Ontological foundation of Nash Equilibrium

Pelosse, Yohan

February 2011

Ontological Foundation of Nash Equilibrium*

Yohan Pelosse

e-mail: yohanpelosse@googlemail.com

Gate Lyon-St-Etienne CNRS UMR 5824
93 Chemin des mouilles
69130 Ecully (France)
Tel.: +33 (0) 4 72 86 60 60 Fax.: +33 (0) 4 72 86 60 90

Abstract: In the classical definition of a game, the players' hierarchies of beliefs are not part of the description. So, how can a player determine a rational choice if beliefs are initially nonexistent in his mind? We address this question in a three-valued Kripke semantics wherein statements about whether a strategy or a belief of a player is rational are initially indeterminate i.e. neither true, nor false. This "rationalistic" Kripke structure permits to study the "mental states" of players when they consider the perspectives or decision problems of the others, in order to form their own beliefs. In our main Theorem we provide necessary and sufficient conditions for Nash equilibrium in an n-person game. This proves that the initial indeterminism of the game model, together with the free will of rational players are at the origin of this concept. This equivalence result has several implications. First, this demonstrates that a Nash equilibrium is *not* an interactive solution concept but an intrinsic principle of decision making used by each player to shape his/her own beliefs. Second, this shows that a rational choice must be viewed *in statu nascendi* i.e. conceived as a genuine "act of creation" *ex nihilo*, rather than as a pre-determined decision, arising from an underlying history of the game.

Keywords and phrases: Free will, Indeterminism, Lukasiewicz's three-valued logic, Player's perspective, Rationalistic frames, Relational truth-values, Three-valued Kripke semantics.

1. Introduction

The object of game theory, as originally formulated by von Neumann and Morgenstern (1944) is to predict the behavior of rational players. The starting point for most game theory, which is usually thought of as the embodiment of "rational behavior", is the Nash equilibrium solution concept (Nash, 1950). A Nash equilibrium is defined as a n-tuple of strategies or strategy profile (one strategy for each player) if each player's strategy is optimal against the others' strategies. A common view is that the interactive epistemology[1] under which rational individuals play such an "equilibrium point" is quite demanding. (see e.g., Aumann and Brandenburger (1995), Bernheim (1984), Pearce (1984)).
The goal of this paper is to answer the central foundational problem posed to game theory :

*This paper is the first part of the initial version of "The intrinsic quantum nature of Nash equilibrium mixtures" presented at the "Quantum Physics meets TARK" workshop (Groningen, July 15, 2011) and Maastricht seminar (Meteor-ETBC, November 2011). I am grateful to these seminar audiences for helpful comments and stimulating discussions.

[1]For a thoroughful survey on the epistemic foundation of game theory, see e.g., Battigalli and Bonnanno (1999).

"*Why equilibrium?*" (see Sabourian and Juang 2008). We shall answer this question by taking the formal definition of the classical game model as it stands; that is, we will examine the game model when players have no pre-existing hierarchies of beliefs in their minds. Hence, we comply with the original tenseless definition of a game by excluding the possibility that players may have formed their beliefs about the others' strategies in the past. Instead, we will analyze how players endowed of a *bona fide* free will, without initial beliefs, can determine their choice of a strategy. Our main result proves that the free will assumption of the game model, together with the classical rationality postulate permits to constructively derive the Nash equilibrium concept as a *self-referential* principle of determination of each player. This is therefore in stark contrast with the conventional wisdom, which claims that processes by which individuals reach decisions are *a priori* completely absent of classical game theory.

More precisely, the methodology pursued in the paper is as follows. We will exhibit a process of reasoning leading each player to determine a Nash equilibrium by incorporating in the analysis the initial indeterminism of the game model. To do so, we start by asking how a rational player with free will, may form his beliefs in order to discover a rational strategy in the classical game model i.e. how should a rational player define his prior beliefs *ex nihilo*? The gist of our approach builds on the following two observations:

(i) *The classical game model is complete in the sense that its complete description is given by the strategy sets, the outcome map, and the payoff functions, and*;

(ii) *If we do not append some extraneous hierarchies of beliefs i.e. players have free will, then rationality is a relativistic or relational concept in the sense that it consists of making an optimal choice that has to be justifiable by an initially nonexistent belief.*

Taken together, (i) and (ii) imply that atomic statements like $A^i :=$"*strategy a is optimal in the game G for player i*" are generally neither absolutely "true", nor absolutely " false" but indeterminate.[2] Hence, a non-classical logic—the three-valued logic of Lukasiewicz (1930)—enters the picture of the game model in its own right because this model does not (generally) contain the answers to questions like "*what constitutes a rational behavior?*".

So far, the implications of the assumption of rationality has been explored in epistemic models in which the statements are either true or false. So, let us stress that the use of a non-classical logic to analyze the classical game model is *not* an extraneous postulate, but is rather its inevitable formal consequence. This is so, since the notion of (hierarchies of) beliefs are not part of the description of the game. So how could we assign some sharp truth-values to some statements that pertain on undefined objects?

In fact we could make a strong case for the use of such a non-classical logic by noting that the assertions made in game theory like "*player i will choose action a in the experiment*" have to be considered as *future contingents* i.e. assertions about a future choice of a player are of the same nature as the well-known Aristotle "tomorrow's sea battle". As argued by Kripke (1975), the three-valued logic of Lukasiewicz models the *lack* of sharp truth-values on statements. Hence, unlike the "fuzzy logics", it is important to understand that the "third truth-value" cannot be construed as reflecting the ignorance of a player about statements that are unknown.[3] In game theory, this non-classical logic captures the "ontological open-

[2] Indeed, recall that in a game a strategy is generally neither strongly dominant, nor a never best reply.

[3] This view is consistent with the idea that the truth of a proposition consists in representing an *actual* state of affairs. This is usually referred to as the "correspondence theory of truth". To put Aristotle's argument in modern terms, he claimed that some statements about the future have no truth value today because

ness" of the players behavior in the game model i.e., the fact that the behavior is not settled in the game model as long as it is not determined by the players themselves. This is so, since all the assertions formulated in the game model are examples of the "future contingents" of Aristotle.[4] It is well-known that a standard way to capture in a rigorous semantic the indeterminacy of assertions that pertain on future events is to use the three-valued logic of Lukasiewicz.[5] So, we can also understand the use of the three-valued logic to model the absence of initial players' strategies and beliefs in the classical game model as the result of the Aristotle's idea that contingent statements about the future have no determinate truth value. The bottom line is that the lack of a sharp truth-value is part of the description of the classical game model, *on a par with* the strategy sets, the outcome map, and the payoff functions. Hence, the absence of any hierarchies of beliefs in the game model means that the truth of what is rational for a player is *not* relative to a particular history of the game.

Note the immediate implication for the notion of "uncertainty" in games; if players are free to choose their strategies and beliefs, then a pair of strategies-beliefs are under-determined until a player determines such objects himself. Thus, the game model explains how a definite rational strategy *comes into being* in the mind of the player. This means that a rational strategy will be unpredictable before it is determined: the future choice of a strategy and a belief of a player is uncertain because it does not yet exist in the mind of the player. Thus this uncertainty has an *ontological* character.

This intial under-determination of the game model—an objective indeterminism—implies that players can choose their beliefs freely. Note that this initial assumption of an open future, i.e., the future can be changed by the choices the players make, is therefore very much in line with the "tense logic" of Prior (1967), and the theory of logic and philosophy of action of Belnap et al. (2001), or Horty (2001), who analyze the choices made by agents living in a world with an "indeterministic causal structure ".[6] Finally, there is also a methodological reason for the use of this non-classical logic in the game model. Assuming the existence of extraneous pre-existing beliefs implies that there is no room for a real free will of players in game theory. Arguably, this seems a quite suspect assumption, if the theory aims at predicting the actual behavior of individuals. [7]

---

of indeterminism. Or the opposite that indeterminism must be true because we cannot find a consistent valuation of all possible statements about the future in a quantum-mechanical universe.

[4] Aristotle, in De interpretatione IX, for instance, held that only those propositions about the future which are either necessarily true, or necessarily false, or "predetermined" in some way have a determinate truth-value. The initial under-determination of the classical game model conflicts with this necessity of sharp truth-values.

[5] In the three-valued logic of Lukasiewicz the law of the excluded middle does not hold. This has been criticized by Prior (1953). However, as we shall discuss, in the present three-valued Kripke model, this property is in fact the direct consequence of the fact that statements that pertain on the rational choice of a player can only take on *relativistic.* truth-values. Again, this is due to the fact that our use of the Lukasiewicz three-valued logic captures the ontological/real inexistence of any pre-existing beliefs in the minds of players, not their epistemic plausibility.

[6] It is therefore worth noting that the present tenseless structure of the game model precludes any evaluation of a game-theoretic statement by singling out a particular history, a path of a tree-like structure i.e. a branching time structure a la Belnap.

[7] Note that the absence of free will is at odds with all the current experimental findings in physics. In particular, quantum theory and (special) relativity make correct predictions in already well tested situations by assuming that experimentalists can freely (independently) choose measurement settings.

One could say that the presence of a non-classical logic marks the transition from an epistemic game theory (see, e.g., Harsanyi (1973), Aumann (1987) and Brandenburger and Dekel (1989)) to an ontological analysis of game theory. At first sight, this "ontological approach" of games might therefore seem in stark contrast with the usual epistemic approach to game theory, where the introduction of pre-existing hierarchies of beliefs is a prerequisite. To our way of thinking, this is rather its direct corollary, for as well put by Brandenburger, an important contribution of the epistemic program has precisely been to demonstrate "*that there isn't one right set of assumptions to make about a game.*" [Brandenburger, p.490, 2007].

So, the initial lack of pre-existing strategies and beliefs in the classical game model raises the central question: How will a player ascribe a relative truth-value, true, to a particular rational strategy?

The answer is simple; in the game model, each player possesses only pieces of a puzzle made of contingent statements about the optimality of a strategy. Playing a Nash equilibrium is the picture on the box—the principle so to speak—to guide their assembly.[8] Alternatively put, since atomic statements are generally neither true nor false, each player i who "speaks" a "rationalistic language" about rational strategies or beliefs can find that the relative-truth values of contingent statements like "*strategy a is rational for player i is true if and only if the components of the strategy profile b of the other players are mutually rational for each other player* $j \neq i$*, when player i adopts the "meta-decision problem" of these players simultaneously*" is true. [9] Of course, such a tautology corresponds precisely to a Nash equilibrium of the game.

This is the central idea of the paper: The initial indeterminacy of what constitutes a rational strategy leads each player to break the Gordian Knot of endless chain of contingent statements by "self-interacting" in a Nash equilibrium. To recap: On one hand, the underdetermination of the classical game model endows players of free will, in the sense that the definition of a game does not impose players to hold some particular beliefs. On the other hand, players are free to choose between possible strategies. Add rationality and a player becomes forced to follow a Nash equilibrium in order to arrive at determining a (rational) strategy. A Nash equilibrium appears then as a *self-interactive* principle accounting for the mental process by which *each* player can reach a decision in a game, from an initially nonexistent strategy-belief pair in his mind.

