

Game Theory, Game Situations and Rational Expectations: A Dennettian View

Cyril Hédoin*

University of Reims Champagne-Ardenne (France)

This version: 5 February 2016

Abstract: This article provides a theoretical and philosophical analysis of the account of rational expectations in games recently developed by Aumann and Dreze (Aumann and Dreze 2008) on the basis of the correlated equilibrium solution concept. Aumann and Dreze identify a player's rational expectation with his conditional payoff to a correlated equilibrium in a given "game situation". This definition depends on the satisfaction of several assumptions in an epistemic game-theoretic framework: the Information Partition Assumption, the Common Prior Assumption and the Common Knowledge of Bayesian Rationality Assumption. I evaluate these three assumptions on the basis of a particular view about intentionality, Daniel Dennett's intentional stance functionalism. Once rational expectations are interpreted along this Dennettian view, one is no longer committed with endowing the players with implausible cognitive abilities. They rather reflect and explain real patterns at the behavioral level. I argue that as an instantiation of externalism in the philosophy of mind, the Dennettian view provides a plausible defense of the Information Partition Assumption and also offers a new – though not entirely convincing – interpretation of the Common Prior Assumption. However, it fails to provide a satisfactory rationale for the Common Knowledge of Bayesian Rationality Assumption.

Keywords: Rational expectations – Epistemic game theory – Daniel Dennett – Correlated Equilibrium – Externalism

1. Introduction

While discussions about rational expectations are pervasive in macroeconomics, they are surprisingly scarce in the microeconomic context of game theory. However, since macroeconomic variables are obviously a function of the economic agents' choices at the microeconomic level, expectations about the former necessarily depend on expectations about the latter. As a consequence, the appraisal of the rational expectation hypothesis must proceed through a critical examination of the status of the concepts of expectations and beliefs in a strategic context where agents must solve coordination problems. This article proposes to

* Full Professor of Economics, economics and management research center REGARDS. Contact: cyril.hedoin@univ-reims.fr

Paper prepared for the 16th international conference of the Charles Gide Association for the Study of Economic Thought.

tackle the issue of the status of rational expectations in a game-theoretic framework on the basis of Aumann and Dreze's (2008) formalization of *game situations*: "a game played in a special context" and where "a player's expectation depends upon the context – the 'situation'". Aumann and Dreze identify a player's rational expectation with his conditional payoff to a correlated equilibrium in a given game situation. This definition of rational expectations basically relies on three key epistemic assumptions: the Information Partition Assumption (IPA), the Common Prior Assumption (CPA) and the Common Knowledge of Bayesian Rationality Assumption (CKBRA).

The paper investigates the ontological and methodological status of the players' expectations (and thus, of these three assumptions) and evaluates the relevance of the "rational" expectations hypothesis on the basis of an "externalist" view about intentionality that builds on Daniel Dennett's (1987) intentional-stance functionalism and its interpretation within economics by Don Ross (2005). Once rational expectations are interpreted along this Dennettian view, one is no longer committed to endowing the players with implausible cognitive abilities. They rather reflect and explain real patterns at the behavioral level. I argue that as an instantiation of externalism in the philosophy of mind, the Dennettian view provides a plausible defense of the IPA and also offers a new – though not entirely convincing – interpretation of the CPA. However, it fails to provide a satisfactory rationale for the CKBRA.

The article is organized as follows: the second section presents Aumann and Dreze's game theoretic account of rational expectations and provides an explicit epistemic framework for it. In the process, I provide a full description of the three key assumptions of IPA, CPA and CKBRA. The third section characterizes the Dennettian view of intentionality. It briefly surveys the general view called "externalism" in the philosophy of mind and focuses more specifically on Dennett's intentional stance functionalism, the latter being a peculiar instantiation of the former. The fourth, fifth and sixth sections respectively deal with the IPA, the CPA and the CKBRA on the basis of the Dennettian view. The seventh section concludes.

2. Aumann and Dreze's Game-Theoretic Account of Rational Expectations

Aumann and Dreze's article "Rational Expectations in Games" (Aumann and Dreze 2008) is one of the few attempts to explicitly characterize the rational expectation hypothesis in a game-theoretic framework. It builds on several of Aumann's key contributions in game theory and interactive epistemology, in particular on Aumann (1987) where the solution concept of correlated equilibrium is formally linked to the assumption of Bayesian rationality. Aumann and Dreze's contribution is significant for while the rational expectation hypothesis is pervasive in macroeconomics, its meaning from a microeconomic point of view has rarely been investigated. As a result, even if the hypothesis is formally well-defined in macroeconomic models, it is not clear what it precisely entails in terms of the agents' reasoning abilities and knowledge of others' reasoning abilities. Since the value of any relevant macroeconomic variable is necessarily a function of the agents' behaviors and the latter are partially due to the agents' expectations about others' behaviors, it follows that the definition of rational expectations with respect to relevant macroeconomic variables

necessarily depends on our ability to characterize *such expectations in terms of interactive epistemology*.¹

Cristina Bicchieri (1993) insightfully notes that the rational expectation hypothesis results from the conjunction of two logically independent assumptions regarding the epistemic rationality of the agents. The first (“strong subjective rational belief”) states that the agents use all the relevant information and do not make systematic (*i.e.* correlated) mistakes while according to the second (“objectively rational belief”) the belief of any agent is correct (*i.e.* corresponds to the objective probability distribution). However, according to Bicchieri (1993, 25), the rational expectation hypothesis “gives no account of how this coincidence comes about, as there is no plausible theory of how the agents “learn” to be epistemically rational in the sense specified by [the objectively rational belief assumption]”. As I explain below, even though Aumann and Dreze’s account does not offer any hint regarding how the agents may learn the objective probability distribution, it provides a clear-cut statement of the conjunction of the two epistemic assumptions underlined by Bicchieri but also shows that this conjunction is not sufficient in itself. I start by sketching Aumann and Dreze’s account that I then reformulate in an explicit epistemic framework.

Aumann and Dreze’s main purpose is to characterize an agent’s rational expectation in terms of the payoff that he can rationally expect to have in a particular *game situation*, *i.e.* “a game played in a specific context” where “a player’s expectation depend upon the context – the situation” (Aumann and Dreze 2008, 72). More precisely, they propose to define a player’s rational expectation as his *conditional payoff* to a correlated equilibrium in a given game situation. Consider a generic strategic interaction that we describe through some game $G: \langle N, \{S_i, u_i\}_{i \in N} \rangle$ where, as usual, N is a set of $n \geq 2$ players $i = (1, \dots, n)$, S_i is the finite set of player i ’s pure strategies and $u_i: S \rightarrow \mathbb{R}$ i ’s utility function mapping any strategy profile belonging to $S = \prod_i S_i$ onto some real number. A game situation corresponds to any particular instantiation of G where each player possesses some specific (private) information about anything that is relevant from his point of view, in particular the other players’ choices and beliefs. Let characterize such a game situation as an epistemic game $\Gamma_w: \langle G, \mathcal{I}, w \rangle$ where $\mathcal{I}$ is an information structure or a “broad theory of G ” that specifies what each player knows and believes about others and how he reasons on the basis of this information. A correlated equilibrium in G corresponds to a correlated distribution of strategy profiles defined by some variable $f(\cdot)$ such that each player maximizes his expected utility in each strategy profile given the information available to him. A player’s conditional payoff to a correlated equilibrium is defined as the player’s expected utility at an information set, *i.e.* what he expects to gain by implementing his strategy in the corresponding strategy profile given the information available to him, assuming that others play along the correlated equilibrium. Therefore, a player’s rational expectation corresponds to what he can expect to gain if he implements the strategy constitutive of a given correlated equilibrium defined by the function $f(\cdot)$, conditional on the information he has about his own behavior (possibly among other things).

