neumann and Morgenstern would have disagreed with the definition of subjective beliefs over strategy choices – it might be Harsanyi rather than KL who misinterprets the contribution of game theory for economics: game theory should indeed be a prescriptive theory without any reference to the other’s rationality. KL (1983, p. 1377) seems indeed to look for the kind of ‘handbook for the good player’ that Von Neumann and Morgenstern intended to write.

### 3. Rational priors and common belief in rationality

If game theory ought to describe the choices of players who hold ‘rational’ beliefs, game theorists should study how players could form such rational beliefs. Harsanyi’s suggestion seems unsatisfactory, since the priors must be defined from the equilibrium: this means that the beliefs that the players are supposed to use when maximising their expected utility (i.e. the inputs of their deliberation) precisely correspond to the strategy profile they will implement (i.e. the output of their deliberation). Such players are therefore Bayesian in a tautological way, and deliberation is vacuous (see Levi 1998, p.180 for a similar argument).

As mentioned above, a central condition for determining rational beliefs is that players mutually expect each other to be rational. Believing in each other’s rationality indeed implies that I cannot believe that you could play a non-rational action. But if I also believe that you believe that I cannot play a non-rational action, I should also revise my beliefs about your beliefs about my actions, and thus revise what would be a rational action with those new beliefs. Luce and Raiffa already described the kind of belief revision that would result from CBR:

To determine such a subjective *a priori* distribution, the decision maker might imagine a series of simple hypothetical side bets whose payoffs depend upon the strategy his adversary employs. This is easier said than done, however, since the decision maker cannot ignore the possibility that his adversary will attempt to hypothecate such a procedure for him and will adjust his choice of strategy accordingly. [...] one can argue that the decision maker should keep on modifying the *a priori* distribution until this alleged indirect feedback no longer produces any change – until there exists an equilibrium in the decision’s maker mind. (Luce and Raiffa, 1957, p.306)

The first mathematical analysis of the concept of common knowledge can be found in Friedell (1969), but it is usually Lewis (1969) who is credited by game theorists for formulating the concept of common knowledge. While they do not quote each other’s

work, they also both offer a possible explanation of the emergence of common knowledge (common ‘opinion’ for Friedell) based on a form of simulation routine<sup>9</sup> (Friedell 1969, 31; Lewis 1969, figure 26). In the folk history of game theory, Aumann is nonetheless credited as the first author to formalise *rigorously* the concept of common knowledge.<sup>10</sup>

Aumann (1974) initially developed the notion of correlated equilibrium as a generalisation of mixed Nash equilibrium, in which randomisations need not be independent (both players can for instance randomise their strategies based on the *same* coin toss). He suggested the possibility that players hold *subjective* beliefs, i.e. that they should not necessarily hold the same beliefs about the probability of occurrence of the public event. Aumann then started to investigate under which conditions players could maintain different priors, although they should be led to update their priors once they know that the other has different beliefs (see Aumann’s (2005, pp.697-698) interview by Hart). The *agreement theorem* (Aumann 1976) states that, if two individuals have common priors, and that their posteriors about an event are common knowledge (we both know what probability we assign to the event, we both know that we know it, etc.), then those beliefs must be the same: they cannot agree to disagree.<sup>11</sup>

The agreement theorem pioneered the field of *interactive epistemology* (see Battigalli and Bonanno (1999) and Brandenburger (2007) for reviews), defined by Aumann as follows:

<sup>9</sup> ‘Simulation’ refers to a specific theory of *mindreading*, according to which the individual forms her beliefs about the other’s actions by imagining what she would do if she were in the others’ shoes. See Guala (2016, pp.89-101) and Larrouy and Lecouteux (2017) for a game-theoretic analysis of simulation thinking.

<sup>10</sup> See Vanderschraaf and Sillari (2014) for a history of the concept and a review of the different formalisations.

<sup>11</sup> The basic intuition of the theorem is the following. If we share a common prior and receive some private information, then our posteriors may differ if the information was different. However, if we both know our ‘initial’ posteriors, then we must revise them to take into account the fact that the other seems to have received an information that led her to revise her beliefs in a different way. The theorem states that this process of belief revision will necessarily lead the players to hold the same posterior belief.



Formal Interactive Epistemology deals with the logic of knowledge and belief when there is more than one agent or “player”. One is interested not only in each person’s knowledge about substantive matters, but also in his knowledge about the others’ knowledge (Aumann, 1999, p.263)

An important contribution of the 1976 paper was the *semantic* definition of common knowledge in terms of ‘partition structure’, i.e. ‘information sets’ defined as partition of the set of ‘states of the world’. This contrasts with the *syntactic* formalism used by Lewis and Friedell, based on logical operators such as ‘and’, ‘or’, and on knowledge operators (see Aumann 1979a, p.264; see also Cubitt and Sugden (2003) on the distinctive features of Lewis approach). The *state-based approach* works as follows. A state of the world  $\omega$  describes for each player the set of states of affairs he ‘deems possible’.<sup>12</sup> A state of the world is ‘all-inclusive’, since it describes what the players know and believe about the world, their knowledge and beliefs about the other players, as well as their own action in that state – all the relevant elements about the state of the world are included in its description. If every player is informed about the ‘deeming possible’ operator for each of the players, then we can derive – just as with Harsanyi’s type-based approach – the infinite belief hierarchies of the players. At the state of the world  $\omega$ , we can indeed derive the different states the player deems possible, what she deems possible about the sets of states that the other players deem possible, and so on. Aumann’s state-based approach can be seen as ‘a *restatement* of Harsanyi structure’ (Aumann 2010, p.24, emphasis in original). This equivalence was formally stated by Brandenburger and Dekel (1993), who show that we can always transform one model into another, i.e. that a type can be described within the framework of the state-based approach, while the belief hierarchies induced in the state-based model can be encoded as types in the type-based model.

12 The ‘deeming possible’ operator was introduced by Kripke (1963), and is slightly more general than Aumann’s model. Samet (1990) characterises the restrictions that should be put on the ‘deeming possible’ operator to obtain Aumann’s model. See also Bacharach (1985).

It is in this context that game theorists started to investigate the implications of common belief<sup>13</sup> in rationality for game theory (e.g. Milgrom 1981, Brandenburger and Dekel 1987a, Monderer and Samet 1989, Rubinstein 1989), and notably the definition of ‘admissible’ strategies under the assumption of CBR. This is precisely the object of Bernheim (1984) and Pearce (1984) papers on *rationalisability*.<sup>14</sup> They both consider the strategies of other players as uncertain events, and (unlike Harsanyi) are very critical of Nash equilibrium. According to Bernheim:

when we think in terms of maximising utility subject to expectations rather than realizations, it becomes clear that the Nash hypothesis [that players expect that the Nash outcome will occur], far from being a consequence of rationality, arises from certain restrictions on agents’ expectations which may or may not be plausible [...] We are then quite naturally led to ask: are there any restrictions of individuals’ expectations (and hence choices) which are required by rationality alone, rather than by (subjective) plausibility? (Bernheim 1984, p.1007)

Bernheim and Pearce then both explicitly endorse SEUT to model the choices of the players, by adding the additional condition that the prior beliefs of the players must be consistent with the *common knowledge* of the rationality of the players:

<sup>13</sup> I will not distinguish between common belief and common knowledge in this paper: the main reason is that the notion of common knowledge introduced in most of the papers that will be discussed below correspond to a notion of common belief with probability 1.