Our main Theorem gives a formal statement of this idea by proving that the assumption of free will of the classical game allows to obtain some necessary and sufficient conditions for a Nash equilibrium. This shifts the usual interpretation of an equilibrium: Instead of describing the end-point of complex (social) strategic interactions, a Nash equilibrium accounts for the *internal process of choice* wherein each individual self-interacts in a consistent way in order to determine his *own* rational behavior. Arguably, this motivates a neuroeconomics approach and process-based models of decision making in games (see, e.g., Giacomo et al. (2008)): unlike conventional wisdom, our main result shows that cognition is in fact an inherent feature of the rational choice paradigm. Moreover, this responds to the longstanding

[8] More precisely, the picture says ' "*If you pick this statement, then you must also pick this one and vice versa*". Hence, the picture does *not* guide the player on a particular Nash equilibrium.

[9] Hereafter, we will use the terms "relative" or "contingent", interchangeably.

criticisms addressed to this solution concept (see e.g., Bernheim (1984) and Pearce (1984)).[10] This new way of looking at Nash equilibrium is in stark contrast with the conventional wisdom. In the earlier view, a Nash equilibrium is interpreted as an "*interactive*" equilibrium point which has an "*as if*" interpretation (Nash [1950, p. 21-23]). Instead, we prove that the Nash equilibrium concept is a "*self-interactive principle of determination*" followed by each player in order to unravel the initial indeterminism of the game model, as a consequence of his free will.[11] A direct consequence of this result is that if one wants to continue to view a Nash equilibrium as an "interactive or focal equilibrium point" (the term is taken from Schelling [1960, p.57]), then, the earlier studies of correlation in games à la Aumann (1974) become essential; if players play in such a focal equilibrium point, we must give a story for why the players have determined the same intrinsic Nash equilibrium.[12]

No necessary conditions for Nash equilibrium have been given to date. In particular, Aumann and Brandenburger [1995, p.1163] provide some "tight" epistemic sufficient conditions which are not necessary. As they point out "*It is always possible for the players to blunder into a Nash equilibrium "by accident"*." Here, the very notion of free will—the absence of any pre-determined beliefs—leads each rational player to self-interact in a Nash equilibrium. Hence, players must freely *determine* the same "intrinsic" equilibrium in order to effectively "blunder" into such a focal equilibrium point.[13]

The paper is organized as follows. The next section is an informal discussion of the main result to follow. The formal treatment is in Sections 3-8. Section 9 concludes. The proof of the main result has been relegated to an Appendix.

2. Preview of the main result

To fix ideas, consider the game of Figure 1. In this game, Ann chooses the row, Bob chooses the column, Charlie chooses the matrix. Each player must (simultaneously and independently) and *freely* choose to go to the North Pole or the South Pole in the sense that players do not have any pre-specified hierarchies of beliefs i.e. they have free will. For Ann and Bob, this is simply the "Battle of the Sexes;" their payoffs are no affected by Charlie's choice. For the sake of discussion, let us focus on the following Nash equilibrium of this game $(\frac{2}{3}\mathtt{N} \oplus \frac{1}{3}\mathtt{S}, \frac{1}{3}\mathtt{n} \oplus \frac{2}{3}\mathtt{s}, N)$. First, consider the implication of formally incorporating the Free Will assumption on the behavior of each rational player. Let us focus our discussion on Ann. The application of our first result (Theorem 1) states that if Ann is rational and has no pre-specified beliefs, then she will play an equilibrium mixed strategy i.e., $\frac{2}{3}\mathtt{N} \oplus \frac{1}{3}\mathtt{S}$ in the game[14]: In the classical model, the absence of pre-existing rational strategies and beliefs

[10] In their work on rationalizability, Bernheim (1984) and Pearce (1984) have argued that Nash equilibrium behavior cannot be deduced solely from assumptions regarding the rationality of players and their knowledge of the rationality of their opponents.

[11] This might explain why the epistemic literature has regarded a Nash equilibrium as a particular case lacking of a crisp foundation (see, Brandenburger (2007)).

[12] More precisely, an extrinsic or intrinsic (see, Brandenburger and Friedenberg (2008)) source of correlation across players is needed in order to explain how a group of independent players with free will can determine the same *intrinsic* equilibrium point.

[13] Of course, this claim is valid for games where different equilibria have different strategy combinations for all players. As shown below, the extension of this result to all games is straightforward.

[14] The Theorem applies to any finite n-person normal-form game.

TABLE 1
Figure 1.

	n	s		n	s
N	2,1,0	0,0,0	N	2,1,0	0,0,0
S	0,0,0	1,2,0	S	0,0,0	1,2,0
	N			S	

implies that Ann must literally create what she deems as being rational. Doing so means that she must *simultaneously* put herself in the shoes of Bob and Charlie while she is considering her own decision problem. Since Ann is rational, she must be rational in her own perspective A_A as well as when she is considering Bob and Charlie's decision problems at $A_{B\otimes C}$. The gist is that Ann being a single person, she can only hold some self-referential strategy-belief. Hence, the very existence of a rational strategy for Ann coincides with a logical equivalence of a strategy-belief pair. To see this, note that when she has to choose a destination all statements like "The North Pole is rational","*The South Pole for Charlie and the North for Bob are rational beliefs for Ann*" or "*The South Pole for Charlie and the North Pole for Bob are rational*" and so on, are all indeterminate statements. In general, the rationality of Ann leads her to determine the truth-value of statements like "*The (mixed) strategy σ_A is rational*" is true because the statement "*The (mixed) strategy σ_C for Charlie and the (mixed) strategy σ_B for Bob are rational strategies*" is true. Hence, this in turn requires that Ann determines whether statements like "*The (mixed) strategy σ_C for Charlie and the (mixed) strategy σ_B for Bob are rational strategies*" are true or not.
This raises the question: What are the "mental states" that Ann has to take in order to break this endless chain of reasoning?
In order to cut this Gordian Knot, Ann must *simultaneously* put herself, "in the shoes" of Bob and Charlie (the meta-perspective $A_{B\otimes C}$) *and* in her own shoes (her own perspective A_A). Here is a visual way to grasp this situation more clearly.
Imagine that Ann is sitting simultaneously in two different *transparent* "cubicles" (the term is taken from Kohlberg and Mertens [1989, p.1005]) A_A and $A_{B\otimes C}$. Now, note that the transparent walls of the cubicles allow Ann to determine that the statement "$\frac{1}{3}\mathtt{n} \oplus \frac{2}{3}\mathtt{s}$ *is rational for Bob and the South Pole is rational for Charlie*" is true in her cubicle $A_{B\otimes C}$, because Ann can check—by taking a look at cubicle A_A—that the statement "$\frac{2}{3}\mathtt{N} \oplus \frac{1}{3}\mathtt{S}$ *is rational for Ann*" is indeed true in the corresponding cubicle, A_A, relative to the statement that "$\frac{2}{3}\mathtt{N} \oplus \frac{1}{3}\mathtt{S}$ *is rational for Bob and the South Pole is rational for Charlie*" is true in cubicle $A_{B\otimes C}$, whose she knows to be true by looking from her cubicle A_A, and so on. The bottom line of this story is thus that the mere determination of a rational strategy induces Ann to find a Nash equilibrium of the game.
Alternatively, we can also examine the situation of Ann if she adopts (simultaneously) the perspectives of all player, $A_{A\otimes B\otimes C}$. In this case, Ann is sitting in her "big cubicle" in which she simultaneously assigns truth-values to statements for all players. By definition, in this big glass cubicle she can only make *absolute* statements about all players. But Ann is unable to check—by looking through the windows of her cubicle—that a given meta-statement, i.e., a profile of strategies, is indeed rational relative to itself. For this she needs the perspective of another cubicle. Formally, Ann cannot check that a given profile is indeed a fixed point of the combined best-response mapping.

To recap: Cutting the Gordian Knot of an endless chain of indeterminate relativistic statements makes Ann's free will determine a rational strategy by interacting with herself (self-interacting) in a Nash equilibrium. This is the gist of our first result.
Below, we exclude games where some players may use the same strategy in two different Nash equilibria. As shown in the formal analysis, this additional restriction can be easily dropped. We then have the following (Theorem 1):[15] *Suppose we have a finite n-person game in strategic form where each player is rational without any pre-specified hierarchies of beliefs and knows that the others are rational. Then σ is a Nash equilibrium profile in the game being played if and only if each player determines the same intrinsic Nash equilibrium, σ, by using at least two perspectives.*
Theorem 1 gives some *necessary* and *sufficient* conditions for a Nash equilibrium when players can freely choose their strategies and beliefs: When players are not endowed with beliefs, they must form them by mentally self-interacting in a Nash equilibrium. So, this result provides the minimal (normative) process describing how the expectations are computed by the players and how the equilibrium state can be concretely achieved. Alternatively put, this theorem proves that a Nash equilibrium is the inevitable decision-theoretic principle used by each rational player endowed of free will i.e. without any pre-existing beliefs in his mind. Let us stress that Theorem 1 gives necessary conditions for a Nash equilibrium viewed as an interactive solution concept. This is so, since players cannot "blunder" into such a focal point if one of the players does not "self-interact" mentally in the same equilibrium point in order to form his initially nonexistent beliefs. However, the main message of Theorem 1 is that a Nash equilibrium is *not* an interactive solution concept but a *self-interactive* principle, which provides player with a guidance to shape his beliefs. To see this more clearly, note that one could transport the Aumann-Brandenburger's result (1995) in the mind of a single player, and interpret the induced Nash equilibrium as belonging to the player. But, even in this case, this would just indicate that the induced "intrinsic" Nash equilibrium follows from the common knowledge of the complete *pre-determinated* conjectures of the player with itself, while our result proves that this is the very *absence* of such pre-existing conjectures that leads to this result. Hence, what is new in Theorem 1 is that this "internal process of reasoning" is directly implied by rationality, when we refrain from assuming some extraneous beliefs. Equivalently put, without assuming a complete pre-determination of the players' conjectures i.e. when we incorporate the explicit under-determination of the game model, then the lack of deliberation of a player in a Nash equilibrium leads to the *absence* of choice. In short, an "(intrinsic) Nash equilibrium" accounts for the transition from the initially non-existent states of mind of a player to an effective *mental* choice of a rational strategy. The upshot is that Theorem 1 is therefore at odds with the seminal literature see e.g., Berhneim (1984) and Pearce (1984), where rationality alone does not imply Nash equilibrium, when interpreted as a "focal point". Indeed, our equivalence result allows to answer one of the questions posed by the foundations of game theory : "*Why should we expect Nash equilibrium play (players choosing best response strategies to the choices of others)*?" Samuelson (2002). Our answer is that if players have *bona fide* free will, then we do not need to assume some underlying form of communication between players to justify the "*assumptions that agents are rational*

[15] All our results could be extended to Euclidean games and some other usual classes of games with infinite strategy spaces. This restriction allows us to avoid all measure-theoretic issues.

and they all have common knowledge of such rationality and beliefs", since in this case the existence of a rational choice of one single player implies his self-interaction into a Nash equilibrium. Note in particular, that unlike the extent literature see e.g. Bernheim (1984), Pearce (1984) and Aumann and Brandenburger (1995), Theorem 1 states that in games with a *unique* Nash equilibrium, rational players endowed of free will have to mentally "blunder" on this unique equilibrium.