¹ Informally, the rational expectation hypothesis is generally stated by ambiguous sentences like “the individuals know the relevant macroeconomic theory”, “the agents correctly predict the value of macroeconomic variables” or even more loosely “one cannot be fooled systematically”. Of course, the hypothesis has a quite clear formal expression, namely that for any agent i and for any macroeconomic variable X , i ’s expectation at time t of the value of X at time $t+1$ corresponds to the actual value of X at $t+1$ on average, *i.e.* $E(X^{t+1}) = X^{t+1} + \varepsilon_0$ where ε_0 is a random error variable of mean 0 and E the expectation operator.

As a simple illustration, consider the following hawk-dove game:

Figure 1

	H	D
H	0 ; 0	5 ; 1
D	1 ; 5	4 ; 4

This game has two Nash equilibria in pure-strategy, yielding (1, 5) and (5, 1) and one in mixed-strategy where each player plays H with probability $\frac{1}{2}$, thus yielding (5/2, 5/2). However, there are also a wealth of correlated equilibria, such as the one defined by the following probability distribution over the strategy profiles: $f([D, D]) = 1/3$, $f([H, D]) = 1/3$, $f([D, H]) = 1/3$, yielding (3, 3). This distribution constitutes a correlated equilibrium as it is not difficult to find that it is optimal for a Bayesian rational player to implement the strategy corresponding to each strategy profile given the conditional probabilities derived on the basis of $f(\cdot)$.² According to Aumann and Dreze's definition, a player's rational expectation when he plays H is then 5, while it is 5/2 when he plays D.³

There are several ways through which we can generalize this example and formalize the notion of game situation. This depends on how we represent the information structure $\mathcal{I}$ in the epistemic game Γ_w . Aumann and Dreze (2008) characterize the information structure in terms of a type space T . Each player i is endowed with a finite set of types T_i where each type t_i specifies a) the player's choice and b) the player's belief regarding the type of the other players. As each player's type defines a first-order belief over others' types, it is easy to see that we can associate to any type profile $(t_1, t_2, \dots, t_n)$ an infinite belief hierarchy that specifies all the higher-order beliefs of the players (*i.e.* what Row believes about what Column believes about what... he will play). A more natural but formally equivalent representation is in terms of a state space Ω . A state (or possible world) $w \in \Omega$ is an exhaustive description of everything that is relevant for the players and the modeler. In a game-theoretic context, it specifies in particular the players' choices, their beliefs about others' choices, their beliefs about others' beliefs and so on. Basically, at a given state, no uncertainty remains relatively to the value of any relevant variable.⁴ Formally, one can define a state w simply as a specific

² Consider Row player (the same reasoning applies for Column player). Obviously, it is optimal for him to play H in the profile (H, D) as in this case he knows with certainty that Column plays D. When Row plays D, he assigns a conditional probability of $\frac{1}{2}$ to Column playing H and of $\frac{1}{2}$ to column playing D. Then, by playing D as indicated, his expected gain is 5/2. If he plays H instead, his expected gain is also 5/2, thus he has no incentive to deviate. Therefore, the probability distribution defines a (weak) correlated equilibrium.

³ An interesting result discussed by Aumann and Dreze is that the players' rational expectations may be mutually inconsistent in the sense that the resulting payoff profile is infeasible, *i.e.* it is outside the convex hull of the possible payoff vectors. This shows that rational expectations in a strategic context do not entail efficiency or even consistency when players have differential information.

⁴ The state characterization (like the type characterization) corresponds to a semantic model in the sense that it consists to assigning a truth value to a list of propositions. A state is thus a list of propositions (about the players' behavior, about their beliefs, and so on) that are true. Such a semantic framework has a syntactic counterpart which is built from a language consisting in atomic propositions, logical connectives and modal operators. By combining this language with a set of axioms, we can derive a set of theorems. A syntax is sound and complete with respect to a given semantic model (or class of models) if all theorems in the syntax are valid in the semantic model (*i.e.* true at every states) and all valid propositions in the semantic model can be proved as theorems in the syntax respectively. In economics, it is usual to left the syntax implicit. See however Aumann (1999) and Bacharach (1994) for a discussion of the relationship between syntax and semantics in game theory.

type profile $(t_1, t_2, \dots, t_n)$, *i.e.* $\Omega = T_1 \times T_2 \times \dots \times T_n$, which means that we can also ascribe to each state a strategy profile and a belief hierarchy.

A second component of the information structure is a vector of prior probabilities functions $\{p_i(\cdot)\}_{i \in N}$ defined over the state space Ω . Formally, the function $p_i(\cdot)$ assigns to each state w a prior probability on the basis of which each player updates his belief conditioning on the information received. As I explain below, the proper economic interpretation for the prior function is unclear but it still plays a crucial role in the derivation of Aumann and Dreze's results. The third component of the information structure consists in a vector of accessibility relations $\{R_i\}_{i \in N}$ which states for each state w and for each player i which are the states w' that are accessible, denoted as wR_iw' (*i.e.* w' is accessible for i from w). The accessibility relation can be interpreted in several ways; in the epistemic context which is relevant here, the appropriate interpretation is in terms of epistemic possibility: wR_iw' if and only if when at state w , i considers possible to be at w' . I denote $R_i(\cdot)$ the corresponding possibility operator where $R_i(w)$ is the set of states w' that are accessible for i at w . The players' posterior probabilities $p_{i,w}(\cdot)$ are accordingly defined over $R_i(\cdot)$ using Bayes' rule. The tuple $\mathcal{J} = \langle \Omega, \{p_i, R_i\}_{i \in N} \rangle$ is thus the *broad theory* of game G , the complete description of whatever may happen or could have happened in G . Finally, denote $\mathbf{w} \in \Omega$ the actual state, *i.e.* the way the game is actually played and what the players actually know and believe. A game situation then is formalized as an epistemic game $\Gamma_{\mathbf{w}} = \langle G, \langle \Omega, \{p_i, R_i\}_{i \in N} \rangle, \mathbf{w} \rangle$.