<sup>14</sup> Werner Böge and some of his students in mathematics at Heidelberg developed in the 1970s a formal model of Bayesian games (Böge 1974, Armbruster and Böge 1979, Böge and Eisele 1979). They seem to be the first ones to formally define the notion of common belief in rationality (although they did not use the term, see Perea 2014, pp.14-15). Armbruster and Böge (1979) and Böge and Eisele (1979) propose a recursive procedure identifying all the choices that rational players could make under an assumption of common belief in rationality, which is essentially the same than in Pearce (1984) – neither Bernheim nor Pearce however refer to those papers.

For strategic games in normal form, it is natural to proceed on the basis of two premises: (1) agents view their opponents' choices as uncertain events, and (2) all agents abide by Savage's axioms of individual rationality, and this fact is common knowledge (Bernheim 1984, p.1011)

Assumption (A1): When a player lacks an objective probability distribution over another player's choice of strategy, he forms a subjective prior that does not contradict any of the information at his disposal.

Assumption (A2): Each player maximises his expected utility relative to his subjective priors regarding the strategic choices of others.

Assumption (A3): The structure of the game (including all participants' strategies and payoffs, and the fact that each player satisfies Assumptions (A1) and (A2)) is common knowledge (Pearce 1984, pp.1031-1032).

Bernheim calls a choice *rationalisable* if there is a 'consistent system of beliefs' such that this choice is an optimal strategy. More specifically, a strategy of  $i$  is rationalisable if it is 'justifiable' (i.e. it is optimal given some belief over the strategies of  $j$ ), that all the strategies that  $i$  believes  $j$  could play are justifiable, that all the strategies that  $i$  believes  $j$  believes  $i$  could play are justifiable, and so on. Pearce on the other hand calls a choice rationalisable if and only if it survives a specific recursive elimination procedure.<sup>15</sup> Benheim (1986) and Tan and Werlang (1988) later showed the formal equivalence between rationalizability and common belief in rationality: if there is common belief in the rationality of the players, then the players can only play rationalisable strategies, and they must believe that the other players play rationalisable strategies.

Brandenburger and Dekel (1987b) however notes that 'rationalisability is more closely related to an equilibrium approach than one might think at first' (p.1392). They

<sup>15</sup> In the first step, we start with the whole set of strategies for every player  $i$ . In the second step, we only keep the strategies  $s_i$  for which there exists a belief about the choice of  $j$  such that  $s_i$  is optimal. We continue the same process with the remaining strategies. Bernheim and Pearce's approaches lead to the exact same set of rationalisable strategies.

indeed establish an equivalence between rationalisability<sup>16</sup> and the solution concept of *a posteriori equilibrium* (Aumann 1974): the direct implication is that common belief in rationality is equivalent to equilibrium behaviour. The improvement upon Harsanyi's initial approach is that the resulting equilibrium is not necessarily a Nash equilibrium: if the players share common priors and their rationality is common knowledge, then the resulting equilibrium is a correlated equilibrium (Aumann 1987), while Nash requires an additional condition of 'informational independence' (see Brandenburger and Dekel 1987b, pp.1400-1401). Brandenburger and Dekel (1987b) conclude that their analysis 'provides a formal decision-theoretic justification for using equilibrium concepts in game theory' (p.1392).

More recently, Aumann and Drèze (2008) refined the analysis of consistent beliefs under CBR by analysing beliefs at the *interim* rather than *prior* stage (i.e. once the players received some private information about their type). Their objective is to characterise the value of an arbitrary *n*-person game (i.e. the payoff that the player should expect), under the assumptions of CKR and common priors. One of their main results (Theorem B) is that 'the rational expectations in a game *G* are precisely the conditional payoffs to correlated equilibria in the doubled game  $2G$ ' – meaning, in terms of strategies, that CKR and common priors lead to a correlated equilibrium in the doubled game  $2G$ .<sup>17</sup>

Interestingly, Aumann's papers directly echo Harsanyi's reply to Kadane and Larkey. Although Aumann does not quote Harsanyi (1982a), he refers to KL critique in

<sup>16</sup> Their paper distinguishes between 'independent' rationalisability (as in Bernheim and Pearce, since the priors are supposed to be distributed independently) and 'correlated' rationalisability (which was a case considered by Böge, though his 1979 papers are not mentioned by Brandenburger and Dekel). Correlated rationalisability is equivalent to *a posteriori* equilibrium, while they introduce the refinement of conditionally independent *a posteriori* equilibrium for the equivalence with independent rationalisability.

<sup>17</sup> The doubled game  $2G$  is defined as the game in which every strategy is available twice (doubling strategies can have a significant impact when priors are correlated, because it allows more subtle beliefs). The difference with Aumann's 1987 paper is that rational expectations are defined at the *interim* stage rather than the prior stage (see Aumann and Drèze, 2008, p.80). This is why the resulting equilibrium is slightly different between the two papers.

the 1987 paper (Aumann 1987, p.17), as well as in two papers with Drèze (2008, p.81 on rational expectations, and 2009, p.14 on the definition of subjective probability in games).

The theory presented in this paper differs from much of the strategic game literature in that it does not deal with equilibria; rather, it recommends that a player in a game, like in one-person decision problem, should simply maximize against her subjective probabilities for the other players' actions. [...] Joseph B. Kadane and Patrick D. Larkey (1982) expressed a similar view, but unlike Luce and Raiffa (1957), they eschewed the “strategic aspects.” This ignores *the* fundamental insight of game theory, an insight that is captured by the idea of rational expectations introduced here: that a rational player must take into account that the players reason about each other in deciding how to play. (Aumann and Drèze, 2008, p.81, emphasis in original)

Although Aumann explicitly rejects the use of equilibrium solution concepts as the *basis* of prior probabilities, the equivalence between CBR and equilibrium play implies that his objection to KL is essentially the same as Harsanyi's: the object of game theory is the study of decision making when players have coherent beliefs, i.e. when their beliefs form an equilibrium. It is also worth noting that, in the two papers with Drèze (Aumann and Drèze 2008, 2009), KL suggestion of ‘[looking] to other disciplines such as cognitive psychology for predictive theories of decisional behavior’ (Kadane and Larkey 1982a, p.118) is contrasted with Luce and Raiffa's suggestion of assigning a subjective probability to the choice of others that takes into account the ‘strategic aspects’ of the game (1957, p.306). Aumann (2005, p.730) indeed acknowledged that Luce and Raiffa's *Games and Decisions* ‘had a big influence’ on him, and it may not be so surprising that his work on epistemic game theory followed Luce and Raiffa's intuition that, in a game, beliefs should be such that ‘there exists an *equilibrium* in the decision maker's mind’ (Luce and Raiffa 1957, p.306, emphasis added).

#### **4. Bayesian players or theorists?**

We have seen thus far that EGT intends to model the strategic interaction of Bayesian rational agents, when their rationality is common belief. The players have a prior belief

about the strategy choices of the others, but CBR restricts the set of beliefs from which this distribution can be drawn (e.g. I cannot believe that you cooperate in a Prisoner's Dilemma). This means that Bayes rationality plus CBR determine the set of priors I could hold. Since CBR implies equilibrium behaviour through the definition of *ad hoc* beliefs, equilibria are interpreted in terms of coherent beliefs:

The idea of common knowledge really enables the “right” formulation of correlated equilibrium. It's not some kind of esoteric extension of Nash equilibrium. Rather, it says that if people simply respond optimally to their information – and this is commonly known – then you get correlated equilibrium. The “equilibrium” part of this is not the point. Correlated equilibrium is nothing more than just common knowledge of rationality, together with common priors (Aumann 2005, p.698)

The two key ingredients of EGT are therefore (i) the existence of a prior distribution over the set of strategy profiles, and (ii) the compatibility of this distribution with CBR.<sup>18</sup> Given the prior distribution, we can derive the players' subjective beliefs about the strategy of the others,<sup>19</sup> and the players may then maximise their subjective expected

<sup>18</sup> Although CBR is routinely used by epistemic game theorists (and considered by Perea (2014, p.13) as its ‘central idea’), it is not constitutive of the epistemic program. Aumann and Brandenburger (1995) for instance show that mutual knowledge of rationality combined with common priors and common knowledge of players' conjectures imply that their beliefs form a Nash equilibrium. The important point here is that EGT looks for a combination of epistemic conditions such that a single ‘equilibrium’ of beliefs remain. CBR is the most efficient condition to generate such consistent beliefs, but other conditions can also work (common priors and the common knowledge of players' beliefs imply for instance the convergence to a single profile by Aumann's agreement theorem).