What if we want to continue viewing the Nash equilibrium as an interactive solution concept? Theorem 1 indicates that the usual interpretation of "equilibrium points" construed as "focal points" arises when all players follow a same pattern of reasoning in their deliberation (or determination) of their own behavior. So, Theorem 1 lends support to the various interpretations of a Nash equilibrium as a norm of behavior, a convention or a focal point, see e.g. Bicchieri (2006), Lewis (1967), Sugden (1986), Schelling (1960).[16]

Another important consequence of Theorem 1 is to give a conceptual principle for capturing the role played by reasons in rational decision making. Theorem 1 says that a rational player with free will does *not* seek to forecast the behavior of other players *per se*, but rather seeks a '(rational) *reason* to act in a particular way. Hence, a Nash equilibrium encapsulates a *reason-based theory of rational choice*, which establishes a connection between formal rational choice theory and philosophical work on reasons (see Dietrich and List (2011)).

The incorporation of players' free will entails that a Nash equilibrium occurs in the *mind* of *each* individual. From this perspective Theorem 1 complements Perea (2007). Unlike Perea, our analysis of an epistemic model from a single player's perspective is not the starting point of our analysis. Instead, it is the direct consequence of the very absence in the classical game model of any pre-specified beliefs. Hence, our main result shows that the absence of any pre-determined hierarchies of beliefs shifts the usual interpretation of an equilibrium as a "social" or interactive equilibrium point to an "individual" equilibrium point. This is so since without assuming some extraneous beliefs in the classical game model, Theorem 1 uncovers an *equivalence* between the choice of a rational player endowed of free will and his self-interaction in a Nash equilibrium.

Historically, von Neumann did not receive Nash's idea positively. (See Shubik [1992, p.155]). Nash's proposal seemed to be at odds with the "*enormous variety of observed stable social structures*" [1955, p.25] that von Neumann perceived as a natural state of affairs. Unexpectedly, our main result indicates that a Nash equilibrium responds to von Neumann's picture: By setting as its task to resolve what a rational individual should play, Nash responds in one shot to the problem of the multiplicity of the "*standards of behavior*" [1944, p.42] posed by von Neumann: The introduction of a "determination principle"—the so-called Nash equilibrium concept—guides each player in his resolution of the initial indeterminacy of the game model, thereby inducing each player to "self-interact" in one of those "*standards of behavior*".

[16]For example, in Schelling (1960) agents focus on a particular equilibrium because it is more concspicuous than the others.

3. Rationalistic models of a game

3.1. Finite games

Let $N = \{1, ..., n\}$ be a finite set of players with $n \geq 2$. An n-person finite normal-form game of complete information, interpreted as one-shot games is given by $G = \langle S^1, ..., S^n; \pi^1, ..., \pi^n \rangle$ where each set S^i consists of m_i pure strategies, with typical element, s^i, available to player i, and $\pi^i : \Pi_{i=1}^n S^i \to \mathbb{R}$ is i's utility function. We adopt the convention that $S = \Pi_{i=1}^n S^i$, $S^{-i} = \Pi_{j \neq i} S^j$ and $S^J = \Pi_{j \in J} S^j$ where $J \subset N$. The set of mixed strategies of player i is thus the $(m_i - 1)-$dimensional unit simplex $\Delta^i = \{\sigma^i \in \mathbb{R}_+^{m_i} : \sum_{s^i \in S^i} \sigma^i(s^i) = 1\}$ and $\Delta = \times_{i=1}^N \Delta^i$ is the polyhedron of mixed-strategy combinations $\sigma = (\sigma^1, ..., \sigma^n)$ in the game. We identify each pure strategy $s^i \in S^i$ with the corresponding unit vector $a_{s^i} \in \Delta^i$. When $J \subset N$, we set $\Delta^J = \times_{j \in J} \Delta^j$ with $\sigma^{-J} := (\sigma^k)_{k \notin J} \in \Delta^J$ and as usual $\Delta^{-i} := \Delta^{N \setminus \{i\}}$. We extend π^i to Δ in the usual way: $\pi^i(\sigma^i, \sigma^{-i}) = \sum_{s^i \in S^i} \sum_{s^{-i} \in S^{-i}} \sigma^i(s^i) \sigma^{-i}(s^{-i}) \pi^i(s^i, s^{-i})$ with $\sigma^{-i} \in \Delta^{-i}$. The mapping $\pi^J : S^J \to \mathbb{R}^J$ gives the payoffs of pure strategy combinations of players $j \in J$, and its extension $\pi^J : \Delta^J \to \mathbb{R}^J$ is defined in the obvious manner. The support of some mixed strategy $\sigma^i \in \Delta^i$ is denoted by $\mathtt{supp}(\sigma^i) = \{s^i \in S^i : \sigma^i(s^i) > 0\}$.
We will also use the usual notions of domination. Consider a pure strategy $s^i \in S^i$. We say that $s^i \in S^i$ **strictly dominates** or is **strongly dominant** if

$$\forall s^{-i} \in S^{-i}, \pi^i(s^i, s^{-i}) > \pi^i(s'^i, s^{-i}), \forall s'^i \in S^i \setminus \left\{s^i\right\}.$$

We say that a strategy $\sigma^i \in \Delta^i$ is **never a best response** for i if

$$\forall \sigma^{-i} \in \Delta^{-i}, \exists \sigma'^i \in \Delta^i, \pi^i(\sigma'^i, \sigma^{-i}) > \pi^i(\sigma^i, \sigma^{-i}).$$

Last, for each strategy combination $\sigma \in \Delta$,

$$\mathtt{BR}_i(\sigma) = \left\{\sigma^i \in \Delta^i : \pi^i(\sigma^i, \sigma^{-i}) \geq \pi^i(\sigma'^i, \sigma^{-i}) \ \forall \sigma'^i \in \Delta^i\right\},$$

denotes the mixed best replies of player $i \in N$. In the rest of the paper we always consider the mixed-strategy extension of G.

3.2. Rationalistic frames of a player

The idea that players can shape their beliefs by putting themselves "into the shoes of others" is not new. For example, Luce and Raiffa [1957, p.306] were among the earliest to suggest such a process:

"The problem of individual decision making under uncertainty can be considered a one-person game against a neutral nature. Some of these ideas can be applied indirectly to individual decision making ... where the adversary is not neutral but a true adversary. ... One modus operandi for the decision maker is to generate an a priori probability distribution over the ... pure strategies ... of his adversary by taking into account both the strategic aspects of the game and ... psychological information ... about his adversary, and to choose an act which is best against this ... distribution."

The structure defined below aims at giving a formal content to this idea. According to Luce and Raiffa, a player adopts *two points of views* by treating, simultaneously in a mutually consistent way, the decision problems of all the other players to form his beliefs on one hand and his own decision problem on the other hand. However, although this canonical dual structure may sound pretty natural at first sight, there is no reason, *a priori*, to exclude some other structures of reasoning. Note that the idea that game-theoretic solution concepts should be evaluated by defining a classes of models has been explored by Stalnaker (1994). The Kripke model developped in this paper is thus very much in line with this methodology. The following definitions set the stage of a general structure which allow to encompass the set of *all* possible structures of reasoning. In our analysis, each player in a game can adopt an arbitrary set of multiple viewpoints about the other players—some of them may even overlap. We do so, via models that have now become standard for modal logics, *viz.* so-called Kripke models (see, Kripke (1963)). Given a game $G = \langle S^1, ..., S^n; \pi^1, ..., \pi^n \rangle$, a **rationalistic frame** for player i in G is the structure, $\mathcal{F}_G^i = \langle W, \mathcal{W}, \mathcal{R}^i \rangle$, where $W = \{w^{i_1}, w^{i_2}, ..., w^{i_n}\}$ represents the (non-empty) set of the n **atomic (informational) perspectives** corresponding to the n decision problems that can be considered by player i in the game G. Each w^{i_j} represents player i when he is mentally considering the decision problem—the perspective—of player j. We want a structure encompassing all the Kripke models that a player could possibly adopt in a game. So, we do not impose any requirement on $\mathcal{W}$ i.e., player i may consider some "non-partitional" perspectives and $\mathcal{W}$ can be any class of m (nonempty) subsets of W whose union is W.[17] Thus, if player i has a frame $\mathcal{F}_G^i$ with m different perspectives, then $M = \{S_l \subseteq N : l = 1, ..., m, \bigcup_{l=1}^m S_l = N\}$ represents the induced class of subsets of N. It is therefore natural to refer to each non-singleton cell $w^{i_J} := \{w^{i_j} : j \in J \subseteq M\} \in \mathcal{W}$ as the **meta-perspective** J of player i in G. Each cell $w^{i_J} \in \mathcal{W}$ must be thought of as the viewpoint of player i when he is simultaneously treating,(all) the decision problem(s) of player(s) $j \in J$ independently so that they are mutually rational. Hereafter, the generic term "**perspective**" will be employed for an atomic or meta-perspective.[18] The above abstract structure captures *all* the possible combinations of viewpoints that can be taken by a player in a game. Given a particular class M, we refer to the resulting structure $\mathcal{F}_G^i$ as a M-**frame**. As a particular case, a M-frame where $M = \{i, -i\}$ formalizes the above Luce and Raiffa's suggestion.

The interpretation of a M-frame is clear. This formalizes the fact that a player can adopt the decision problems of the other players in order to form some initially nonexistent beliefs in a consistent way, with itself. So a M-frame has a purely self-referential interpretation, which captures the aforementioned Luce and Raiffa's idea of introspection of a player in a game. So, from this perspective, a M-frame does *not* model a player who is trying to guess or forecast the strategies of the other players. Rather, each player is seeking a (rational) *reason* to act in a particular way. Finally, note the implications of the self-referential nature of a M-frame. When a (rational) player $i \notin J$ adopts a M-frame with a perspective w^{i_J}, he continues to be a rational player. This means that player i has to be rational *in the role of* other players J. This entails that the formation of the beliefs of player i at w^{i_J}, coincides with the determination of a rational strategy profile σ^J, whose all components will be *mutually rational* for players J or equivalently, for player i when he adopts the decision problems of players J.