As explained above, Aumann and Dreze characterize a player's rational expectation as his conditional payoff to a correlated equilibrium in a given game situation $\Gamma_{\mathbf{w}}$ based on game G . I denote $\mathbf{s}(w) = (s_1, \dots, s_n)$ the strategy profile that is implemented at w and $E[u_i(\mathbf{s}(w)) | R_i(w)]$ player i 's expected payoff when he plays his part in $\mathbf{s}(w)$ conditional on his information at w (defined by his possibility operator R_i and his posterior probability $p_{i,w}$). Now, the broad theory $\mathcal{J}$ implements a correlated equilibrium in G only if for all $w \in \Omega$, all players i and any strategy $s'_i \neq s_i$,

$$(1) \quad E[u_i(\mathbf{s}(w)) | R_i(w)] \geq E[u_i(s'_i, \mathbf{s}_{-i}) | R_i(w)], \text{ with } \mathbf{s}_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n).$$

Then, in any game situation $\Gamma_{\mathbf{w}}$, a player i 's rational expectation corresponds to $E[u_i(\mathbf{s}(w)) | R_i(w)]$, *i.e.* his expected payoff at the actual world $\mathbf{w}$. This definition of rational expectations depends on several assumptions that are related to the players' rationality, both practical and epistemic. Actually, these assumptions follow from a theorem that has been established by Aumann (1987) regarding the sufficient conditions for a correlated equilibrium to be played.⁵ For the rest of the paper, I will call it "Aumann's Theorem":

Aumann's Theorem – Consider a broad theory $\mathcal{J}$ of game G such that all players have a common prior $\mathbf{p} = p_1 = \dots = p_n$. Then, the probability distribution of strategy profiles $\mathbf{s}(w)$ is a correlated distribution $f(\cdot)$ such that $f(\mathbf{s}) = \mathbf{p}(w)$. Moreover, if all players are Bayesian rational at all states $w \in \Omega$, then $f(\cdot)$ corresponds to a correlated equilibrium in G .

Proof – See Aumann (1987) and Gintis (2009, 138-9).

⁵ See also Gintis (2009, 138-9).

Basically, since each state w specifies a unique strategy profile $s(w)$ to be implemented, it follows that the probability distribution of these profiles is defined by the common prior p . In other words, the common prior p in $\mathcal{J}$ implements the correlated equilibrium $f(\cdot)$ in G . Aumann's Theorem makes two explicit requirements: first, there must be a common prior over the state space; second, all the players must be Bayesian rational at all states of the world:

CPA (Common Prior Assumption): $\forall i \in N: p_i = p$.

CKBRA (Common Knowledge of Bayesian Rationality): $\forall i \in N, \forall w \in \Omega: E[u_i(s(w)) | R_i(w)] \geq E[u_i(s_i', s_{-i}) | R_i(w)]$.

These two assumptions are mathematically straightforward. CPA states that all the players share the same *ex ante* belief (*i.e.* before receiving any private or public information) over what can happen in G . Another way to state this assumption is that there is an (possibly tacit) agreement among the players regarding the fundamental features of the social world. CKBRA indicates that the players are Bayesian rational, that everyone knows that, that everyone knows that everyone knows that, and so on *ad infinitum*.⁶ Aumann's Theorem also relies on a third, implicit assumption according to which each player must have an *information partition* I_i over Ω . This is derived from the properties of the possibility operators R_i :

IPA (Information Partition Assumption): The possibility operators R_i have the following properties,⁷

- a) $\forall i \in N, \forall w \in \Omega: w \in R_i(w)$.
- b) $\forall i \in N, \forall w, w' \in \Omega: \text{if } R_i(w) \neq R_i(w'), \text{ then } R_i(w) \cap R_i(w') = \emptyset$.

Property (a) states that a player always considers the actual state of the world to be possible. Property (b) is a direct statement that each player has a partition over Ω : if at w , i considers w'' as possible but w' as impossible, then at w' he cannot also consider w'' possible. Therefore, Ω is divided into cells with no intersection.⁸

The foundations of rational expectations in games are thus now explicit: the players must have an information partition and a common prior, and Bayesian rationality must be common knowledge. Though they are somewhat standard in economics (especially in information economics), these assumptions are also all controversial. I will present and discuss a rationale

⁶ A player is Bayesian rational if he maximizes his expected utility given his subjective beliefs and if he uses Bayes' law to update his beliefs conditional on some information.

⁷ IPA could also be stated in terms of the properties of the accessibility relation R_i from which the possibility operator is derived. In this case, R_i must be reflexive (wR_iw), transitive (if wR_iw' and $w'R_iw''$, then wR_iw'') and Euclidean (if wR_iw' and wR_iw'' , then $w'R_iw''$).

⁸ Formally, property (b) is sufficient for the players to have a partition I_i over Ω . However, Aumann's Theorem as well as Aumann and Dreze's account are couched in terms of both belief and knowledge, *i.e.* probability 1 beliefs that are true. Property (a) is then required to make sure that what the player believes with probability 1 is indeed true. We could dispense with this requirement if instead we choose to frame the whole discussion in terms of beliefs uniquely. I will not go into these subtleties as they do not affect the main points of my argument.

for them based on a Dennettian view of intentionality in sections 4, 5 and 6. Before, I characterize this Dennettian view and explain why it is relevant here.

3. A Dennettian View of Intentionality: Externalism and Intentional-Stance Functionalism

It is important to emphasize that Aumann and Dreze's do argue neither for the *empirical* plausibility nor for the *theoretical* relevance of rational expectations. Their goal is purely formal: to provide a mathematical characterization of rational expectations in explicitly strategic contexts. I shall argue however that this mathematical characterization could potentially enhance both the empirical plausibility and the theoretical relevance of the rational expectation hypothesis, provided we adopt a specific view about intentionality that I call the "Dennettian view".

The rational expectation hypothesis has been disputed both at the theoretical and the empirical levels. I have already mentioned one objection made by Bicchieri (1993): the rational expectation hypothesis does not provide any argument for the assumption that the agents' subjective beliefs always match the objective probability distribution of events in the world. There are clearly grounds to doubt this assumption even for exogenous events that do not depend on the agents' choices and beliefs. Doubts are only strengthened when we consider events whose probability distribution is endogenously determined. Moreover, the rational expectation hypothesis builds on the strong assumption that errors tend to cancel out, *i.e.* agents may make mistakes (choices that do not maximize expected utility) but these mistakes are randomly distributed such that they do not play any role at the aggregate level. This is dubious from a purely empirical point of view, at least until we show that there are institutional structures that have the property to generate some form of "ecological rationality" (Smith 2009). One may argue that Aumann and Dreze's characterization of rational expectations in games suffers from the same problem. Consider this line of argument: according to Aumann and Dreze, from her point of view in some game situation Γ_w , Ann rationally expects to gain $E[u_{Ann}(s(w)) | R_{Ann}(w)]$. Such expectation however depends on the fact that a specific correlated equilibrium corresponding to a common prior p is implemented. But there are typically many correlated equilibria in a game.⁹ Why should Ann have any particular reason to expect that this particular equilibrium will be played and not any other possible one? Why should she even expect that *any* correlated equilibrium will be played? We could provide many specific answers to these questions: maybe Ann has previously agreed with Bob that the strategy profile to be implemented should be a function of the result of a coin toss or of the weather. Maybe Ann has observed that in the past, people's behaviors were correlated to some external signal (*e.g.* in most places on the planet, people seem to stop at red traffic lights but not at green ones). Or maybe this is due to a purely genetic disposition that has programmed Ann to identify some asymmetries in an interaction and to adopt a particular behavior on this basis (Skyrms 1996, chap. 4). These are all *proximate*