19 If we allow for correlated priors (as in Aumann 1987), then player *i*'s belief about player *j*'s action can be conditioned on player *i*'s own action. This does not however mean that beliefs are action-dependent (see Hammond, 2009), i.e. that *i* believes that his action could directly influence the action of the others, since the players do not genuinely ‘choose’ in Aumann's model – their action is indeed given by the state of the world.

utility. I will now argue that this approach – despite its ‘Bayesian’ flavour – is essentially the same than classical game theory, and still faces some of its main difficulties.

Recall that a major drawback of CGT is that the theory does not provide any clear-cut criterion to discriminate between different equilibria. Unfortunately, EGT also fails in providing an explanation for the selection of specific equilibria. Consider for instance the Hi-Lo game (see Bacharach, 2006, chapter 1 for a detailed discussion):

Hi-Lo	H	L
H	(10;10)	(0;0)
L	(0;0)	(5;5)

This coordination game is puzzling for CGT, since HH seems to be the obviously ‘right’ equilibrium, while LL is also a Nash equilibrium (and there is no straightforward epistemic requirement suggesting why a rational player should unambiguously play H). The problem of multiplicity of Nash equilibria (HH, LL, and in mixed strategies) is however not solved by EGT. The theory can only say that, under CBR, several consistent priors are possible, including the common belief that both players will play L. The only ‘solution’ that EGT can offer to escape the puzzle is to suggest that the players are more likely to hold the beliefs that HH will be implemented (but for reasons that are necessarily *external* to the model). By imposing CBR, we can rationalise the implementation of an equilibrium profile. CBR however can only select internally consistent beliefs (i.e. beliefs that do not contradict CBR), and does not offer a *constructive* method that could identify to which priors the players could converge.

The second fundamental issue of CGT – and one of the main motivations behind the development of EGT – is the inability of CGT to explain *why* players should play an equilibrium. What EGT achieved is to establish an equivalence between CBR and equilibrium play: it is therefore rational for a player to choose an equilibrium profile, because it maximises her subjective expected utility. The complementary issue (which was not solved by Harsanyi, see *supra*) to determine why a Bayes rational player would

form such beliefs is that they are the only beliefs *compatible* with CBR. It seems therefore that EGT, unlike classical game theory, succeeds in explaining why players should play equilibrium strategies (although this equilibrium need not be a Nash equilibrium). We can however wonder whether assuming that players believe in CBR is ‘more plausible’ than assuming that they ought to play a Nash equilibrium (this is indeed the argument used by Perea (2012, p.2) in the first – chronologically speaking – textbook on EGT to justify its relevance against CGT). A careful look at the formal structure of epistemic models reveals that beliefs are essentially static, and that players are necessarily ‘frozen’ (Binmore 1993, p.330) into a specific equilibrium (e.g. correlated equilibrium if we assume common priors). This means that CBR does not *imply* equilibrium play, but that CBR *is equivalent to* equilibrium play. Assuming CBR (as in EGT) or equilibrium play (as in CGT) are therefore essentially the same thing. If we consider that CGT is unable to justify why players should play equilibrium strategies, then EGT is also probably unable to justify why players should believe in CBR.

Consider for instance a prior distribution in the Hi-Lo game such that HH will occur with probability 1/2 and LL with probability 1/2. The narrative is the following. At the prior stage, the players only know what profile could be implemented and with which probability (the distribution over the set of states of the world). There is then a random move by ‘Nature’, i.e. a fair coin toss whose result is commonly observed, that determines the true state of the world. Each player receives her private information (e.g. I observe ‘Heads’, which reveals that the true state of the world is HH, that you play H, that I play H, and that all of this is common knowledge). The players may then update their beliefs given their private information (if I observe ‘Heads’, I believe that the true state of the world is HH, and therefore believe that you play H). They finally both maximise their expected payoff given those updated beliefs (I play H, because it is a best reply to my belief that you play H). At the posterior stage, the true state of the world (HH) is realised. Since it was rational for us to follow the recommendation of the toss, our strategy choices correspond to the true state of the world. Since nobody had an interest in deviating from the suggestion, the distribution is a correlated equilibrium. Reciprocally, a correlated equilibrium implies that my updated beliefs (my *posteriors*, upon which I maximise my expected utility) *must be equal* to the true state of the world, i.e. the actual strategy profile that is implemented.



The issue at stake is that beliefs are necessarily static, and already correspond to the output of my reasoning: they cannot therefore be used as an *input* of this very same reasoning (as would require a Bayesian explanation of strategic behaviours). There is no actual maximisation process: the players already know *before playing* (once they updated their beliefs to form their posteriors) what equilibrium will be implemented. The ‘Nature move’ in Aumann’s model reveals to the players the true state of the world (or more generally a partition of the states of the world containing the true state). If we are at a correlated equilibrium, nobody has an incentive in deviating. This sounds quite similar to suggesting to players that they ought to play a Nash equilibrium in CGT: once I believe that all the others follow the recommendation, it is indeed rational for me (in the sense of SEUT) to play the Nash equilibrium.

This leads to a more fundamental issue in EGT, i.e. what prior beliefs do represent. The division between the prior, interim, and posterior stages suggests that players first have a prior belief about the outcome of the game, then receive some private information at the interim stage, update their beliefs to form their posterior beliefs, and *in fine* maximise their subjective expected utility. As argued in detail by Binmore (1993), this story mistakenly suggests that there is some actual ‘learning’ at the interim stage, and that moving from the prior to the posterior beliefs represents some kind of information acquisition. The prior beliefs are however just a way to *encode* the set of posteriors beliefs, and ‘serves essentially as an indexing system that keeps track of the library of [...] posteriors’ (Binmore 1993, p.234). According to Binmore’s reading of Savage (e.g. Binmore 2009, pp.128-132), priors are constructed *from* the set of posteriors, such that they can be used analytically to recover (by Bayes’ rule) the posteriors of the players. This means that, unlike the narrative of EGT, there is no hypothetical prior stage in which players actually hold the prior beliefs of the model: they are mere ‘artifacts of a notational device to represent the infinite hierarchies of beliefs on  $X$  of the players, i.e., their “posteriors” at the true state of nature’ (Gul 1998, p.925).

It is not surprising that the previous point echoes Harsanyi’s definition of types, as a way to *encode* infinite belief hierarchies. Harsanyi’s type-based approach has indeed been very influential in modelling strategic uncertainty (as discussed above, Aumann’s state based approach can indeed be interpreted as a restatement of Harsanyi’s, see also

Aumann and Drèze 2008, p.82). This ambiguity was already salient in KL critique, who misinterpreted the nature of the prior distribution in Harsanyi's model:

Harsanyi introduces special assumptions (e.g., a "basic probability distribution" that influences players' priors) that compromise a purely subjective view of probability (Kadane and Larkey 1982a, p.115)

[Kadane and Larkey] seem to think that the "basic probability distribution" [...] is meant to be a factor "that influences the players' prior." In actual fact, the prior distribution is merely a mathematical artefact introduced into the model so that the *subjective* probability distributions used by the players in deciding their strategies can be replaced by *objective* (conditional) probability distributions in order to obtain a game model admitting of analysis by the usual analytic method of game theory (Harsanyi 1982a, pp.122-123, emphasis in original)

The methodological similarity between Harsanyi and EGT is made explicit by Aumann and Brandenburger (1995):

An interactive belief system is not a prescriptive model; it does not suggest actions to the players. Rather, it is a formal framework – a language – for *talking* about actions, payoffs, and beliefs. [...]