[17] That M-frames must necessarily verify that $\bigcup_{l=1}^m S_l = N$ for each player i is obvious for otherwise, some players' decision problem(s) are not even considered by i.

[18] Note however that in a pure rationalistic world it would be more natural to confine the analysis to partitions of W so as to exclude frames wherein players hold some redundant (overlapping) perspectives.

As usual, the "knowledge" of player i at perspective w^{i_J}—we shall give a formal definition of this term in due course—is represented by the binary accessibility relation, $R^{i_J} \in \mathcal{R}^i$ over the cells of $\mathcal{W}$ i.e., $R^{i_J} \subseteq \mathcal{W} \times \mathcal{W}$. Each R^{i_J} is assumed to be a reflexive and antisymmetric accessibility relation.[19] Note that the antisymmetric property is consistent with the notion of perspective: If player i is in perspective w^{i_J}, then he cannot R^{i_J}-access w^{i_J} from another perspective $w^{i_K} \neq w^{i_J}$. [20] Hereafter, each structure $\langle w^{i_J}, R^{i_J} \rangle$, with $w^{i_J} \in \mathcal{W}$ is referred to as the $J-$**frame** of player i. Given the importance of the notion of a rationalistic frame in this paper, we record this notion in the following definition.

Definition 1 *We say that $\mathcal{F}^i_G = \langle \mathcal{W}, \mathcal{R}^i \rangle$ is the M-**rationalistic frame** of player i in game G if each J-frame of i for $J \in M$, $\langle w^{i_J}, R^{i_J} \rangle$, with $w^{i_J} \in \mathcal{W}$ is such that $R^{i_J} \in \mathcal{R}^i$ is a reflexive and antisymmetric binary accessibility relation over $\mathcal{W}$.*

We refer to the rationalistic M-frame $\mathcal{F}^i_G = \langle \mathcal{W}, \mathcal{R}^i \rangle$ with $\mathcal{W} = \{w^{i_i}, w^{i_{-i}}\}$ as the **canonical rationalistic frame** of player i. Hereafter, this frame will be thought of as whole class of M-frames with two perspectives i.e., the equivalence class of all frames with $|M| = 2$ such that $S_1 \bigcup S_2 = N$.

3.3. The logical mental language of players

Probability theory is an integral part of game theory because randomized strategies play a significant role in the existence of "equilibrium points". Thus, there is a natural correspondence in game theory between propositions used to describe a player i's behavior and his mixed strategies of the unit simplex spanned by S^i, and between logical operations on the propositions and set-theoretical operations on the corresponding sets and between logical operations on the propositions and set-theoretical operations on the corresponding sets. More specifically, , in a game G, the decision problems solved by player i at perspective $w^{i_J} \in \mathcal{W}$ pertain to "rationalistic" statements on what constitutes a rational strategy for each player $j \in J$ in G. Hence, for each player $j \in N$, we define a set of statements, $\mathbb{A}^j$. Each element $A^j \in \mathbb{A}^j$ represents a statement like $A^j :=$ "*strategy σ^j is optimal*[21] *in G*", without any reference to the strategies of players $k \neq j$. We shall refer to statements A^j as the **atomic statements** of player j in G. In the sequel, we set $\mathbb{A} = \mathbb{A}^1 \bigcup \mathbb{A}^2, ..., \bigcup \mathbb{A}^n$. We shall use the usual metalingustic abbreviation: $A \wedge A'$ for $\neg(\neg A \vee \neg A')$ with $A, A' \in \mathbb{A}$. We will introduce some modalities in due course.

In a rationalistic frame, each player i abides by the non-cooperative axioms of game theory. In the (important) particular case of the canonical model where player i has a $-i$-frame, i can *simultaneously* and *independently* handle the decision problems of players $j \neq i$ by considering the viewpoints of those players with formulae A^{-i} like "*strategies σ^j, with $j \neq i$ are mutually rational between players $j \neq i$*", without any reference to strategy $\sigma^i \in \Delta^i$. Each strategy can be thought of as an hypothetical statement and rational responses (the best replies) are interpreted by the material implication $\rightarrow$. Hence, A^{-i} can be construed as a conjunction of conditional statements . Of course, our use of the Lukasiewicz three-valued logic

[19] A relation $R \subseteq S \times S$ on a set S is reflexive if $\forall x \in S, xRx$ and antisymmetric in that xRy and yRx imply $x = y$.

[20] This is only a consistency requirement. It does not play a role in any of our results.

[21] We use the term "optimal" and "rational" interchangeably.

requires that we replace this classical connective by the extended implication of Lukasiewicz. Thus, as it is customary in this literature, we shall use the following (metalinguistic) abbreviation: $A \to_{\text{L}} B$ for $\neg A \vee B$ where $\to_{\text{L}}$ stands for the *Lukasiewicz implication*. The truth table of this connective is shown in Table 2. We extend the domain of relativistic valuations by considering the set of formulae $\mathbb{F}(w)$ obtained from the set of atomic statements $\mathbb{A}(w)$ by closing with respect to negation, disjunction, conjunction $\{\neg, \vee, \wedge\}$ and some modalities that we shall define in due course. With this rationalistic language and connectives in mind, we can therefore write compound statements A^{-i} as $A^{-i} := \bigwedge_{j \in N \setminus \{i\}} (\wedge_{k \in N \setminus \{i,j\}} A^k \to_{\text{L}} A^j)$. Let $\mathbb{A}^{-i}$ denote the set of rationalistic formulae like A^{-i}. More generally, if player i has a J-rationalistic frame, then he adopts simultaneously the viewpoints of the subgroup $J \subseteq N$ of players in a mutually consistent way. Thus, in such a perspective w^{i_J}, player i writes formulae A^J like $A^J :=$"*strategies σ^j, with $j \in J$ are mutually rational for players J*", without any reference to strategy profile $\sigma^{-J} \in \Delta^{-J}$. Given a subset $J \subseteq N$ of players, $\mathbb{A}^J$ denotes the set of formulae like $A^J := \bigwedge_{j \in J} (\wedge_{k \in J \setminus \{j\}} A^k \to_{\text{L}} A^j)$. The important point, is to keep in mind that when player $i \notin J$ considers formulae A^J at perspectives w^{i_J}, he is not trying to guess the others' strategies, but rather he is forming his *own* (subjective) beliefs with itself i.e. he is trying to find a rational reason to make a particular choice.

4. Relativistic valuations and (relative) truth

4.1. Free will, relativistic valuations and "context of reasoning"

The mental language of a rational player formalizes the correspondence between propositions used to describe a player i's behavior and his mixed strategies of the unit simplex spanned by S^i. So, for example, in the "number guessing game", the property P that a player's pure strategy is below 11 corresponds to the subset of pure strategies $\widetilde{S}^i \subset S^i$ for which P is true, Its negation P, $\neg P$, corresponds to the complement $\widetilde{S}^i$ of the set $\widetilde{S}^i$, the points for which P is false (i.e., the number lies outside the specified range).
The reason for mentioning such elementary matters is that the situation becomes radically different if we incorporate the free will of players i.e. the absence of any pre-existing strategies and beliefs in their minds; in this case, prior to a choice in an actual experiment, a statement like $P :=$ "*choosing a number below* 11 *is rational for i*" is initially neither true, nor false, since player i has not yet determined such a statement as being true or false in his mind i.e. he has free will.
As illustrated in the above example, the discussion of how a player can assign a sharp truth-value to such initially indeterminate statements leads to a Kripke model where the truth-value at one perspective of a player will depend on the truth-value at another perspective i.e. the truth-values have only a *relational* meaning. Alternatively put, we formalize the idea that the truth-value of a statement made at one perspective depends upon the the other truth-values assigned at the other perspectives. The upshot is thus that instead of having the usual "absolute" valuation mapping, $V : \mathbb{A} \to \{0, 1\}$, that assigns a sharp truth-value to any given statement, we have some *contextual relativistic valuations*. The term "relativistic" captures the fact the sharp-truth values at one perspective will depend on the truth-value assigned by the player at his other perspectives. Contextuability refers to the fact that the valuation mappings depend on the particular M-frame used by the player i.e. on the number

and the particular class of perspectives $\mathcal{W}$ adopted by a player.
As a result, given a M-**rationalistic frame**, it is then convenient to denote the set of formulae $\mathbb{A}^L$ considered at perspective $w^{i_L} \in \mathcal{W}$ by $\mathbb{A}(w^{i_L})$ and a statement A^L by $A(w^{i_L})$. For models where $w^{i_J}, w^{i_L} \in \mathcal{W}$ we set $A^{-J} := (A(w^{i_L}))_{L \neq J}$. Let $\mathcal{F}_G^i = \langle \mathcal{W}, \mathcal{R}^i \rangle$ be a frame for player i in a game G with a pair of (ordered) perspectives $(w^{i_J}, w^{i_K}) \in \mathcal{W} \times \mathcal{W}$ and $\overline{A(w^{i_J}, w^{i_K})} := (A(w))_{w \in \mathcal{W} \setminus \{w^{i_J}, w^{i_K}\}} \in \times_{w \in \mathcal{W} \setminus \{w^{i_J}, w^{i_K}\}} \mathbb{A}(w)$.

Definition 2 *Let $\mathcal{F}_G^i = \langle \mathcal{W}, \mathcal{R}^i \rangle$ be a rationalistic frame for player i with (w^{i_J}, w^{i_K}) a pair of (ordered) perspectives such that $w^{i_J} \neq w^{i_K}$. A $\overline{A(w^{i_J}, w^{i_K})}$*−***relativistic valuation*** *is a map,*

$$V_{w^{i_J}, w^{i_K}}(\cdot, \cdot; \overline{A(w^{i_J}, w^{i_K})}) : \mathbb{A}(w^{i_J}) \times \mathbb{A}(w^{i_K}) \rightarrow \{0, \boldsymbol{i}, 1\},$$

where $\mathtt{i}$ means "indeterminate" in the sense of the three-value logic of Lukasiewicz (1930) (see the three-value truth tables for this logic in Table 2). Again, the absence of a truth-value $\mathtt{i}$ formalizes the fact that the choice of a strategy or beliefs do not initially exist in the minds of players, as stipulated by the classical game model. The relativistic valuation has also a clear-cut interpretation. For example, in the case of a canonical frame, $V_{w^{i_i}, w^{i_{-i}}}(A^{\sigma^i}, A^{\sigma^{-i}})$ gives the truth-value of statement A^{σ^i} when i considers his own decision problem at w^{i_i}, given the truth-value of statement $A^{\sigma^{-i}}$ that has been determined (or not) at $w^{i_{-i}}$, when he puts himself into "the shoes of the others". Of course, in the absence of any introspection, the definition of the classical game model will imply that $V_{w^{i_i}, w^{i_{-i}}}(A^{\sigma^i}, A^{\sigma^{-i}}) = \mathtt{i}$ (see Section 5).
Henceforth, $\mathbb{V}_{\mathfrak{r}}$ denotes the class of all such (extended) relativistic valuations. Given a rationalistic frame $\mathcal{F}_G^i$ for player i in G, we call a **Kripkean model** for player i the structure, $\mathcal{M}_G^i = \langle \mathcal{F}_G^i, \mathbb{V}_{\mathfrak{r}} \rangle$. In the special case where $\mathcal{F}_G^i$ is the canonical frame, we say that $\mathcal{M}_G^i$ is the **canonical Kripkean model** of player i in G. In this case, the corresponding $\overline{A(w^{i_J}, w^{i_{-J}})}$−relativistic valuations for $J = i, -i$, is denoted by $V_{w^{i_J}, w^{i_{-J}}}(\cdot, \cdot)$ and called the **canonical relativistic valuation** of player i_J.