⁹ Consider the game described by Fig. 1 above. It is easy to see that there are an infinity of correlated equilibria in this game because any probability distribution of the two Nash equilibria [H, D] and [D, H] is a correlated equilibrium. More generally, since any Nash equilibrium is a correlated equilibrium while the reverse is not true, the number of correlated equilibria in a game is always even or greater than the number of Nash equilibria.

explanations. But the only ultimate and general one has already been provided: Ann's rational expectation is grounded on the fact that the IPA, CPA and CKBRA are all satisfied.

If one wants to make sense of rational expectations in games both from a theoretical and empirical point of view, then the task is to provide a rationale to these three assumptions. There are many arguments to reject all of them that I will examine in due course. If these arguments are valid, then we should not expect people to empirically have rational expectations but we should also be skeptical regarding the methodological and the theoretical value of the rational expectation hypothesis. I shall argue that the Dennettian view of intentionality provides maybe the best defense of the three assumptions even if ultimately there are reasons to consider that it is not entirely successful. Indeed, expectations and beliefs are what philosophers of mind call *intentional states*.¹⁰ The ontological status and the methodological implications of rational expectations thus depend on one's views about intentionality and intentional states. As argued by Ross [(2005); (2014)] in the case of revealed preference theory, Daniel Dennett's (1987) intentional-stance functionalism provides an interesting account of the nature of intentional states paving the way for a reinterpretation of economic theory. There are reasons to think that the same may be true in the case of rational expectations as characterized above.

Before describing Dennett's account further, it might be helpful to consider a key distinction in the philosophy of mind between *internalism* and *externalism*, Dennett's intentional-stance functionalism being a peculiar instance of the latter. The distinction concerns the nature and the origin of the semantic content of intentional states. Any intentional state (also called intentional attitude) is of the general form $F(x)$ where F is the type or mode of state (belief, desire, intention ...) and x the propositional content, *i.e.* what the state is "about". For instance, the fact that I believe that the Golden State Warriors won the NBA Finals in 2015 is an intentional state $F(x)$ where F corresponds to the type "belief" (a cognitive attitude with a mind-to-world direction of fit) and x is the object of the belief, *i.e.* that "the Golden State Warriors won the NBA Finals in 2015". A key feature of such intentional attitudes is their "aboutness": they are about something that is distinct and in some way external to the person who holds the attitude. In other words, intentional attitudes represent non-mental properties or states of affairs. The propositional content of an intentional state also represents the "conditions of satisfaction" of this state: the fit between the intentional state and the actual state of affairs is achieved if and only if the latter matches with the propositional content. In the case of a belief, a belief F that x is true if and only if x actually holds as a state of affairs. The debate between internalism and externalism about the propositional content of intentional states concerns the way the meaning of the propositional content x in $F(x)$ is determined. Internalism holds that the meaning of x is intrinsic to the entity (the person) that holds the state $F(x)$. Alternatively, we might say that the meaning of x supervenes on the intrinsic properties of the entity. Externalism holds precisely the converse: the meaning and even the very existence of the propositional content is partially determined by the relationship of the entity with its environment. In other words, the semantic content of intentional states depends on the whole situation in which these states are embedded.

¹⁰ The same is true for preferences, assuming that they represent desires and wants.

Though the debate is far from having been completely settled, externalism has acquired a dominant position in the philosophy of mind and in the cognitive sciences.¹¹ For some kinds of intentional states, the case for externalism is almost straightforward (Lau and Deutsch 2014). This is the case for instance regarding knowledge about the external world: I can know that the Warriors won the 2015 NBA Finals only if indeed it is the case that Warriors won the 2015 NBA Finals. The disagreement between internalists and externalists rather holds for intentional states that do not imply a veridical content. Since this paper is concerned with the status of (rational) expectations, it is especially important to state what externalism for beliefs implies. Consider this modified version of Hilary Putnam's (1975) "Twin Earth example": suppose that somewhere in the Universe (or in one among many universes if one holds that there are parallel worlds) there is a planet almost identical to ours that we call "Twin Earth". Suppose that the only difference between Earth and Twin Earth is that on the latter people do not play basketball exactly in the same way as we play it on Earth (we may assume for instance that tackles are legal and that you can walk with the ball without making it rebounding). Suppose finally that the NBA Finals take place each year on Twin Earth to determine which team will be the champion in the same way as on Earth. Now, someone committed to externalism about beliefs will hold that the content of any belief related to the game of basketball (such as which team will win the NBA Finals this year) has no intrinsic meaning but instead depends on whether the entity having this belief is on Earth or on Twin Earth. For instance, Bob and Twin-Bob (Bob's counterpart on Twin Earth) may hold exactly identical beliefs about states of affairs related to basketball, *e.g.* both may believe that the Warriors will win the 2016 NBA Finals. However, according to externalism, Bob's and Twin-Bob's beliefs are not the same because they refer to qualitatively different things. Meaning is not "in the head" but instead is determined by the whole relationship between the believer and his environment.

I do not intend here to evaluate the virtues of externalism with respect to beliefs against various forms of internalism. As I said above, externalism has currently the upper-hand among philosophers of mind and cognitive scientists (at least those concerned with such philosophical issues). This fact is by itself sufficient to justify that we investigate the implications of externalism for economics. The point is thus to reinterpret the meaning of rational expectations in games from an externalist point of view and to determine whether this reinterpretation reinforces the rational expectation hypothesis, from a methodological, a theoretical and possibly an empirical perspectives. At the same time, it is clear that externalism is not a strongly committing doctrine: we can all be externalists regarding the content of intentional states and yet defend a great diversity of positions about a range of issues such as the mind-body problem and the nature of consciousness, mental causation or the problem of intentionality. Dennett's intentional-stance functionalism is a particular instantiation of externalism whose significance for economics has already been argued by Ross (2005). As Ross convincingly shows (and as Dennett has punctually remarked), there is

¹¹ Burge (1986) and Putnam (1975) provide two well-known and early defenses of externalism about the content of intentional states, building on a related but different form of externalism, semantic externalism. Crane (2015) and Searle (2004) present introductory remarks about the externalism/internalism debate, both endorsing internalism. Dennett explicitly endorses externalism at several places, see especially his article "True Believers: The Intentional Strategy and Why it Works", reprinted in Dennett (1987, 13-35).