The idea of belief system is due to John Harsanyi (1967-68), who introduced the concept of *I-game* to enable a coherent formulation of games in which the players needn't know each other's payoff functions. I-games are just like belief systems, except that in I-games a player's type does not determine his action (only his payoff function). (Aumann and Brandenburger, 1995, pp.1174-1175, emphasis in original)

If the prior distribution is only a tool for the game theorist, can we consider that the players maximise their subjective utility *based on this prior distribution*? According to Aumann and Brandenburger, 'it's not clear that [the players] need concern themselves with the structure of the model' (1995, p.1175). If we interpret prior beliefs literally, and since the

prior distribution is defined over the whole set of strategy profiles, then it means that, at the prior stage, players have subjective beliefs *about their own actions*. Spohn (1977) however suggests that ‘anyone will find it absurd to assume that someone has subjective probabilities for things which are under his control and which he can actualize as he pleases’ (p.115), since this would seriously undermine the claim that the agent is free to choose.<sup>20</sup> Aumann (1987) is aware of this point but emphasises that the probability distribution is only a modelling device at the disposal of an outside observer. This however does not allow the game theorist to assume that the players *have access to this probability distribution*. Indeed, as argued by Kadane and Seidenfeld (1992), a player  $j$  should not form a subjective belief over her own action, but should not also believe that another player  $k$  could believe that  $j$  could form a subjective belief about  $j$ ’s own action. As a result, the primitive of analysis should not be a single prior distribution over the whole set of strategy profiles, but rather a collection of individual beliefs about the choice of others (one distribution per player). This may sound a point of detail, but it turns out that focusing on the prior distribution rather than the collection of conditional distributions imposes quite strong conditions on the set of beliefs allowed for the players.

The reason why game theorists work with a prior distribution rather than the collection of beliefs is its tractability (we have a single ‘objective’ distribution, from which we could derive the  $n$  ‘subjective’ beliefs of the players). Given a collection of  $n$  conditional distributions  $P(X_{-i}|X_i)$  – each one describing the belief of a player  $i$  about the action  $x_{-i} \in X_{-i}$  of the other players, when  $i$  plays  $x_i \in X_i$  – nothing guarantees that we can build a prior distribution  $P(X)$  from which we could derive all the conditional distributions (see e.g. Arnold and Press 1989).<sup>21</sup> This problem is discussed at length by

<sup>20</sup> The important point for the argument is that the prior distribution is a formal tool to encode the beliefs of the players, and not an actual belief held by the player. The argument about the possibility of self-prediction is a disputed topic (see e.g. Joyce 2002, Rabinowicz 2002) but is not of primary importance here.

<sup>21</sup> Consider a coordination game in which we can choose either ‘Left’ or ‘Right’. Suppose that I believe you always choose the same side than me (Left when I choose Left, Right when I choose Right), but that you believe that I always choose the opposite side (Left when you choose Right, Right when you choose Left). It turns out that it is impossible to define here

Harsanyi in part III of the 1967-68 paper, and he recognises that it may not be possible for a player  $j$ , with her beliefs about the types (and hence beliefs) of other players, to derive the beliefs of all the players from a common prior distribution. The ‘Harsanyi doctrine’ is that a player can always encode her beliefs about the game in a prior distribution over the set of outcomes. This is however distinct from the ‘common prior assumption’ (as in Aumann 1987), according to which (i) each player  $j$  can deduce her beliefs about the beliefs of the other players from a *common* distribution  $p_j$ , and (ii) those priors  $p_j$  are identical, i.e.  $p_j = p_i = p$ . While the second part of the proposition is an *empirical* statement (that players share a same ‘background’, and therefore that differences in beliefs can only result from differences in information at the interim stage), the first part is strictly mathematical. Mertens and Zamir (1985, Theorem 4.8) show that, when there exists a coherent prior  $p_j$ , then  $j$  can derive it from her own information, and the distribution is common knowledge. This suggests that, as soon as a coherent prior exists, the players can identify it, and then play as described in the model. However, Mertens and Zamir (1985, example 4.3) also gives a very simple example of a 2x2 game in which there is no coherent prior. Nyarko (2010) and Hellman and Samet (2012) later showed that this problem could be much more general, and that *in most games* with arbitrary beliefs, we cannot derive a coherent prior!

To summarize thus far: EGT claims to take the players’ beliefs as the input of the model, and to describe how they reason when maximising their subjective expected utility. EGT is supposed to be more plausible than CGT, in which players ought to play equilibrium strategies without any individualistic foundations, and are unable to discriminate between multiple equilibria. I have argued (i) that EGT is unable to solve the equilibrium selection problem, (ii) that CBR is equivalent to (rather than implies) equilibrium play, and therefore that beliefs cannot be treated as inputs of the deliberation, and (iii) that the primitive of the model is not the players’ beliefs, but a prior distribution over the set of

a distribution  $P$  over the set of possible outcomes from which we could derive our individual beliefs. It means that I *cannot* define a prior distribution that encodes my beliefs about the game. Larrouy and Lecouteux (2017, pp.329-332) provide a similar illustration in a prisoner’s dilemma.

strategy profiles (meaning that some combinations of individual beliefs *cannot be incorporated* in an EGT model).

A serious issue for EGT is thus that the *narrative* of the model (that beliefs are the primitive, and used as inputs of the decision process) is not coherent with its *formal structure*. This can be exemplified by Stalnaker's (1994) interpretation of the purpose of EGT, as providing a *semantics* for solution concepts, in an analogous way that logicians use semantic models to justify a modal logic.<sup>22</sup> The main object of game theory is the study of solution concepts (and we assume that players ultimately implement the equilibrium profile), and the object of EGT is to establish theorems about sufficient conditions for equilibrium play. The beliefs of the players are not the primitive of the analysis, but are defined *ex post* in order to evaluate various solution concepts.

Aumann's semantic approach is currently the main approach in EGT, although – as recognised by Aumann himself – ‘the semantic formalism appears to involve nontrivial conceptual issues’ (Aumann 1999, p.265). The ‘states of the world’ and ‘partition structures’ are indeed rather obscure objects (though they are perfectly well defined mathematically), and are hardly interpretable in economic terms<sup>23</sup> (see Board 2004, pp.50-52). The ‘states of the world’ are indeed all-inclusive, unlike Savage's ‘states of the world’ in which players *react* to the states of the world, and do not *affect* them. As pointed out above, the model is *static*, it merely ‘says’ what players do. It is neither descriptive (it does not explain how people choose) nor normative (it does not tell what people ought to do), it is *analytic*. It is a mathematically convenient framework to represent a complex problem involving infinite hierarchies of beliefs, and ‘to enable us –

<sup>22</sup> I thank an anonymous referee for pointing out this reference.

<sup>23</sup> A reason why the semantic approach has not been replaced by the ‘hierarchical’ approach of Mertens and Zamir (1985) and Brandenburger and Dekel (1993), which generalises Harsanyi's type-based approach, and in which the states of the world have a more traditional interpretation, is that the hierarchical approach tends to generate an increasing analytical complexity. Aumann (1999) for instance describes it as ‘cumbersome and far from transparent [...] the hierarchy construction is so convoluted that we present it here with some diffidence’ (pp.265-295).

the analysts – to discuss the things we want to discuss: actions, payoffs, beliefs, rationality, equilibrium, and so on’ (Aumann and Brandenburger 1995, p.1175, emphasis in original). It is a very convenient model to *talk* about games, but it does not say anything about how people *do* or *should* behave.