Definition 3 *We say that A^J is* ***absolutely true*** *(resp. false) in $\mathcal{M}_G^i$ if*

$$V_{w^{i_J}, w^{i_K}}(A^J, A^K; \overline{A(w^{i_J}, w^{i_K})}) = 1, \ \textit{(resp. } 0\textit{)},$$

$\forall A^K, \forall \overline{A(w^{i_J}, w^{i_K})}$, *and* $\forall R^{i_J} \in \mathcal{R}^i$.

In short, statement A^J is an absolute truth when the component of the corresponding strategy profile σ^J are always mutually rational when player i adopts the role of players J in *any* possible M-frame with $w^{i_J} \in \mathcal{W}$. The most trivial example is when statement A^j is absolutely true which is tantamount to saying that σ^j is a strongly dominant action in the game for player j. In other words, a statement is always true when it does not depend on the other truth-values assigned to the statements at the other perspectives. In this case, a player at perspective w^{i_J} does not need to consider the other perspectives via the accessibility relation R^{i_J}.
Note that this notion of "absolute truth" can be construed as reflecting a *lack* of free will from the part of the players. Indeed, in this case, players do not need to reason to determine what is their best course of action; their rational choice is imposed from the outset by the game structure. Of course, the existence of absolute (true or false) truths are highly non

generic. Hence, we can say that players have generically a real free will in games, in the sense that they can freely change their mind on different pairs of strategies-beliefs, and their resulting choices made in an experiment are not a function of some history. As a result, the under-determination of the game model calls for the more general notion of a"determined relative (sharp) truth-values". Such a notion will permit to formally capture the transition from an initially nonexistent sharp truth-value to the assignment of a sharp truth-value to statement about what can be regarded as a rational choice in the game being played. This notion is formally defined, in an inductive manner, as follows.

Definition 4 *(Determined relatively true statements) Let $\mathcal{F}_G^i = \langle w^{i_L}, R^{i_L} \rangle_{L \in M}$ be the M-rationalistic frame of i in a Kripkean model $\mathcal{M}_G^i$ and set $A^{-J} := (A(w^{i_L}))_{w^{i_L} \in \mathcal{W} \setminus \{w^{i_J}\}}$. Suppose each component of profile $(A(w^{i_L}))_{w^{i_L} \in \mathcal{W}}$ is neither absolutely true, nor absolutely false. We say that $A(w^{i_J})$ is **determined** as being A^{-J}-**relatively true** in $\mathcal{M}_G^i$ at w^{i_J} if:*
(1) $w^{i_J} R^{i_J} w^{i_L}$ hold $\forall L$;
(2) $V_{w^{i_J}, w^{i_K}}(A^J, A^K; \overline{A(w^{i_J}, w^{i_K})}) = 1$ and;
(3) Properties (1)-(2) are satisfied for each statement $A(w^{i_L})$, $w^{i_L} \in \mathcal{W}$ of profile $A^{-J} := (A(w^{i_L}))_{w^{i_L} \in \mathcal{W} \setminus \{w^{i_J}\}}$.

This definition is crucial for it models the fact that the initially nonexistent future rational choice of belief in the mind of a player, as implied by the definition of a game "comes into being" in the mind of the player. In words, this says that a statement A^J made by player i when he considers the decision problems of players $J \subset N$ i.e. i is at perspective w^{i_J}, is determined as being A^{-J}-relatively true if i knows that the profile of statements A^{-J} formulated at the other perspective of i's M-frame are also determined as being (A^J-relatively) true. Intuitively, this definition allows to give a formal content to the idea that truth-values that pertain on statements in the game model can only have a "becoming" interpretation; the timeless structure of the game model precludes that the choice made by a player is a function of the past—as in the case of the introduction of a hierarchy of beliefs representing the players' prior experiences. This notion of a "determined truth" of a rational action is therefore consistent with the old philosophical idea that statements in games must be treated as "future contingents": truth-values only exist for describing the *actual state of mind* of a player. The above inductive definition therefore leads to the following Lemma which is a mere restatement of the above definition.

Lemma 1 *Statement $A(w^{i_J}) \in \mathbb{A}(w^{i_J})$ is determined as being A^{-J}-relatively true in $\mathcal{M}_G^i$ at w^{i_J} with $A^{-J} := (A(w^{i_L}))_{w^{i_L} \in \mathcal{W} \setminus \{w^{i_J}\}}$ iff there exists a set of valuations, $V_{w^{i_K}, w^{i_L}} \in \mathbb{V}_{\mathfrak{r}}$, such that profile $(A^J(w^{i_J}), A^{-J})$ verifies that $\forall K$, $w^{i_K} R^{i_K} w^{i_L}$ hold $\forall L$, and*

$$V_{w^{i_K}, w^{i_L}}(A^K, A^L; \overline{A(w^{i_K}, w^{i_L})}) = 1,$$

$\forall (w^{i_K}, w^{i_L}) \in \mathcal{W} \times \mathcal{W}$ with $K \neq L$.

Proof. Obvious and therefore omitted.
Accordingly, statement $A(w^{i_J})$ is determined as **being relatively false** if $\neg A^J(w^{i_J})$ has been determined as being relatively true. As a result, we have that

$$V_{w^{i_K}, w^{i_L}}(A^K, A^L; \overline{A(w^{i_K}, w^{i_L})}) = \mathtt{i},$$

$\forall(w^{i_K}, w^{i_L}) \in \mathcal{W} \times \mathcal{W}$ with $K \neq L$ if A^K is neither absolutely true nor absolutely false and if it has not been determined as being true or false. Notice that the indetermination $w(A^N) = \mathtt{i}$ is always true as we rule out "trivial" games where some players have a strongly dominant action.

Thereafter, $\mathbb{F}$ denotes the set of formulae obtained from the set of *all* atomic statements by closing with respect to usual connectives and modal operators that we shall define below. Abusing of notations, let denote the set of **absolute valuations** by $w : \mathbb{F} \to \{0, \mathtt{i}, 1\}$, whenever $w \in \mathcal{W}$. In the sequel $\mathbb{V}_{\mathfrak{a}}$ denote the set of all such mappings.

The following notion of the determination of a statement "as being true" (i.e. rational) in a game formalizes the von Neumann and Morgerstern' s aforementioned observation that:

"given the same physical background different established orders of society or accepted standards of behavior can be built" [1944, p.42].

Definition 5 *(Truth and indeterminate formulae) In $\mathcal{M}^i_G = \langle \mathcal{F}^i_G, \mathbb{V}_{\mathfrak{a}}, \mathbb{V}_{\mathfrak{r}} \rangle$, a formula $F \in \mathbb{F}$ is true (resp. false) at $w \in \mathcal{W}$, denoted $w(F) = 1$, iff F is absolutely true or if F has been determined as being (relatively) true. A formula $F \in \mathbb{F}$ is false in $\mathcal{M}^i_G$ iff $\neg F$ is true. A formula F is indeterminate in $\mathcal{M}^i_G$, iff F is neither true nor false i.e., $w(F) = \mathtt{i}$.*

According to the above definition, a formula is rational i.e. true in the mind of player i if this formula is always rational, or if the introspection of player i implies that the formula is true, *given* his context of reasoning i.e. which profile of other formulae has been determined as being true at some other perspectives of i's M-frame. Hereafter, the notation $\mathcal{M}^i_G \vDash F$, (resp. $\mathcal{M}^i_G \vDash \neg F$,) means that formula F is true (resp. false) in $\mathcal{M}^i_G$. The upshot is thus that a strategy profile $\sigma^J = (\sigma^j)_{j \in J}$ (or equivalently, a profile of beliefs of i) is considered to be true by i because the underlying game G induces a deterministic behavior from players J (in the sense that an outside observer could predict with certainty profile σ^J), or because the game in under-determined in the sense that the outcomes of the game in an experiment depends on the context of reasoning of the players. For example, in the canonical model, if a pure strategy, σ^i, is represented by a relatively true statement, this means that the observation of this strategy in experiments is highly contextual, as it depends on which particular belief (or equivalently, strategy profile) σ^{-i} has been determined as being rational by i in his mind, when he puts himself into the shoes of the others. Clearly, this contextuality of reasoning will permit to understand why the von Neumann and Morgerstern's assertion of a variety of "standards of behavior" is inevitable when players have free will. The Lukasiewicz's logic abandons the law of the (future) excluded middle ($p \vee \neg p$) (see Table 2). This particular property, which violates the idea of Aristotle's future contingents has often been criticized (see e.g. Prior, 1953). However, in a game model, the failure of this law falls out naturally from the initial relativistic nature of the rationality concept: if the truth-values of a rationalistic statement A^i (and $\neg A^i$) need to be determined i.e. created, by player i, then the whole disjunction needs also to be determined by player i. Alternatively put, considering the disjunction A^i or $\neg A^i$ as a tautology, would entail that one of the two statements, "*strategy s^i of player i is rational*" or "*strategy s^i of player i is not rational*", has already been pre-determined as being true in the mind of player i, thereby contradicting his/her free will. Otherwise stated, the law of the excluded middle is meaningless in the game model, for the truth-value of any proposition used in the present logic is *not* determined by the truth-values of its parts. This is so, since the rationality concept implies that the truth-values have

TABLE 2
[Lukasiewicz' three-valued semantics]

$\vee$	1	i	0	$\wedge$	1	i	0	$\rightarrow_{\text{L}}$	1	i	0
1	1	1	1	1	1	i	0	1	1	i	0
i	1	i	i	i	i	i	0	i	1	1	i
0	1	i	0	0	0	0	0	0	1	1	1

(generically) only a relational meaning.[22]

5. Relativistic valuations of the classical game model

We now complete the definition of a rationalistic Kripkean model of a game G. First, some notations. For any *pair* of formulae $(F', F) \in \mathbb{F} \times \mathbb{F}$, let $\mathfrak{L} : \mathbb{F} \times \mathbb{F} \rightarrow \{0, \mathtt{i}, 1\}$ denote the truth function for the Lukasiewicz implication $\rightarrow_{\text{L}}$ (see Table 2). For example, in the canonical Kripkean model $\mathcal{M}_G^i$ of a game G, when $A^{-i} \in \mathbb{A}^{-i}$, $\mathfrak{L}(A^{-i}, A^i)$ is the truth value of the formula, $\wedge_{j \neq i} A^j \rightarrow_{\text{L}} A^i$. Hence, the mapping $\mathfrak{L}$ assigns a relative truth-value, $0, \mathtt{i}$ or 1 to every contingent statement "*σ^i is rational for player i in G given that all strategies σ^j, with $j \neq i$ are mutually rational between players $j \neq i$*".
Now, the definitions of the relativistic and absolute valuations fall out automatically as a consequence of the formal under-determination of the classical game model. The following properties (i)-(iii) summarize the logical interpretations of the classical game model.