a deep affinity between Dennett's views about what he calls "intentional systems" and the way agents are formalized in economics, especially in decision theory and game theory.¹²

Dennett's account is not easy to recapitulate because it has several ramifications and subtleties. I will confine myself to the way Dennett interprets intentional states and particularly beliefs. As its name indicates, intentional-stance functionalism is a specific form of functionalism. The latter is itself a broad view according to which mental events (including intentional states) supervene on physical events and that the relationship between the two kinds of events is a functional one: a given physical event (or state) E is a mental event (a state) in virtue of its function in the overall behavior of the entity. An interesting implication of functionalism is the relative independence between the characteristics of the physical structure of the entity (the hardware) and its functional activity (the software): how a particular functional activity is implemented is secondary and in principle any functional activity can be implemented by any physical structure.¹³ Functionalism takes several shapes and has for instance led some scholars to lean toward "eliminativism", *i.e.* the doctrine holding that the categories of folk psychology (beliefs, desires) correspond to an erroneous theory of mind and that scientific discussions should refrain from using them. Though it has been frequently associated to such an eliminativist endeavor, Dennett's intentional stance functionalism firmly rejects eliminativism.¹⁴ Quite the contrary, Dennett's account can be seen as an attempt to rehabilitate the categories of folk psychology not only from an epistemological point of view (*i.e.* as useful concepts to predict people's behavior) but also from an ontological point of view.

The nature of intentional states such as beliefs is deeply related to the existence of what Dennett calls "intentional systems": "What is it to be a true believer is to be an intentional system, a system whose behavior is reliably and voluminously predictable via the intentional strategy" (Dennett 1987, 15). The intentional strategy consists precisely in predicting one's behavior through the attribution of intentional states. Thus, reformulating

- a) An entity E has belief B(E) only if E is an intentional system S.
- b) E is an intentional system S only if its behavior is reliably predictable via the intentional strategy I.
- c) The intentional strategy I consists in predicting E's behavior by attributing to E some belief B(E) (possibly among other intentional states).
- d) Therefore: An entity E has belief B(E) only if E's behavior can be predicted by attributing to E belief B(E) (possibly among other intentional states).

The intentional strategy for predicting someone's behavior is what Dennett also calls the *intentional stance*. It corresponds to one of the three available epistemological postures (along with the physical stance and the design stance) to predict and explain a system behavior.

¹² Dennett has not written anything specific about economics and its methodology. However, he notes that his "intentional system theory" overlaps with branches of economics and especially game theory at several places. See for instance his article "Three Kinds of Intentional Psychology" reprinted in Dennett (1987, 58).

¹³ This point of view is held in particular by what is sometimes called "computer functionalism" or "strong artificial intelligence" (Searle 2004). The Church-Turing thesis and the related notion of Turing machines are at the core of computer functionalism as the former holds that any computable function is computable by a universal Turing machine (the Turing machine that computes all the functions computed by all Turing machines). If one assumes that the mind is a computable function (a premise which is of course debatable), then the Church-Turing thesis strongly indicates that strong artificial intelligence is possible.

¹⁴ For an explicit departure from eliminativism, see for instance Dennett (1987, 227-235).

Taking the intentional stance toward someone (or something) consists in explaining and predicting the latter's behavior through the attribution of intentional states. The intentional stance has an apparently strong instrumentalist flavor that corresponds to what can be called the "Dennettian method" for behavioral explanation (Ross 2002, 154) or "methodological intentional-stance functionalism" (Ross 2005).

However, a purely instrumentalist reading of Dennett's account would indicate that belief attribution is merely based on a falsifiable theory of the mind and that we should give it up provided we are able to show that this theory is false or unnecessary (with respect to some parsimony criterion). This is precisely the eliminativists' position that Dennett rejects. Consider indeed hypothetical Martians who are observing Humans and are trying to predict our future on the basis of superhuman abilities making them equivalent to Laplacean superphysicists: "Our imagined Martians might be able to predict the future of the human race by Laplacean methods, but if they did not also see us as intentional systems, they would be missing something perfectly objective: the *patterns* in human behavior that are describable from the intentional stance, and only from that stance, and that support particular generalizations and predictions" (Dennett 1987, 25, emphasis in original). Dennett's point is that the intentional stance is not merely instrumental; it is the *only* way to observe *real* behavioral patterns. In this sense, the intentional stance is not a theory that might be proved to be wrong as eliminativists would hold. It is constitutive of real patterns that we cannot characterize but in terms of beliefs, desires and other kinds of mental states. According to Dennett's "ontological intentional-stance functionalism" (Ross 2005), there is thus nothing more in the fact that the entity E has the belief that ϕ than the fact that E's behavior can be interpreted and predicted (by E itself or others) from the intentional stance through the ascription to E of the belief that ϕ . This is a form of realism, though a mild one since in many cases one's mental states will be partially indeterminate from the intentional stance (Dennett 1991).¹⁵

We are now in the position to characterize the Dennettian view of intentionality, especially of beliefs. On this view the semantic content of beliefs and any other intentional states is not intrinsic to the entity holding them (Externalism). It depends on the functional relationship between the entity and its environment (Functionalism). This semantic content is fixed by adopting the intentional stance: this entity has the beliefs and the other intentional states that make its behavior the most understandable and predictable, assuming that the entity is endowed with at least a minimal form of rationality (*i.e.* we attribute to the entity the beliefs and desires it *ought to* have given its behavior and the environmental context) (Intentional Strategy). Moreover, this is all there is to have a belief though in many cases the precise content of the belief will be indeterminate (Mild Realism). For the rest of the paper, the Dennettian view will denote the conjunction of Externalism, Functionalism, Intentional Strategy and Mild Realism.

¹⁵ This indetermination has a strong formal similarity with Quine's radical translation problem, as Dennett notes at several places.

4. Externalism and the Information Partition Assumption

I now examine the IPA, CPA and CKBRA on the basis of the Dennettian view presented above, starting with the IPA. Recall that the IPA states that each player i has an information partition I_i over the state space Ω . To have an information partition means that each state w belongs to one and only one set defined by the possibility operator $R_i(\cdot)$. As a consequence, ignoring the limit case where $R_i(w) = \Omega$ for all $w \in \Omega$, at any w the player i has some information such that the state space can be divided into the states w' that he knows are impossible (*i.e.* $w' \notin R_i(w)$) on the one hand and the set of states w' he knows are possible (*i.e.* $w' \in R_i(w)$) on the other hand. Moreover, we have assumed that $w \in R_i(w)$ which means that i necessarily knows at w that w is possible. It follows that the partition I_i defines i 's *knowledge* in any given game situation. To see this point, consider the following definitions.

Definition 1 – An event $E \subseteq \Omega$ is a subset of states such that a given proposition is true at all $w \in E$.