## 5. Game theory without common belief in rationality?

In the *Theory of Games and Economic Behaviour*, Von Neumann and Morgenstern highlighted that the major difference between game theory and one-person decision theory is that:

from the point of view of the player 1 who chooses a variable [...] the other variable can certainly not be considered as a chance event. The other variable [...] is dependent upon the will of the other player, which must be regarded in the same light of ‘rationality’ as one’s own (Von Neumann and Morgenstern, 1953, p.99)

This suggests that a Bayesian approach to game theory should incorporate into individual beliefs this ‘symmetry’ in terms of rationality. Following Harsanyi’s work on games of incomplete information, this problem seemed to have been solved by the introduction of CBR.

I however argued that the whole literature on Bayesian games (including Harsanyi pioneering work in 1967-68) is founded on a contradiction between the *narrative* of the theory and its *formal* structure. That is, the economic interpretation of Bayesian games is not compatible with the mathematical representation of Bayesian games. The prior beliefs are indeed an analytical tool used by the game theorist – as an outside observer – to represent the choices of the agents, but nothing is said about whether the players could compute their choices from this distribution. The ‘Bayesian’ aspect of Bayesian game theory is therefore a mathematical representation: game theorists are studying consistent system of beliefs, suppose that the players will eventually conform to those beliefs, and rationalise the outcome by attributing to the players the prior beliefs that they use to model them. This seems very similar to what classical game theory tried to do: define solution

concepts, and suppose that the players will implement an equilibrium. This would suggest that game theorists became Bayesian, but not the players in their models.

This development of Bayesian game theory might be the consequence of the transformation of neoclassical economics towards a consistency (rather than maximisation) view of rationality during the second half of the 20<sup>th</sup> century (Giocoli 2003). We can indeed notice that, in parallel to Aumann's semantic approach, several 'Bayesian' models of games were developed in the late 1970s, emphasising the *dynamic* processes involved in the revision of the players' beliefs. Kadane and Larkey (1982a) refers for instance to the works of Sanghvi and Sobel (1976), Sanghvi (1978), and Eliashberg (1980, 1981) – to which we could add Shefrin's (1978, 1981) informational equilibrium – but those papers have been largely overlooked in the literature and are barely cited. This commitment to an analytical, *static*, study of games on the contrary deprives the players from any significant role. The object of (Bayesian) game theory should indeed be how players form their beliefs in a strategic interaction, and how they choose given their beliefs. In a semantic model *à la* Aumann, the players are frozen in an equilibrium state: once Nature reveals the state of the world (including the players' own actions), there is no choice to make – everything has already been described in the state of the world. If a player 'wishes to do something else, he is heartily welcome to it; but then the state is different' (Aumann and Brandenburger 1995, p.1174). Such a model does not say anything about *how* a player comes to form coherent beliefs, and is just a mathematical representation of a group of individuals at a correlated equilibrium.

As already pointed out by Kadane and Larkey (1983), game theory is 'speculative-normative', and 'results accumulate in a mammoth deductive edifice that neither cumulatively improves our understanding of how the world works nor cumulatively improves our ability to intervene in the world' (p.1377). The general confusion between what the models actually mean, and what the theorists claim they mean, perfectly illustrates KL's (1983) point about the fuzziness of the scientific object of game theory. While its proponents claim that EGT introduces into game-theoretic analysis the fact that economic agents *reason* about each other's objectives and beliefs – and therefore that it becomes more plausible than CGT – it simultaneously imposes drastic (and unrealistic) conditions on the epistemic states of the players. Formally, the theory is internally

consistent and thus perfectly valid. However, the assumptions of the model (CBR and common priors mainly) are routinely and casually defended on ‘empirical’ grounds. Aumann (2010, 30) for instance sees common priors as a ‘reasonable’ assumption, while CBR is ‘some sort of *benchmark*’ (emphasis in original). Just like Harsanyi’s defence of rational choice theory, ‘everything else is calculated as a *departure* from the benchmark. It’s like a no-friction assumption in physical systems, or the no-air resistance assumption’ (Aumann 2010, p.30, emphasis in original). Note however that no experimental test of those assumptions is suggested. They are taken for granted, just as the idea that players are at least approximately rational. But if the scientific validity of EGT is an empirical matter, then we can seriously doubt of the soundness of an assumption like CBR.

Brandenburger (2014) acknowledges that, ‘as yet, EGT is devoid of empirical content’ (p. xxii). It becomes necessary to ‘look beyond the mathematics’, and explaining the beliefs of the players requires ‘[looking] to outside factors, such as the environment in which strategic interaction takes place, the prior experiences of the players, their personalities, etc.’ (Brandenburger 2014, p.xxxi). The development of EGT has followed thus far a research program in interactive epistemology. What ultimately matters is the definition of coherent system of beliefs – the role of the players is trivialised, since they simply play what the state of the world says they play. If we consider that game theory should be about how players reason and form their beliefs in games, then we should probably leave the assumption of CBR. Borrowing Giocoli’s (2003) expressions, we may interpret such reorientation as an ‘escape from perfect foresight’, in opposition to the current ‘escape from psychology’ pursued by EGT. KL’s early suggestion was that game theorists should turn to *psychology* to offer epistemic foundations of strategic behaviours. Harsanyi vigorously objected to this idea, and stressed that game theory should be based on the postulate of mutually expected rationality, which was later formalised through CBR. I argued that this line of research makes the deliberation of the agent, who knows the outcome of the game before actually playing it, vacuous: imposing CBR thus embodies the players with perfect foresight. Leaving CBR is thus necessary to escape perfect foresight, with the objective of an empirically testable theory of games. Rather ironically, both Brandenburger (2010) and Perea (2014) quote Morgenstern’s (1935) battle of wits between Holmes and Moriarty to illustrate the need for ‘models in which agents *reason* about the possible choices and beliefs of other agents’ (Perea 2014, p.3,



emphasis in original). Morgenstern however used this story to illustrate that reciprocal influences among players (such as conditioning my beliefs and actions on my beliefs about your beliefs about me) make the principle of maximisation inadequate, and that economics should consider the limited cognitive abilities of the agents (Innocenti 1995, pp.212-214, see also Giocoli 2003, p.163). This is clearly in contradiction with assuming CBR, since solving the kind of indeterminacy faced by Holmes and Moriarty requires an arbitrary move by one of the players. I am not claiming that EGT is ill suited to deal with this kind of problems, but simply that cases of false beliefs and bounded rationality are clearly not the main concern of current research in EGT. We should of course mention level-k theory (Stahl and Wilson 1994, Nagel 1995) and the closely related cognitive hierarchy theory (Camerer *et al* 2004) as epistemic models incorporating the bounded rationality of the players. Those theories however face the difficulty of offering testable assumptions about the anticipations by strategic players of the behaviours of naïve (level-0), non-strategic players. Hargreaves-Heal *et al* (2014) highlight for instance the lack of portability of models of level-0 behaviours, which are often game-specific – completing level-k theory would require developing a general (psychological) theory of naïve behaviours in games.