Definition 6 *Let $\mathcal{F}_G^i = \langle w^{i_J}, R^{i_J} \rangle_{J \in M}$ be a rationalistic frame of player i in G. We say that $\mathcal{M}_G^i = \langle \mathcal{F}_G, \mathbb{V}_{\mathfrak{a}}, \mathbb{V}_{\mathfrak{r}} \rangle$ is a* ***rationalistic Kripkean model*** *for player i in G if*
(i) in every perspective $w^{i_J} \in \mathcal{W}$, the absolute valuation function, $w^{i_J}(\cdot)$, is such that

$$w^{i_J}(A^J) = \begin{cases} 1 & \textit{if} \quad \textit{each } A^j, j \in J \textit{ corresponds to a strongly dominant strategy in } G; \\ 0 & \textit{if each} \quad A^j, j \in J \textit{ corresponds to a never best response in } G; \\ \boldsymbol{i} & \textit{otherwise, if} \quad A^J \textit{ has not been determined as being true or false.} \end{cases}$$

(ii) $\underline{\text{for every pair}}$ $(w^{i_J}, w^{i_K}) \in \mathcal{W} \times \mathcal{W}$ with $w^{i_J} \neg \overline{R^{i_J} w^{i_K}}$ and every $(A^J, A^K) \in \mathbb{A}^J \times \mathbb{A}^K$, the $\overline{A(w^{i_J}, w^{i_K})}-$relative valuation $V_{w^J, w^K}(A^J, A^K; \overline{A(w^{i_J}, w^{i_K})}) = w^{i_J}(A^J)$;
(iii) suppose that $\forall J, w^{i_J} R^{i_J} w^{i_L}$, hold, $\forall w^{i_L} \in \mathcal{W}$, with each component $A(w^{i_L})$ of profile $A^{-J} := (A(w^{i_L}))_{w^{i_L} \in \mathcal{W} \setminus \{w^{i_J}\}}$ determined as being A^{-L}-relatively true. Then, the $\overline{A(w^{i_J}, w^{i_K})}-$ relativistic valuation is such that,

$$V_{w^{i_J}, w^{i_K}}(A^J, A^K; \overline{A(w^{i_J}, w^{i_K})}) = 1$$

iff $A^J \in \arg\max_{A'^J \in \mathbb{A}^J} \mathfrak{L}(\wedge_{L \neq J} A(w^{i_L}), A'^J)$.

Property (i) reflects the under-determination of the game model: in general we cannot answer a question like "*is strategy σ^i is rational?*", unless for the non-generic examples of strongly dominant actions or never best responses. Hence, absolute valuations accurately describe the truth-values—or its absence—of rationalistic statements, as initially given by the game

[22]Note that this explains why the objection of Prior (1953) does not apply in the present case.

model, independently of any "context" of reasoning. Thus, the signification of absolute valuations is clear; when $w^{i_J}(A^i) = 0$ (resp. $\mathtt{i}$, 1), this indicates that the behavior represented by statement A^i will not occur (resp. will be indeterminate, will occur) in all the possible "play" of game G i.e., in all "measurements" that could be made by an outside observer. Alternatively put, the sharp truth-values taken on by the absolute valuation can therefore be understood as the deterministic predictions that could be made by an outside observer. By contrast, the lack of such sharp-truth values, $\mathtt{i}$, is a generic feature of the game model capturing its initial under-determination i.e. the fact that the classical definition of a game does not generally encapsulate enough information to predict whether a given strategy is rational in a game. Property (ii) captures the fact that in the absence of reasoning, a player has initially no pre-existing strategy and beliefs in his mind. By "default", i.e., in the absence of introspection, each $w^{i_K} \in \mathcal{W}$ is *not* accessible from a perspective $w^{i_J} \in \mathcal{W}$ whenever $J \neq K$, which models the fact that player i does not consider the decision problem of players K. Hence, property (ii) expresses the fact that if player i does not even consider the other players' decision problems, then his relativistic valuation is vacuously defined i.e. player i does not incorporate the relativistic aspect of rationality in his reasoning, which accounts for the property (ii) that the relativistic valuation must coincide with the absolute valuation. Property (iii) is a mere restatement of Lemma 1. It says that player i can assign some sharp truth-values to initially nonexistent rational strategies and beliefs whenever he makes the effort to put himself "into the shoes of the others" in order to form his beliefs, in a mutually consistent way.

6. Self-interactive epistemology

A feature of the present "ontological" approach is that a player is naturally led to introspectively reason about himself in order to form his beliefs. This means that the usual machinery for talking about the beliefs and knowledge of players does no longer induce an interactive epistemology, but a "self-interactive epistemology" i.e. the absence of initial hierarchies of beliefs and the impossibility of communication across players imply that we need to talk formally about each players' beliefs and knowkledge about *himself*, when he puts himself into the "shoes of the others", rather than about the others. [23]

Given the accessibility relation R^{i_J}, the possibility correspondence of player i in perspective w^{i_J} is defined as

$$\mathcal{P}^{i_J}(w^{i_J}) = \left\{ w^{'} \in \mathcal{W} : w^{i_J} R^{i_J} w^{'} \right\}.^{24}$$

In order to establish the interpretations of events—that is subset of the perspectives $\mathcal{W}$—as propositions, we need to introduce the following standard definitions.

Given a frame $\mathcal{F}^i_G$, we add a function, $f : \mathbb{A} \to 2^{\mathcal{W}}$[25] that associates with every atomic statement $A \in \mathbb{A}$, the set of perspectives where A is true. For every formula $F \in \mathbb{F}$, the **truth set** of F in $\mathcal{M}^i_G$, denoted by $\|F\|^{\mathcal{M}^i_G}$ is defined recursively as follows:

[23] In epistemic game theory, it is standard to postulate a possibility correspondence $\mathcal{P}$, rather than a binary accessibility relation R. In this paper, it is more convenient to work with R. That the two notions are equivalent is well-known.

[24] Conversely, given a possibility correspondence $\mathcal{P}^{i_J}$, the associated accessibility relation R^{i_J} is obtained as follows: $\forall w^{'} \in \mathcal{W}, w^{i_J} R^{i_J} w^{'}$ iff $w^{'} \in \mathcal{P}^{i_J}(w^{i_J})$.

[25] As there is a finite number of players, the power set of $\mathcal{W}$, $2^{\mathcal{W}}$, is well-defined.

(1) If $F = (A)$ where A is an atomic statement, then $\|A\|^{\mathcal{M}_G^i} = f(A)$;
(2) If $\neg \|A\|^{\mathcal{M}_G^i} = \|\neg A\|^{\mathcal{M}_G^i}$;
(3) If $\left\|A \vee A'\right\|^{\mathcal{M}_G^i} = \|A\|^{\mathcal{M}_G^i} \bigcup \left\|A'\right\|^{\mathcal{M}_G^i}$;
(4) $\|\Box^{i_J} A\|^{\mathcal{M}_G^i} = \left\{w \in \mathcal{W} : \mathcal{P}^{i_J}(w) \subseteq \|A\|^{\mathcal{M}_G^i}\right\}$.

The intended interpretation of $\Box^{i_J} A$ is "*player i knows A at perspective w^{i_J}.*" If $w \in \|F\|^{\mathcal{M}_G^i}$ we say that F is true in $\mathcal{M}_G^i$ at perspective w. Thus, according to (4), player i in perspective w^{i_J} knows F if and only if F is true at every other perspective(s) that i considers at w^{i_J}. When A is an indeterminate statement, we have that $\|\Box^{i_J} A\|^{\mathcal{M}_G^i} = \emptyset$ and we set $w^{i_J}(\Box^{i_J} A) = \mathtt{i}$ in this case. [26] With these definitions in mind, we can formally discuss the knowledge of player i as follows.

If E is the truth set of some formula F (that is $E = \|F\|^{\mathcal{M}_G^i}$), and $K : 2^{\mathcal{W}} \to 2^{\mathcal{W}}$ is the knowledge operator, then $K_{i_J} E$ is the truth set of the formula $\Box^{i_J} E$, that is $K_{i_J} E := \|\Box^{i_J} E\|^{\mathcal{M}_G^i}$. Henceforth, we say that player i in perspective w^{i_J} **knows** event $E \subseteq \mathcal{W}$ (or more precisely, the statement represented by event E) if $w^{i_J} \in K_{i_J} E$. Notice that our definitions imply that the *truth axiom*[27] (also called *axiom of knowledge*) holds: If player i has determined a formula F as being true, then by definition F is true at every perspective of i. This justifies our use of the term knowledge rather than belief. Note also that if a formula F is absolutely true, then player i does not need to determine this formula and we have that $\|\Box^{i_J} F\|^{\mathcal{M}_G^i} = \{w^{i_J}\}$ whenever $w^{i_J} \neg R^{i_J} w'$ for all $w' \neq w^{i_J}$.[28] Let $M := \{J : w^{i_J} \in \mathcal{W}\}$. For every event E, $K_{\otimes_{J \in M} i_J} E := \bigcap_{J \in M} K_{i_J} E$, is the event that every player i_J knows E. When $w \in K_{\otimes_{J \in M} i_J} E$ we say that E is **mutual knowledge**, at perspective w. Thus an event E is known by player $i = \otimes_{J \in M} i_J$ if player i knows E in each of his perspectives, $w^{i_J} \in \mathcal{W}$. The common knowledge operator K_*^i is then defined as follows. For any operator $K_{\otimes_{J \in M} i_J}$ define $K^k_{\otimes_{J \in M} i_J}$, the k^{th} iteration of $K_{\otimes_{J \in M} i_J}$, as follows: For all E, $k \geq 1$, $K^0_{\otimes_{J \in M} i_J} E = E$ and $K^k_{\otimes_{J \in M} i_J} = K_{\otimes_{J \in M} i_J} K^{k-1}_{\otimes_{J \in M} i_J} E$. The event that E is **commonly known** by i is defined by

$$K_*^i E = \bigcap_{k \geq 1} K^k_{\otimes_{J \in M} i_J} E.$$

Thus, an event E is commonly known by player i in a rationalistic frame $\mathcal{F}_G^i = \{\mathcal{W}, \mathcal{R}^i\}$ if player i knows it in each of his perspective w^{i_J}, player i knows that he knows it in every of his perspectives, and so on, *ad infinitum*.[29]

[26] Whether we set $w^{i_J}(\Box^{i_J} A) = \mathtt{i}$ or 0 does not play a role in this paper.