Definition 2 – The event that player i knows an event E is denoted $K_i E$ and is defined as follows: $K_i E = \{w \mid R_i(w) \subseteq E\}$.

According to Definition 2, i knows E at w if and only if all the states i considers as possible belong to E . Note that $K_i E$ is itself an event since it corresponds to set of states where the corresponding proposition is true. Accordingly, K_i is called a “knowledge operator”. On the basis of these definitions and of the IPA, it can be shown that the knowledge operator satisfies the following axioms for any events E and F :¹⁶

- (K1) $K_i \Omega = \Omega$
- (K2) $E \subseteq F \rightarrow K_i E \subseteq K_i F$
- (K3) $K_i E \subseteq E$
- (K4) $K_i E = K_i K_i E$
- (K5) $\neg K_i E \subseteq K_i \neg K_i E$

The first two axioms are axioms of (logical) omniscience as K1 states that one knows everything that is necessarily true and K2 that one necessarily knows the logical consequences of what he knows. K3 is generally known as the truth axiom and is constitutive of the definition of knowledge: one can only know things that are true. K4 and K5 are sometimes called axioms of transparency (or of positive introspection) and of wisdom (or of negative introspection) respectively. The former states that one knows that he knows and the latter that when one does not know something he knows this (or equivalently, that if one does not know that he does not know something, then he knows this something). It is now largely established that these five axioms are the semantic equivalent to the well-known S5 modal logic at the syntactic level (*e.g.* Stalnaker (2006)). From this point of view, even though K1 and K2 lead to the problem of logical omniscience, they are generally considered as quite standard by logicians.¹⁷ Axioms K3, K4 and K5 are more disputable, particularly in the context of

¹⁶ See for instance Binmore (2007) or Gintis (2009) for simple derivations.

¹⁷ K1 and K2 alone correspond to the modal logic K which is taken as the generic system of modal logic. Logical omniscience may be regarded as problematic as it endows agents with excessively strong cognitive and

economics. K3 implies that what one believes with certainty is necessarily true, thus forbidding the possibility that beliefs with probability one are false. At the same time, K3 is required as soon as one wants to deal with problems of information in terms of knowledge as it is generally the case in economics. K4 and especially K5 cannot be defended on this basis: they arguably define very strong requirements of epistemic rationality that seem hard to ground on a reasonable psychological and/or economic rationale. Still, like K3, they directly follow in Aumann and Dreze's account from the fact that it is implicitly assumed that each player knows his "type" t_i . Since a state w corresponds to a profile of types $(t_1, \dots, t_n)$, it follows easily from this that if any player i knows the event that he is of type t_i then he is indeed of type t_i (K3). Moreover, since to be of type t_i implies to know his type, knowing his type implies to know this (K4). Finally, since i knows that he is of type t_i , he knows when his type does not know something (K5). IPA is thus implied by the type-space approach used by Aumann and Dreze.

I shall argue that the Dennettian view provides a convincing case for the IPA in a game-theoretic context. This claim can be supported by several arguments. First, a case can be made for the fact that the axioms of S5 system in epistemic logic and thus the IPA are innocuous in a small set of specific contexts (Lismont and Mongin 1994). For instance, it has been argued that the use of the so-called "Kripke structures" is justified for the study of systems with distributed knowledge, *i.e.* systems where distributed information is computed by parallel processors (Halpern and Moses 1990). Relatedly and more significantly, the use of information partitions seems to follow naturally from an "external view" of knowledge, *i.e.* "knowledge as ascribed by the scientist rather than computed by the agent" (Lismont and Mongin 1994, 91). This external view of knowledge is clearly deeply related to externalism about intentional states. As I have noted above, externalism about knowledge is almost uncontroversial because knowledge implies a veridical content: one can only know something that is true and truthfulness implies a relationship between a mental state and some "external" state of affairs. In this sense, the ascription of knowledge to an agent (possibly by the agent himself) implies a reference to the environment. On this view, there is nothing like an "inner state of knowledge" and all the axioms of the S5 modal logic (including the axiom of wisdom¹⁸) seem to be defensible on this basis. However, this argument clearly depends on a definition of knowledge as true belief that may be rejected for independent reasons.

A second argument still relies on an externalist view of knowledge but also more specifically on Dennett's intentional-stance functionalism. The uneasiness with the axioms of the S5 modal logic in an epistemic context stems from the fact that the computation of all the required information is cognitively out of reach for normally rational humans. In particular, the omniscience axioms (K1 and K2) appear to be particularly strong as they endow the agents with perfect logical abilities. The axioms of transparency and wisdom (K4 and K5) assume that the agents know what they *would* know and not know in any (counterfactual)

computational abilities. See Bacharach (1994), Halpern and Pucella (2011) and Sillari (2008), among others, for discussions of this problem in economics, computational science and philosophy respectively.

¹⁸ If the proposition that "Ann does not know that ϕ " is true, then Ann can *potentially* have the knowledge that she does not know that ϕ . Now, from an external point of view, the ascription of such a knowledge is permissible as long as nothing in Ann's relationship with the external world indicates (through her behavior for instance) that she is ignorant of her ignorance. The latter is a possibility of course, but it is in principle always possible to change our description of this relationship (formally, by redefining the state space and the corresponding information partitions) such that Ann "knows" her ignorance. This point is related to Dennett's remark that intentional states may be indeterminate and Quine's radical translation problem.

state of the world. In other words, they depend on the fact that the agents are able to compute all the information, including the information that is not actually available. Such a computational load seems to be intractable given the cognitive capacities of any normally intelligent human. There are two complementary answers to this objection building on externalism and intentional-stance functionalism. On the one hand, on an externalist reading of knowledge and other intentional states, we need not assume that the agents “really” make all the computations that are reflected in a given epistemic model. That is, we do not require that the agents are able to make explicit their reasoning and the knowledge on which it is based, or that they have some inner, privileged and/or conscious access to the underlying computational processes. The distinction between *explicit* and *implicit* knowledge widely used in the literature on the so-called “awareness structures” captures this point.¹⁹ Under this terminology, an agent explicitly knows that ϕ if and only if he implicitly knows that ϕ and he is aware of ϕ , where the latter could here mean “being conscious of ϕ ”. The K1-K5 axioms are clearly problematic in terms of explicit knowledge for the reasons just stated. However, there is *a priori* no reason to reject these axioms once it is acknowledged that they are about implicit knowledge. From the standpoint of (computer) functionalism, the use of knowledge in the sense of *implicit* knowledge is quite natural as the computation associated to the intentional state ascribed to the agent is not an “inner” one.