Another research program in EGT would be to study the role of institutions and social norms, to understand how players' beliefs evolve in different social contexts, and how institutions could shape our beliefs. With this objective in mind, SEUT could still give a decision-theoretic foundation to game theory, but doing so requires (i) dropping the assumption of CBR, and (ii) investigating how players form their beliefs in a strategic interaction. Note that I do not claim that SEUT is the *only* acceptable way to look at games,<sup>24</sup> nor that it is founded on solid axiomatic grounds (see Mariotti (1995, 1997)), but – quite pragmatically – that it constitutes a very convenient and useful approach to look at decision problems. If we consider that game theory should be about the strategic interactions of players who expect (or at least deem possible) each other to act rationally,

<sup>24</sup> We could also consider e.g. team reasoning (see Lecouteux 2018 for a literature review), Stackelberg reasoning (Colman and Bacharach 1997), or the theory of social situations (Greenberg 1990, Mariotti 1997) as alternative approaches.

then it means that a theory of games should describe how players adjust their initial beliefs (which may be seen as derived from their ‘gut feelings’) to take this strategic dimension into account. The formation of the priors thus becomes ‘the central unknowns of the theory’:

What brings me to have the prior probabilities that I do for your deciding on one option and another is a question not answered (and rarely asked) by the Bayesian theory of games. The absence of an independent account of what is in the players’ priors is a grave lacuna. There are many games for which, once the priors are given, the identities of the rational acts follow trivially, and then game theory itself is trivialized if it is merely assumed that the priors are such and such. To avoid this trivialization by Bayesianization, we must take the content of the priors in such cases to be the central unknowns of the theory, endogenous to it. (Bacharach and Hurley 1991, p.26)

This does not necessarily require an important injection of psychology into game theory – in line with Brandenburger’s (2014, p.xxii) suggestion of referring to the ‘theory of mind’ literature in cognitive sciences, Larrouy and Lecouteux (2017) for instance offer an axiomatization of *simulation thinking* which can explain the emergence of common belief of events in a similar vein than Friedell (1969) and Lewis (1969), and rationalise the coordination on focal points. Such a theory of belief formation would require investigating how players choose in ‘large worlds’ (in the sense of Savage), i.e. how players interpret a choice problem and focus on isolated features of the large world (Savage 1954, pp.82-91), so as to identify salient outcomes and focal points. Developing a genuinely intersubjective and dynamic theory of games thus requires considering explicitly the role of inductive reasoning in decision making, similarly to Lewis’ analysis of convention (Cubitt and Sugden 2003).

## References

- Armbruster, W., & Böge, W. (1979). Bayesian game theory. *Game theory and related topics*, 17, 28.
- Arnold, B. C., & Press, S. J. (1989). Compatible conditional distributions. *Journal of the American Statistical Association*, 84(405), 152-156.
- Aumann, R. J. (1974). Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1), 67-96.
- Aumann, R. J. (1976). Agreeing to disagree. *The annals of statistics*, 1236-1239.
- Aumann, R. J. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica: Journal of the Econometric Society*, 1-18.
- Aumann, R. J. (1999). Interactive epistemology I: knowledge. *International Journal of Game Theory*, 28(3), 263-300.
- Aumann, R. J. (2005). An Interview with Robert Aumann, by Sergiu Hart. *Macroeconomic Dynamics* 9: 683-740.
- Aumann, R. J. (2010). In V. Hendricks & O. Roy (Eds.), *Epistemic logic: Five questions* (pp. 21—33). Copenhagen: Automatic Press/VIP.
- Aumann, R. J., & Brandenburger, A. (1995). Epistemic conditions for Nash equilibrium. *Econometrica: Journal of the Econometric Society*, 1161-1180.
- Aumann, R. J., & Dreze, J. H. (2008). Rational expectations in games. *The American Economic Review*, 72-86.

- Aumann, R. J., & Dreze, J. H. (2009). Assessing strategic risk. *American Economic Journal: Microeconomics*, 1(1), 1-16.
- Aumann, R. J. and Maschler, M. (1967), Repeated Games with Incomplete Information: A Survey of Recent Results, Mathematica, Inc. Report ST-116, 287–403, Princeton. Reprinted as Chapter II in Aumann and Maschler (1995).
- Aumann, R. J. and Maschler, M. (1968), Repeated Games of Incomplete Information: The Zero-Sum Extensive Case, Mathematica, Inc. Report ST-143, 37–116, Princeton. Reprinted as Chapter IV in Aumann and Maschler (1995).
- Aumann, R. J. and Maschler, M. (1995), Repeated Games of Incomplete Information, MIT Press, Cambridge, MA.
- Bacharach, M. (1985). Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory*, 37(1), 167-190.
- Bacharach, M. (2006). *Beyond individual choice: teams and frames in game theory*. Princeton University Press.
- Bacharach, M., & Hurley, S. (1991). Introduction. In M. Bacharach & S. Hurley (Eds.), *Issues and advances in the foundations of decision theory* (pp. 1-38). Oxford, UK: Blackwell.
- Battigalli, P., & Bonanno, G. (1999). Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53(2), 149-225.
- Bernheim, B. D. (1984). Rationalizable strategic behavior. *Econometrica: Journal of the Econometric Society*, 1007-1028.
- Bernheim, B. D. (1986). Axiomatic characterizations of rational choice in strategic environments. *The scandinavian journal of economics*, 473-488.

- Binmore, K. G. (1992). *Fun and games: A text on game theory*. Lexington, MA: Heath.
- Binmore, K. G. (1993). De-Bayesing Game Theory. In Binmore, K. G., Kirman, A. P., & Tani, P. (Eds.). *Frontiers of game theory*. MIT Press.
- Binmore, K. G. (2009). *Rational decisions*. Princeton, NJ: Princeton University Press.
- Board, O. (2004). Dynamic interactive epistemology. *Games and Economic Behavior*, 49(1), 49-80.
- Böge, W. (1974). Gedanken über die angewandte Mathematik. In *Mathematiker über die Mathematik* (pp. 288-328). Springer Berlin Heidelberg.
- Böge, W., & Eisele, T. (1979). On solutions of Bayesian games. *International Journal of Game Theory*, 8(4), 193-215.
- Brandenburger, A. (2007). The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35(4), 465-492.
- Brandenburger, A. (2010). Origins of epistemic game theory. In V. Hendricks & O. Roy (Eds.), *Epistemic logic: Five questions* (pp. 59–69). Copenhagen: Automatic Press/VIP.
- Brandenburger, A. (2014). *The language of game theory: putting epistemics into the mathematics of games* (Vol. 5). World scientific.
- Brandenburger, A., & Dekel, E. (1987a). Common knowledge with probability 1. *Journal of Mathematical Economics*, 16(3), 237-245.

- Brandenburger, A., & Dekel, E. (1987b). Rationalizability and correlated equilibria. *Econometrica: Journal of the Econometric Society*, 1391-1402.
- Brandenburger, A., & Dekel, E. (1993). Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, 59(1), 189-198.
- Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3), 861-898.
- Colman, A. M., & Bacharach, M. (1997). Payoff dominance and the Stackelberg heuristic. *Theory and Decision*, 43(1), 1-19.
- Cubitt, R. P., & Sugden, R. (2003). Common Knowledge, Salience and Convention: a Reconstruction of David Lewis' Game Theory. *Economics & Philosophy*, 19(2), 175-210.
- Eliashberg, J. (1980). *Bayesian Analysis of Competitive Decision-Making Situations*. unpublished mimeo.
- Eliashberg, J. (1981). An investigation of competitive preference structures and posterior performance through a Bayesian decision-theoretic approach. *Management Science*, 27(7), 785-801.
- Friedell, M. F. (1969). On the structure of shared awareness. *Behavioral Science*, 14(1), 28-39.
- Giocoli, N. (2003). *Modeling rational agents: From interwar economics to early modern game theory*. Edward Elgar Publishing.
- Giocoli, N. 2009. "Three alternative (?) stories on the late 20th-century rise of game theory", *Studi e Note di Economia*, 14 (2), 187-210.