[27] Indeed, our framework provides a justification for this axiom. When player i knows that a rationalistic statement A^i is relatively true at w^{i_i} this is precisely because player i has been able to resolve the initial indeterminacy of the game model. From this perspective, this paper contributes to the literature on the foundations of the introspective abilities of an agent (see, Gossner and Tsakas (2010)).

[28] Indeed, recall that a J-frame is defined with R^{i_J} reflexive.

[29] Note that because the number of players and hence perspectives is finite, the common knowledge operator is well-defined.

7. Main result: ontological conditions for Nash equilibrium

Next, we rule out pathological games where Nash equilibria exist in strongly dominant strategies for some players.[30] For ease of exposition we state Theorem 1 for games wherein players do not have the same mixed strategy in two different Nash equilibria.[31] For notational simplicity, we also consider a M-frame $\mathcal{F}_G^i$ where each player i's set of perspectives has been ordered .[32] Finally, event $E^{(\sigma^j)_{j\in N}}$ denotes the set of perspectives of player i where statement $A^{(\sigma^j)_{j\in N}}$ is true.

Theorem 1 *A profile $(\sigma^j)_{j\in N} = (\sigma^{i_L})_{L\in M}$ is a Nash equilibrium in a game being played G if and only if in the model $\mathcal{M}_G^i$ of each player $i \in N$ with frame $\mathcal{F}_G^i = \langle w^{i_L}, R^{i_L} \rangle_{L\in M}$ the statement $A^J :=$ " strategies $\sigma^{i_J} = (\sigma^j)_{j\in J}$, are mutually rational in G for any $i, j \in J \subset M$" has been determined as being relatively true at w^{i_J} in a Kripkean model of player i with at least two perspectives such that the event $E^{(\sigma^j)_{j\in N}}$ is common knowledge in $\mathcal{M}_G^i$ i.e., $K_*^i E^{(\sigma^j)_{j\in N}} = \{w^{i_L} : L \in M\}$ and $|M| \geq 2$.*

Proof. See the Appendix

We have already discussed the various implications of this result in our preview. However, several other important remarks are worth making. Theorem 1 indicates that if we leave players the freedom of their choice *and* beliefs, then we can give a full characterization of the Nash equilibrium—some sufficient and *necessary* conditions. Hence, while the common view is to interpret this concept as an interactive solution concept, Theorem 1 indicates that the notion of Nash equilibrium is a genuine decision-theoretic principle. By contrast, Aumann and Brandenburger (1995)'s result give the tight sufficient conditions if we think of a Nash equilibrium as describing the complex social interactions arising in a many-person world. But, implicit in this result, is some sort of communication between players. Indeed, as well put by Aumann and Brandenburger, "knowledge of what others will do is undoubtly a strong assumption".[1995, p.1176]. So, the *necessary* and *sufficient* conditions of Theorem 1 permit to bypass this additional assumption, which is not part of the description of a game. The bottom line is thus that the Nash equilibrium concept and the correlated equilibrium concept (Aumann, 1974, 1987) complement each others nicely; if one wants to capture the social interactions, the notion of correlated equilibrium trumps the notion of Nash equilibrium by incorporating some explicit form communication across players i.e. the existence of pre-defined beliefs are the shorthand for some "history", while the notion of Nash equilibrium is the appropriate concept in the absence of past communication beyond the game as given.

It is worth noting that the determination of a rational strategy calls for common knowledge of

[30] More precisely, the indetermination of player i arises when exactly $n-2 \geq m \geq 1$ out of $n \geq 3$ players $j \neq i$ have a strongly dominant action. A straightforward modification of the definitions of the notion of relative truth would allow to encompass these situations.

[31] The statement and the proof of Theorem 1 are indeed easily amended for games G wherein there exists different NE, $\sigma, \sigma' \in$ NE(G), with $\sigma^i = \sigma'^i$ for some player $i \in N$. To see this, it suffices to consider the set $\left\{A^{(\sigma'^j)_{j\in N}} : (\sigma'^j)_{j\in N} \in NE(G)\right\}$ modulo $\sim^i$ in $\mathcal{M}_G^i$, for $\sim^i$ defined as $(\sigma^j)_{j\in N} \sim^i (\sigma'^j)_{j\in N}$ iff $\sigma^i = \sigma'^i$.

[32] That is, $\mathcal{W} = \left\{w^{i_L} : L \in M\right\}$ with $M = \{\{1,2,...,j\}, \{j-k, j-(k-1), ..., j+1, j+2, ..., j+k\}, ...\}$ for some suitably chosen $j-1 \geq k \geq 0$.

the strategy choices at all the (mental) perspectives of the Kripkean model of each player.[33] So, the notion of common knowledge in Theorem 1 now appears as a mere consistency condition: it simply requires that a player know that he himself knows ...which strategy or beliefs he holds across his different perspectives. In short, Theorem 1 proves that instead of being an "as if" solution concept, the Nash equilibrium is the inevitable principle whenever one obeys the definition of the game model i.e. once one incorporate the free will of players.
Again, as already outlined in our preview, Theorem 1 *cannot* be construed as a mere restatement of Aumann and Brandenburger's result transported in the mind of a single player. This is so, since our decision-theoretic interpretation of the Nash equilibrium is *necessary* i.e. it arises as the inevitable consequence of the absence of pre-determined beliefs in the game model. Without such a self-interaction, a rational player could just not make any choice. One of the main consequences of our main result is therefore to shift the interpretation of the usual interactive epistemology as belonging entirely to the mind of a *single* player i.e. the epistemology becomes self-referential, rather than interactive. Thus, one of the main interests of our characterization is to show that we cannot think of a Nash equilibrium as an "as if" solution concept, but as a real principle of decision making, when agents have *bona fide* free will i.e. when they are not imposed some extraneous beliefs from the outset.
The essence of the epistemic approach of game theory is to append some hierarchies of beliefs to players in the game model. The introduction of these extra variables has been extremely fruitful to provide some clear-cut epistemic characterizations of many other solution concepts, the Nash equilibrium has remained hermetic to this analysis (see Brandenburger (2007)). Our main result suggests that the reason for this failure lies in the nature of the Nash equilibrium. Instead of describing the complex epistemic relations across players in a game, as the shorthand of their past interactions, a Nash equilibrium is a tenseless and ontological principle of determination. In this, it is therefore very different from the other concepts as it appears to be a principle congenital to the game model.

8. Self-referential rationalistic Kripkean models

A particular case of a M-frame is when $M = \{N\}$. In this case, a player would consider *all* the decision problems at once, simultaneously and in a mutually consistent way. However, as proven by Theorem 1, such a "meta-perspective" encompassing the whole set of decision problems cannot yield the determination of a rational strategy. To see this, let

$$\mathtt{BR} : \Delta \twoheadrightarrow \Delta$$

be the *combined mixed-strategy best-reply correspondence* of the game defined as the Cartesian product of all players' mixed-strategy best-reply correspondences, $\mathtt{BR}(\sigma) = \times_{i \in N} \mathtt{BR}_i(\sigma)$. Then, note that a M-frame with $M = \{N\}$ would reflect a situation where a player has to compute the combined best response of a game (the output), without even having pre-specified in his mind, which profile of strategies will be a fixed point of this mapping (the input). Rephrasing this situation in terms of the above example, this would amount to Ann determining the rational destination of the combined best reply mapping of the game while

[33] Here, common knowledge appears as a necessary condition because player i has to solve for a system of relativistic valuations which characterizes a Nash equilibrium.

she has not even defined in her mind a particular (mixed) destination that could be a fixed point of this mapping. It is quite straightforward to formalize this self-referential mental process of computation of the combined best mapping, through the concept of **self-referential** relativistic valuations in **self-referential** Kripkean models. Here are the definition of this class of models.
A self-referential Kripkean models is a structure where player i's rationalistic frame is a structure, $\mathcal{F}_G^i = \left\langle \left\{ w^{i_N}, w^{\dagger i_N} \right\}, \left\{ R^{i_N}, R^{\dagger i_N} \right\} \right\rangle$. Here $\mathbb{A}^{\dagger N}$ denotes the set of statements $\mathbb{A}^N$ when formulated at the "dual" perspective $w^{\dagger i_N}$. Accordingly, we have a (reflexive) dual binary relation $R^{\dagger i_N} \subseteq \left\{ w^{i_N}, w^{\dagger i_N} \right\} \times \left\{ w^{i_N}, w^{\dagger i_N} \right\}$.

Definition 7 *Let $\mathcal{F}_G^i = \left\langle \left\{ w^{i_N}, w^{\dagger i_N} \right\}, \left\{ R^{i_N}, R^{\dagger i_N} \right\} \right\rangle$ be a **self-referential** frame for player i in a game G. The **right** (resp. **left**) A^N-**self-referential valuation function** of player i in perspective N is the mapping,*

$$V_{w^{i_N}, w^{i \dagger N}}(\cdot, A^N) : \mathbb{A}^N \to \{0, \boldsymbol{i}, 1\}, \ \textit{(resp. } V_{w^{i \dagger N}, w^{i_N}}(A^N, \cdot) : \mathbb{A}^{\dagger N} \to \{0, \boldsymbol{i}, 1\}).$$

Note that the class of "self-referential valuations" is just a variant of the class of relativistic valuations, $\mathbb{V}_{\mathfrak{r}}$ defined in the original structure. Clearly, such referential models capture the idea that player i solves a game by computing "directly" the fixed point of the combined best response correspondence. Thus, intuitively, the notion of a self-referential valuation allows to capture the idea that a rationalistic player i needs to hold a belief about which profile $(\sigma^i)_{i \in N}$ could form a Nash equilibrium before he actually concludes that such a profile is indeed the fixed point of the combined best response correspondence $\mathtt{BR}$.