On the other hand, Dennett’s intentional stance functionalism provides a rationale to start from the assumption of “perfect rationality”: “That is, one starts with the assumption that people believe all the implications of their beliefs and believe no contradictory pairs of beliefs” (Dennett 1987, 21). The first part of this assumption corresponds to axiom K2, while the second corresponds to an axiom that is weaker than the truth axiom K3.²⁰ Moreover, on the intentional strategy, the attribution to an intentional system of possibly false beliefs necessarily “requires a special genealogy, which will be seen to consist in the main in true beliefs... An implication of the intentional strategy, then, is that true believers mainly believe truths” (Dennett 1987, 18-9). Thus, Dennett’s intentional strategy clearly recommends to ascribe beliefs (either true or false) through the ascription of what we can call a “knowledge base”, *i.e.* a set of basic beliefs that are true. This clearly supports the truth axiom K3. Note that these basic beliefs are supported by the agents’ possibility operators R_i as one’s knowledge is formally represented by the set $R_i(w)$ for each state $w \in \Omega$. This requires that these sets do not intersect which is guaranteed by axiom K4 (Bacharach 1993). Indeed, the contrary would imply that one can know contradictions, which is impossible.²¹

It is less clear that the intentional strategy provides an independent support for the axiom of wisdom K5. Without it, the resulting system of axioms (known as S4) allows for the possibility operators R_i to define a topology instead of a partition and Aumann’s Theorem no longer applies. However, we have already seen that K5 is not unreasonable on an externalist

¹⁹ For a survey of awareness models in computational science, philosophy and economics, see Sillari (2008).

²⁰ Formally, $K_i E \subseteq \neg K_i \neg E$. This axiom (generally denoted axiom D in the modal logic literature) applies to an epistemic logic for beliefs instead of knowledge. When D is substituted for K3, the resulting system of modal logic is called KD45.

²¹ Suppose that an agent i has the following information pattern over the space $\Omega = (1, 2, 3)$: $R_i(1) = (1, 2)$ and $R_i(2) = R_i(3) = (2, 3)$. Therefore, the two sets intersect at world 2. However, this pattern is impossible if axiom K4 is satisfied: assume that the actual world is 1; then, i knows that 2 is possible but that 3 is impossible, but at 2 he knows that 3 is possible. Since K4 indicates that i knows what he knows and since one cannot know contradictions (by the axioms of propositional logic and K3), he cannot know that he knows that 3 is both possible and impossible.

understanding of knowledge. Moreover, as shown by Bacharach (1993), it is possible to derive information partitions I_i in an epistemic model from the combination of axioms K1, K2 and K3 only. This derivation is based on the notion of “experiment”: an experiment $e = \{w, \eta, y\}$ consists for any player i at a state w in an observation y generated from a function $\eta: \Omega \rightarrow Y$ that is known by i and mapping the states of the world onto an observation y , with Y the set of observations. On this basis, Bacharach shows that in an epistemic model satisfying axioms K1-K3 and where player i ’s knowledge is defined by the possibility operator R_i , the event that i is an observer in an experiment e implies that i has an information partition over Ω . Intuitively, the idea is that when one observes an experiment e and nothing else, then his information will reflect the known information function $\eta: \Omega \rightarrow Y$. From the perspective of the intentional strategy, one’s knowledge and the accompanying information partition may then be seen to correspond to this underlying information function. Incidentally, the interpretation of this information function leads to similar issues than the interpretation of the CPA, to which I turn now.

5. The Common Prior Assumption and The Intentional Stance

The CPA states that all the players agree on the probability distribution of the states and thus of every relevant characteristic of a given strategic interaction. A player’s prior reflects his “fundamental beliefs” which, in some way, can be seen as being part of the “knowledge base” discussed above. This is on the basis of this prior and of his information partition that an agent is able to form his actual belief $p_{i,w}$ in any particular game situation and thus to derive his rational expectation. The interpretation of the meaning of a player’s prior in an epistemic game is far from being obvious and the assumption that the players have a *common* prior is therefore even more difficult to evaluate. However, the Dennettian view provides a rationale for the CPA, though it ultimately depends on a symmetry hypothesis between the modeler and players that is disputable.

Formally, a player’s prior is a probability measure over a state space. An intuitive interpretation is that a player’s prior represents his *ex ante* belief about a list of propositions *before receiving any private or public information about the situation*. This interpretation is somewhat problematic from an empirical point of view as it is not clear what it is to be devoid of any information. Moreover, both the origin and the nature of this *ex ante* belief are left undefined: is this a purely subjective belief *à la* Savage (Savage 1954) or rather an objective belief that reflects some fundamental features of the strategic interaction? Was this belief “already there” before the strategic encounter or did it result from, say, a prior communication stage where people have explicitly exchanged over a set of propositions?²² Actually, given the fact that the epistemic program in game theory has initially largely been conceived as an

²² Faruk Gul (1998) argues that the case of state space epistemic models should be distinguished from the case of type space epistemic models. According to Gul, the notion of prior is meaningless in the latter case as there is no prior stage that can be associated with a “sensible thought experiment”: “the hierarchy representation interpretation offers no argument to identifying the “priors” of the representation with beliefs at any prior stage” (Gul 1998, 926). In the state space case, a prior may be interpreted as depending on a prior stage which “represents a situation which actually occurred at some previous time” (Gul 1998, 924). However, under this interpretation it becomes problematic to assume that the player’s prior ranges over his own action and that the priors are commonly known. I return on this point in the text. See also Aumann’s (1998) response to Gul’s critique.

attempt to reintroduce decision theory in the context of strategic interactions [(Aumann 1987); (Brandenburger 2014)], the prior-as-subjective-belief interpretation is a plausible one. From this point of view, the “origin” of the belief is outside the scope of the theory of games and maybe of economics as a whole. The problem of this interpretation is that it makes the CPA very hard to sustain as there is *a priori* no reason to expect that people will or should have the same subjective beliefs over everything that is relevant in a strategic interaction.

At the most general level, the CPA states that differences in (posterior) probabilities necessarily express differences in information. In other words, when two agents entertain different beliefs, this must be due to the fact that they do not have the same information. Various arguments have been developed in the economic literature for this assumption. Morris (1995) identifies four kinds of justifications: logical/rational, frequentist, empirical and pragmatic. The frequentist justification is unavailable under a subjectivist interpretation of priors and the empirical justifications are largely unconvincing. The logical/rational justification refers to the so-called “Harsanyi doctrine” (or even the “Harsanyi-Aumann doctrine”). It follows from Harsanyi’s famous demonstration that a game with incomplete information can be transformed into a game of imperfect information with an initial move made by “nature” if and only if the players have a common prior over some state space (Morris 1995, 230). This is on this basis that Aumann (1976) has produced his not less famous “agreement theorem” according to which individuals with a common prior and who have common knowledge of their posterior beliefs must have the same posterior beliefs. Aumann’s agreement theorem makes clear that people with a common prior and the same information cannot “agree to disagree”.²³ The Harsanyi-Aumann doctrine more or less claims that the CPA follows as a property of rationality: if two persons have the same information but different beliefs, then this must be due to someone having made a mistake. The pragmatic justifications are by far the most common, at least in information economics and social choice theory. For instance, it has been noted that normative analysis in terms of social welfare functions becomes difficult if not impossible without assuming a common prior [e.g. (Broome 1989); (Mongin 1995)]. More basically, it is sometimes suggested that giving up the CPA would lead to an “anything goes” methodology. Finally, it could be argued that heterogeneous priors are best captured as parameters of the utility functions and modeled as information processing errors.