- Giocoli, N. (2013). From Wald to Savage: homo economicus becomes a Bayesian statistician. *Journal of the History of the Behavioral Sciences*, 49(1), 63-95.
- Greenberg, J. (1990). *The theory of social situations: an alternative game-theoretic approach*. Cambridge University Press.
- Guala, F. (2016). *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton University Press.
- Gul, F. (1998). A comment on Aumann's Bayesian view. *Econometrica*, 66(4), 923-927.
- Hargreaves Heap, S., Rojo Arjona, D., & Sugden, R. (2014). How Portable Is Level-0 Behavior? A Test of Level-k Theory in Games With Non-Neutral Frames. *Econometrica*, 82(3), 1133-1151.
- Hammond, P. J. (2009). Isolation, assurance and rules: can rational folly supplant foolish rationality?. In K. Basu & R. Kanbur (Eds.), *Arguments for a better world: Essays in honor of Amartya Sen. Volume I: Ethics, welfare, and measurement*. Oxford, NY: Oxford University Press.
- Harsanyi, J. C. (1962). Bargaining in ignorance of the opponent's utility function. *Journal of Conflict Resolution*, 6(1), 29-38.
- Harsanyi, J. C. (1967-1968). Games with incomplete information played by 'Bayesian' players, I-III, *Management Science*, 14(3-5-7), pp. 159-182, 320-334, 486-502.
- Harsanyi, J. C. (1975). The tracing procedure: A Bayesian approach to defining a solution form-person noncooperative games. *International Journal of Game Theory*, 4(2), 61-94.
- Harsanyi, J. C. (1977). *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge England New York: Cambridge University Press.

- Harsanyi, J. C. (1982a). Comment – subjective probability and the theory of games: Comments on Kadane and Larkey's paper. *Management Science*, 28(2), 120-124.
- Harsanyi, J. C. (1982b). Rejoinder to professors Kadane and Larkey. *Management Science*, 28(2), 124-125.
- Harsanyi, J. C., & Selten, R. (1972). A generalized Nash solution for two-person bargaining games with incomplete information. *Management Science*, 18(5-part-2), 80-106.
- Harsanyi, J. C., & Selten, R. (1988). A general theory of equilibrium selection in games. *MIT Press Books*.
- Hellman, Z., & Samet, D. (2012). How common are common priors?. *Games and Economic Behavior*, 74(2), 517-525.
- Hillas, J., & Kohlberg, E. (2002). Foundations of strategic equilibrium. In R. Aumann & S. Hart (Eds.), *Handbook of Game Theory with Economic Applications* (Vol. 3, pp. 1597-1663). Amsterdam: North Holland
- Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, 23(1), 1-25..
- Innocenti, A. (1995). Oskar Morgenstern and the heterodox potentialities of the application of game theory to economics. *Journal of the History of Economic Thought*, 17(2), 205-227.
- Joyce, J. M. (2002). Levi on causal decision theory and the possibility of predicting one's own actions. *Philosophical Studies*, 110(1), 69-102.



- Kadane, J. B., & Larkey, P. D. (1982a). Subjective probability and the theory of games. *Management Science*, 28, 113–120.
- Kadane, J. B., & Larkey, P. D. (1982b). Reply to Professor Harsanyi. *Management Science*, 28, 124.
- Kadane, J. B., & Larkey, P. D. (1983). The confusion of is and ought in game theoretic contexts. *Management Science*, 29(12), 1365-1379.
- Kadane, J. B., & Seidenfeld, T. (1992). Equilibrium, Common Knowledge, and Optimal Sequential Decisions. In Bicchieri, C. & dalla Chiara, M. L. (Eds.). *Knowledge, Belief, and Strategic Interaction* (Cambridge: Cambridge University Press), pp. 27-45
- Kohlberg, E., & Zamir, S. (1974). Repeated games of incomplete information: The symmetric case. *Annals of Statistics*, 2(5), 1040-1041.
- Kohlberg, E. (1975a). Optimal strategies in repeated games with incomplete information. *International Journal of Game Theory*, 4(1), 7-24.
- Kohlberg, E. (1975b). The information revealed in infinitely-repeated games of incomplete information. *International Journal of Game Theory*, 4(2), 57-59.
- Kripke, S. A. (1963). Semantical analysis of modal logic I Normal modal propositional calculi. *Mathematical Logic Quarterly*, 9: 67-96.
- Larrouy, L., & Lecouteux, G. (2017). Mindreading and endogenous beliefs in games. *Journal of Economic Methodology*, 24(3): 318-343.
- Lecouteux, G. (2018). What does ‘we’ want? Team reasoning, game theory, and unselfish behaviours. *Revue d'économie politique*, 128(3): 331-332.

- Levi, I. (1998). Prediction, Bayesian deliberation and correlated equilibrium. In W. Leinfellner & E. Köhler (Eds.), *Game Theory, experience, rationality: foundations of social sciences, economics and ethics in honor of John C. Harsanyi* (pp. 173–185). Dordrecht: Springer.
- Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press
- Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. New York: Wiley.
- Luce, R. D., & Adams, E. W. (1956). The determination of subjective characteristic functions in games with misperceived payoff functions. *Econometrica*, 24(2): 158-171.
- Mariotti, M. (1995). Is Bayesian rationality compatible with strategic rationality? *The Economic Journal*, 105(432), 1099–1109.
- Mariotti, M. (1997). Decisions in games: Why there should be a special exemption from bayesian rationality. *Journal of Economic Methodology*, 4(1), 43–60.
- Mertens, J. F. (1971). The Value of Two-Person Zero-Sum Repeated Games the extensive case. *International Journal of Game Theory*, 1(1), 217-227.
- Mertens, J. F. & Zamir, S (1971). The Value of Two-Person Zero-Sum Repeated Games with Lack of Information on Both Sides. *International Journal of Game Theory*, 1(1), 39-64.
- Mertens, J. F., & Zamir, S. (1985). Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14(1), 1-29.

- Milgrom, P. (1981). An axiomatic characterization of common knowledge. *Econometrica* 49(1): 219-222.
- Monderer, D., & Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1(2), 170-190.
- Morgenstern, O. (1935). Perfect foresight and economic equilibrium. *Selected economic writings of Oskar Morgenstern*, 169-83.
- Myerson, R. B. (1990) *Game Theory. Analysis of Conflict*, Cambridge, Massachussets: Harvard University Press
- Myerson, R. B. (2004). Comments on “Games with Incomplete Information Played by ‘Bayesian’ Players, I–III Harsanyi's Games with Incoplete Information”. *Management Science*, 50, 1818-1824.
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, 85(5), 1313-1326.
- Nyarko, Y. (2010). Most games violate the common priors doctrine. *International Journal of Economic Theory*, 6(1), 189-194.
- Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica: Journal of the Econometric Society*, 1029-1050.
- Perea, A. (2012). *Epistemic game theory: reasoning and choice*. Cambridge University Press.
- Perea, A. (2014). From classical to epistemic game theory. *International Game Theory Review*, 16(01), 1440001.

- Ponssard, J. P. (1975a). A note on the LP formulation of zero-sum sequential games with incomplete information. *International Journal of Game Theory*, 4(1), 1-5.
- Ponssard, J. P. (1975b). Zero-sum games with “almost” perfect information. *Management Science*, 21(7), 794-805.
- Ponssard, J. P., & Zamir, S. (1973). Zero-sum sequential games with incomplete information. *International Journal of Game Theory*, 2(1), 99-107.
- Rabinowicz, W. (2002). Does practical deliberation crowd out self-prediction?. *Erkenntnis*, 57(1), 91-122.
- Rubinstein, A. (1989). The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge". *The American Economic Review*, 385-391.
- Samet, D. (1990). Ignoring ignorance and agreeing to disagree. *Journal of Economic Theory*, 52(1), 190-207.
- Sanghvi, A. P. (1978). Sequential games as stochastic processes. *Stochastic Processes and Their Applications*, 6(3), 323-336.
- Sanghvi, A. P., & Sobel, M. J. (1976). Bayesian games as stochastic processes. *International Journal of Game Theory*, 5(1), 1-22.
- Savage L. J. (1954). *The foundations of statistics*. New-York: John Wiley.
- Schotter, A., & Schwödiauer, G. (1980). Economics and the theory of games: a survey. *Journal of Economic Literature*, 18(2), 479-527.
- Sent, E. M. (2004). Behavioral economics: how psychology made its (limited) way back into economics. *History of Political Economy*, 36(4), 735-760.