9. Concluding remarks

The textbook assumption underlying the analysis of the classical game model is that each player's belief is derived from his past experience, and that this experience is sufficiently extensive that he knows how his opponents will behave. While there is *a priori* no other alternatives, this assumption is at odds with the formal definition of a one-shot game wherein there is *no history* on which players can base their inferences of future play.
A careful analysis of the original game model that complies with this timeless structure brings about a situation which sharply departs from the current approach: The very existence of a choice—an observable choice in an experiment—together with the belief of a player has now to be "created" *ex nihilo*, in an "intrinsic" Nash equilibrium. This situation offers a quite different picture of what it takes to have an equilibrium in a game. The main upshot is that we should rather understand a decision in a game as a"gradual creative act" *ex nihilo*, rather than some deterministic choices accounted for by the players' prior experiences i.e. their pre-determined beliefs.
When it is interpreted as an interactive solution concept, the Aumann and Brandenburger (1995)'s result shows that a Nash equilibrium is the shorthand of a sort of communication between players as it requires some form of common knowledge of the players' conjectures—or the additional epistemic conditions found by Polack (1998). An important consequence of viewing a Nash equilibrium as an intrinsic principle is that the whole set of conditions on the interactive epistemology take now a purely self-referential meaning, which allows to

understand the epistemic conditions for Nash equilibrium as the mere result of some minimal consistency conditions. The upshot is thus that a Nash equilibrium does not call for a notion of "social epistemology" [the term is taken from Gintis, 2009], but of *self-interactive epistemology*.

Our equivalence result provides a clear-cut connection between game theory and logic, tying the Nash equilibrium with the basic notions of truth. This is therefore very much in line with the research program outlined by van Benthem (2007), "Game theory is mainly about global notions like strategic equilibrium in a game (...) this global representation needs fine-structure, of the sort than can be provided by ideas from both computational and philosophical logic".

Finally, the present ontological foundation provides a clear-cut criterion to decide whether some human action is a manifestation of human free will: If we believe in the rationality of humans, then any self-referential process of reasoning should indeed be interpreted as a manifestation of a human free will facing an open future.

10. Appendix

Proof of Theorem 1

Let $(A^J)_{J\in M}$ represent the Nash equilibrium profile $\sigma = (\sigma^{i_L})_{L\in M}$. The proof for self-referential model is a mere restatement of the proof for the general case and is therefore omitted.

Sufficiency. Consider a model with a M-rationalistic frame $\mathcal{F}^i_G = \langle w^{i_J}, R^{i_J}\rangle_{J\in M}$ with $|M| \geq 2$. Suppose that E^σ is common knowledge. Thus, we have the set of binary relations, $R^{i_J} \in \mathcal{R}^i$ such that $w^{i_J}R^{i_J}w^{i_K}$, for all $J \neq K$. We must also have a profile of statements, $(A^L)_{L\in M}$, which has been determined as being relatively true. That is, using Lemma 1, we have a system of relativistic valuation ,

$$A^J \in \arg\max_{A'^J\in\mathbb{A}^J} V_{w^{i_J},w^{i_K}}(A'^J, A^K; A^{-J,-K}), \forall(w^{i_J}, w^{i_K}) \in \mathcal{W}\times\mathcal{W}, J \neq K,$$

whose solution, $(A^L)_{L\in M}$, corresponds to the Nash equilibrium profile of G, $\sigma = (\sigma^{i_L})_{L\in M}$. The solution of the above system of equations is guaranteed by the standard existence theorem of a Nash equilibrium in finite games (Nash, 1950). Hence, if E^σ is common knowledge in the canonical model, then we necessarily have a Nash equilibrium with $\mathcal{M}^i_G,w^{i_K} \vDash A^{\sigma^J}, \forall w^{i_K} \in \mathcal{W}$. When this is true for each $i \in N$, this implies that σ is a NE of the game being played.

Necessity. We first show that A^J is true in a M-rationalistic model of i only if i determines a Nash equilibrium. By definition, the rationalistic Kripkean model with the single perspective w^{i_N} cannot yield the determination of a truth $A^N \in \mathbb{A}^N$ when there is no strongly dominant actions. [34] Thus, we must pick a M-Kripkean model with at least two perspectives. Suppose that there exists an arbitrary $L \in M$ with $R^{i_L} \in \mathcal{R}^i$ such that $w^{i_L}\neg R^{i_L}w^{i_K}$, for some $L \neq K$. Then, by definition, $\emptyset \in \arg\max_{A'^L\in\mathbb{A}^L} V_{w^{i_L},w^{i_K}}(A'^L, A^K; A^{-L,-K})$, for $L \neq K$, since $V_{w^{i_L},w^{i_K}}(A'^L, A^K; A^{-L,-K}) = \mathtt{i}$ whenever $\mathcal{M}^i_G, w^{i_L} \vDash \neg\Box^{i_L} A^{\sigma^K}$. Hence, we have that a statement is determined only if $w^{i_J}R^{i_J}w^{i_K}, \forall J \neq K$. The proof that a determined truth implies

[34] Indeed, recall that in this case $w^{i_N}(A^N) = \mathtt{i}$, $\forall A^N \in \mathbb{A}^N$ and there is no relativistic valuations.

common knowledge is then straightforward. E^{σ} is a self-evident event to every perspective w^{i_J} of player i i.e., $E^{\sigma} = K_{i_J} E^{\sigma}$. This follows since, $w^{i_J} \in \left\| \Box^{i_J} A^{\sigma^{-J}} \to_{\mathrm{L}} A^{\sigma^{-J}} \right\|^{\mathcal{M}^i_G}$. [35] Hence $w^{i_J} \in \left\| \Box^{i_J} A^{(\sigma^i)_{i\in N}} \right\|^{\mathcal{M}^i_G}$, for all $w^{i_J} \in \mathcal{W}$ whenever the profile of propositions, $(A^J)_{J=1}^m$ is the solution of the above set of equations. Thus, it suffices to apply Aumann's Theorem (see Aumann (1976)) to conclude that $E^{(\sigma^i)_{i\in N}}$ is indeed necessarily common knowledge. If σ is a NE of the game being played, this properties must therefore be true in $\mathcal{M}^i_G$ of every $i \in N$. **End of proof**.

References

[1] Aumann, R.J. Agreeing to disagree. Annals of Statistics 4 (1976), 1236-1239.

[2] Aumann, R.J. Correlated equilibrium as an expression of Bayesian rationality, Econometrica 55 (1987), 1-18.

[3] Aumann, R.J., Brandenburger, A. Epistemic conditions for Nash equilibrium, Econometrica 63 (1995), 1161-1180.

[4] Belnap, N., Perloff, M., Xu. M., Facing the Future: Agents and Choices in our Indeterminist World. Oxford University Press, 2001.

[5] Bernheim, D. Rationalizable strategic behaviour, Econometrica 52 (1984), 1007-1028.

[6] Bonanno, G., List, C., Tungodden B., Vallentyne P. (editors), Special issue on "Neuroeconomics", Economics and Philosophy, 24, (2008).

[7] Battigalli, P., Bonnanno, G., recent results on belief, knowledge and the epistemic foundation of game theory. Research in Economics, 53 (1999) 149-225.

[8] Bicchieri, C., The Grammar of Society: the Nature and Dynamics of Social Norms, Cambridge University Press, 2006

[9] Brandenburger. A. The power of paradox: some recent developments in in- teractive epistemology. International Journal of Game Theory, 35 (2007) 465-492.

[10] Brandenburger, A., Dekel, A. Hierarchies of beliefs and common knowledge, J. Econ. Theory 59 (1993) 189-198.

[11] Brandenburger A., Friedenberg, A. Intrinsic Correlation in Games, J. Econ Theory, 141, 2008, 28-67

[12] Brandenburger A., Epistemic Game Theory: An Overview, in The New Palgrave Dictionary of Economics, 2nd edition, Palgrave Macmillan, 2008.

[13] Dietrich, F., List, C., A reason-based theory of rational choice, Noûs (forthcoming).

[14] Gintis, H., The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences. Princeton University Press (2009).

[15] Gossner, 0., Tsakas, E., Reasoning-based introspection, Theory and Decision (forthcoming).

[16] Horty, J. F. (2001). Agency and deontic logic. Oxford University Press, Oxford.

[17] Kolhberg, E., Mertens J.-F.(1986): "On the Strategic Stability of Equilibria," Econometrica,54, 1003-1037.

[35] Recall that the binary relation R^{i_J} is reflexive at every perspective J. Thus, we have that $\forall E \subseteq \mathcal{W}$, $K_{\otimes_{i_J \in M}} E \subseteq E$. which means that the *truth axiom* is satisfied for each player i in $\mathcal{F}^i_G$.

[18] Kripke, S. Semantic analysis of modal logic, Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik 9, (1963) 67-96.

[19] Kripke, S."Outline of a theory of truth", Journal of Philosophy, 72 (1975) 690-716.

[20] Lewis. D. K. Convention: a philosophical study. Harvard University Press, 1969.

[21] Luce, R. D., and H. Raiffa (1957), Games and Decisions, New York: John Wiley.

[22] Lukasiewicz J. Philosophical remarks on many-valued systems of propositional logic (1930), Reprinted in Selected Works (Borkowski, ed.), Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, 1970, 153-179.

[23] Nash, J., Equilibrium points in n-person games, Proc. Natl. Acad. Sci. 36 48-49 (1950).

[24] Pearce. D. Rationalizable strategic behaviour and the problem of perfection. Econometrica, 52:1029-1050, (1984).

[25] Pelosse, Y., The intrinsic quantum nature of Nash equilibrium mixtures. Unpublished Manusript Mimeo (2011).

[26] Prior, A.N., Three-valued logic and future contingents, Philosophical Quaterly 3, 317-326, (1953).

[27] Prior, A. N., (1967) Past, Present and Future, Oxford: Clarendon Press.

[28] Perea, A., A one-person doxastic characterization of Nash strategies (2007), Synthese, Vol. 158, 251-271. (Knowledge, Rationality and Action 341-361).

[29] Polack, B., Epistemic conditions for Nash equilibrium, and common knowledge of rationality.Econometrica, 67, 673-676 (1998).

[30] Sabourian, H., Juang W-T., Evolutionary game theory: Why equilibrium & which equilibrium (2008) in Foundation of the Formal Sciences V: Infinite Games [Studies in Logic Series 11], Edited by Stefan Bold, Benedict Lowe, Thoralf Rash, Johan van Benthem, College Publications, 187-222.

[31] Samuleson, L., Evolution and game theory, Journal of Economic Perspectives 16 (2002) 47-66.

[32] Schelling, T C. The Strategy of Conflict. Harvard University Press. (1960) Cambridge, MA.

[33] Sugden, R., A theory of focal points, The Economic Journal 105, 533-550 (1995).

[34] Stalnaker, R., On the Evaluation of Solution Concepts, Theory and Decision, 37(1994), 49-73. Revised version in Epistemic Logic and the Theory of Games and Decisions, ed. by M. O. L. Bacharach, L.-A. Grard-Varet, P. Mongin and H. S. Shin. Kluwer Academic Publisher, 1997, 345-64.

[35] van Benthem, J., Interview. In V. Hendricks and P. Hansen, eds., Game Theory: 5 Questions. Automatic Press, Copenhagen, 2007, 9-19.

[36] Von Neumann, J., O. Morgenstern, Theory of games and economic behaviour, Princeton University Press, 1944.