- Shefrin, H. M. (1978). Differential information and informational equilibrium. *Economics Letters*, 1(1), 33-37.
- Shefrin, H. M. (1981). Games with self-generating distributions. *The Review of Economic Studies*, 48(3), 511-519.
- Skyrms, B. (1990). *The dynamics of rational deliberation*. Harvard University Press.
- Sorin, S. (1979). A note on the value of zero-sum sequential repeated games with incomplete information. *International Journal of Game Theory*, 8(4), 217-223.
- Spohn, W. (1977). Where Luce and Krantz do really generalize Savage's decision model. *Erkenntnis*, 11(1), 113-134.
- Spohn, W. (1982). How to make sense of game theory. In *Philosophy of Economics* (pp. 239-270). Springer, Berlin, Heidelberg.
- Stahl, D. O., & Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of economic behavior & organization*, 25(3), 309-327.
- Stalnaker, R. (1994). On the evaluation of solution concepts. *Theory and decision*, 37(1), 49-73.
- Sugden, R. (2015). Looking for a psychology for the inner rational agent. *Social Theory and Practice*, 41(4), 579-598.
- Tan, T. C. C., & da Costa Werlang, S. R. (1988). The Bayesian foundations of solution concepts of games. *Journal of Economic Theory*, 45(2), 370-391.

Vanderschraaf, P. & Sillari, G. (2014). Common Knowledge. *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL <<https://plato.stanford.edu/archives/spr2014/entries/common-knowledge/>>.

Von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior*. 3rd edition, Princeton university press.

Zamir, S. (1971). On the relation between finitely and infinitely repeated games with incomplete information. *International Journal of Game Theory*, 1(1), 179-198.

**DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2017**  
*GREDEG Working Papers Released in 2017*

- 2017-01** LAUREN LARROUY & GUILHEM LECOUTEUX  
*Mindreading and Endogenous Beliefs in Games*
- 2017-02** CILEM SELIN HAZIR, FLORA BELLONE & CYRIELLE GAGLIO  
*Local Product Space and Firm Level Churning in Exported Products*
- 2017-03** NICOLAS BRISSET  
*What Do We Learn from Market Design?*
- 2017-04** LISE ARENA, NATHALIE ORIOL & IRYNA VERYZHENKO  
*Exploring Stock Markets Crashes as Socio-Technical Failures*
- 2017-05** IRYNA VERYZHENKO, ETIENNE HARB, WAËL LOUHICHI & NATHALIE ORIOL  
*The Impact of the French Financial Transaction Tax on HFT Activities and Market Quality*
- 2017-06** FRÉDÉRIC MARTY  
*La régulation du secteur des jeux entre Charybde et Scylla*
- 2017-07** ALEXANDRU MONAHOV & THOMAS JOBERT  
*Case Study of the Moldovan Bank Fraud: Is Early Intervention the Best Central Bank Strategy to Avoid Financial Crises?*
- 2017-08** NOBUYUKI HANAKI, EIZO AKIYAMA, YUKIHIKO FUNAKI & RYUICHIRO ISHIKAWA  
*Diversity in Cognitive Ability Enlarges Mispricing in Experimental Asset Markets*
- 2017-09** THIERRY BLAYAC & PATRICE BOUGETTE  
*Should I Go by Bus? The Liberalization of the Long-Distance Bus Industry in France*
- 2017-10** AYMERIC LARDON  
*On the Coalitional Stability of Monopoly Power in Differentiated Bertrand and Cournot Oligopolies*
- 2017-11** ISMAËL RAFAÏ & MIRA TOUMI  
*Pay Attention or Be Paid for Attention? Impact of Incentives on Allocation of Attention*
- 2017-12** GÉRARD MONDELLO  
*Un modèle d'accident unilatéral: incertitude non-radical et estimations différenciées*
- 2017-13** GÉRARD MONDELLO  
*Lenders and Risky Activities: Strict Liability or Negligence Rule?*
- 2017-14** FRÉDÉRIC MARTY  
*Économie des algorithmes et ordre concurrentiel : réflexions sur les abus d'exploitation et les collusions fondés sur des algorithmes de prix*
- 2017-15** AGNÈS FESTRÉ, PIERRE GARROUSTE, ANKINÉE KIRAKOZIAN & MIRA TOUMI  
*The Pen is Might Be Mightier than the Sword: How Third-party Advice or Sanction Impacts on Pro-environmental Behavior*
- 2017-16** EDWARD LORENZ & SOPHIE POMMET  
*Innovation, Credit Constraints and National Banking Systems: A Comparison of Developing Nations*
- 2017-17** PATRICE BOUGETTE & CHRISTOPHE CHARLIER  
*Antidumping and Feed-In Tariffs as Good Buddies? Modeling the EU-China Solar Panel Dispute*
- 2017-18** NOBUYUKI HANAKI, EIZO AKIYAMA & RYUICHIRO ISHIKAWA  
*Behavioral Uncertainty and the Dynamics of Traders' Confidence in their Price Forecasts*

- 2017-19** ADNANE KENDEL, NATHALIE LAZARIC & KEVIN MARÉCHAL  
*What Do People 'Learn By Looking' at Direct Feedback on their Energy Consumption? Results of a Field Study in Southern France*
- 2017-20** NICOLAS VALLOIS & DORIAN JULLIEN  
*Estimating Rationality in Economics: A History of Statistical Methods in Experimental Economics*
- 2017-21** NICOLAS VALLOIS & DORIAN JULLIEN  
*Replication in Experimental Economics: A Historical and Quantitative Approach Focused on Public Good Game Experiments*
- 2017-22** BENJAMIN MONTMARTIN, MARCOS HERRERA & NADINE MASSARD  
*R&D Policy regimes in France: New Evidence from a spatio-temporal Analysis*
- 2017-23** THOMAS BOYER-KASSEM & SÉBASTIEN DUCHÊNE  
*On Discrimination in Health Insurance*
- 2017-24** MICHELA CHESSA & PATRICK LOISEAU  
*On Non-monetary Incentives for the Provision of Public Goods*
- 2017-25** NICOLAS BRISSET  
*Models as Speech Acts: The Telling Case of Financial Models*
- 2017-26** NOBUYUKI HANAKI, EIZO AKIYAMA & RYUICHIRO ISHIKAWA  
*Effects of Eliciting Long-run Price Forecasts on Market Dynamics in Asset Market Experiments*
- 2017-27** CHARLES AYOUBI, MICHELE PEZZONI & FABIANA VISENTIN  
*The Important Thing is not to Win, it is to Take Part: What If Scientists Benefit from Participating in Competitive Grant Races?*
- 2017-28** FRÉDÉRIC MARTY  
*Le prix des services juridiques : entre défaillance de la réglementation et défaillance de marché ?*
- 2017-29** JOAO V. FERREIRA, NOBUYUKI HANAKI & BENOÎT TARROUX  
*On the Roots of the Intrinsic Value of Decision Rights: Evidence from France and Japan*
- 2017-30** GUILHEM LECOUTEUX  
*Bayesian Game Theorists and Non-Bayesian Players*