Neither the logical/rational nor the pragmatic justifications are convincing, at least as they are generally stated. The pragmatic justifications can all be disputed: for instance, the CPA is required for the normative in terms of social welfare functions only if one wants to preserve the criterion of *ex ante* Pareto efficiency. Moreover, there is no more reason to suppose that giving up the CPA will lead to *ad hoc* explanations than the standard practice consisting in specifying utility functions with very specific properties (e.g. homothetic utility functions or utility functions with constant relative risk aversion). The logical/rational justification seems plainly untenable. On the one hand, Bayesian decision theory is completely agnostic regarding the choice of a prior. It is true that Savage’s expected utility theory imposes consistency axioms regarding the choices over uncertain prospects (in particular the independence or “sure-thing” principle). But these axioms fall short of specifying a procedure

²³ See also Geanakoplos (1992) for an accessible discussion of the agreement theorem and its implication. It is in particular shown that a common prior combined with common knowledge of actions negate asymmetric information.

to choose a prior. On the other hand, while the logical/rational justification may seem reasonable or plausible for natural and exogenous events (in which case it would be similar to the frequentist justification), it is hard to accept in the case of endogenous events as they occur in strategic interactions [(Gintis 2009); (Morris 1995)]. Rationality or logic cannot dictate by itself what is the correct prior as this depends on how everyone behaves and what everyone believes. If this is a property of rationality, then it cannot be of *individual* rationality but rather of *collective* rationality.

This last point is confirmed by the fact that the CPA is basically equivalent to assuming that individuals are “like-minded” (Bacharach 1985) or “symmetric reasoners” (Hédoin 2014). In other words, instead of interpreting a prior as an *ex ante* belief and a common prior as a formal agreement over this belief, it may be better to interpret the CPA as an assumption regarding the *reasoning* and *inference modes* of the players. To say that the players have a common prior would then mean that they reason in the same way on the basis of some information rather than that they share a fundamental belief over the probability distribution of events. As I will argue below, the Dennettian view licenses this interpretation and gives to the CPA both a clear ontological and methodological status. But before arguing for this, reconsider briefly Bacharach’s notion of experiment $e = \{w, \eta, y\}$ discussed in section 4. Bacharach assumed that each player i knows his information function $\eta: \Omega \rightarrow Y$ and showed on this basis and other conditions that an information partition I_i can be derived. Now, assume that η is shared and commonly known among the players in what can be called a “common experiment”. It follows that it is commonly known that all players will have the same partition I . As a consequence, any signal $y \in Y$ will lead all players to make exactly the same commonly known inference regarding what is possible. In other words, each signal indicates an event that is common knowledge.²⁴ In substance, this is the Harsanyi-Aumann doctrine but without supposing that the players have a (common or heterogeneous) prior. The common knowledge of the function $\eta: \Omega \rightarrow Y$ captures the fact that the players are η -symmetric reasoners and that this is commonly understood in the population.

The Dennettian view provides support to this last interpretation in the same way that it supports the IPA. The intentional strategy consists in ascribing to an entity a set of intentional states permitting the explanation and the prediction of the entity’s behavior under the assumption that the entity is rational, *i.e.* it has the intentional states it *ought to* have. As it has been already pointed out, this ascription depends on the functional relationship between the entity’s intentional states and its environment. But it clearly also relies on the underlying rationality that is attributed to the entity. At one extreme, an entity behaving authentically randomly cannot be ascribed intentional states as its behavior is unpredictable. Some consistency at the behavioral level is thus required. From this point of view, the intentional strategy proceeds exactly along the same lines than game theory: “It is a sort of holistic logical behaviorism because it deals with the prediction and explanation from belief-desire profiles of the actions of whole systems (either alone in environments or in interaction with other intentional systems), but it treats the individual realizations of the systems as black boxes. The *subject* of all the intentional attributions is the whole system (the person, the

²⁴ We can characterize formally a common knowledge event in an epistemic game through the possibility operators R_i and the related knowledge operators K_i . Define the common possibility operator R^* as the transitive closure of the individual possibility operators R_i . Then the event that the event E is common knowledge is defined as $K^*E = \{w \mid R^*(w) \subseteq E\}$.

animal, or even the corporation or nation)” (Dennett 1987, 58, emphasis in original). Both game theory and the intentional strategy do not care regarding how the system (the entity) are implemented: both characterize (mathematically and intentionally respectively) the behavior of the system as a whole without any reference to the specific intrinsic processes that lead to the behavioral performance. Now, I would argue that the CPA in game theory is, along with the expected utility maximization assumption, a “regulating principle” providing the background against which the behavior of the whole system can be holistically interpreted. In other words, the ascription of beliefs and preferences in a game-theoretic model (respectively reflected by the functions $p_{i,w}(\cdot)$ and $u(\cdot)$) depends on the assumption that the players share a common prior or, equivalently, that they are η -symmetric reasoners.

On this view, the CPA is not a substantive or empirical assumption. It is not even a “theoretical” hypothesis in a standard sense of being a refutable one. On the Dennettian view, the CPA is actually *constitutive* of game-theoretic reasoning in the same way that the intentional stance necessarily relies on an assumption of rationality: using the intentional strategy to predict an entity’s behavior without presupposing that this entity is minimally rational is simply nonsense. Similarly, all epistemic models in game theory have to rely on the CPA because any difference in prior beliefs (or in reasoning modes) can be potentially reflected in an augmented state space with a common prior.²⁵ Dennett’s functionalism departs from eliminativism precisely because Dennett argues, against the eliminativists, that the intentional stance is actually required even in scientific matters. The same is true for the CPA which is a required part of a strategy to explain and predict the agents’ behavior: you simply have to assume that the players are like-minded at the bottom. In this sense, the CPA can be interpreted both along *methodological* and *ontological* intentional-stance functionalisms (see section 3): it is not only a useful and convenient assumption at the methodological level as several pragmatic justifications indicate; it also corresponds to a real pattern common to all strategic interactions between rational systems (individuals but also machines or corporations).

This defense of the CPA builds on a tacit symmetry hypothesis: whatever is true for the observer/modeler is assumed to be true for the players themselves. In some way, this is not surprising as the rational expectation hypothesis is precisely based on such a symmetry: the agents are deemed to be as knowledgeable as the economists who conceived the relevant economic theory. I have interpreted the CPA as a problem-solving device for the modeler: it is a constitutive part of the whole game-theoretic approach to explain the behavior of rational agents. The Dennettian view provides a realist basis for this interpretation: players “really” are like-minded at some fundamental level and this provides the very basis for the modeler to explain their behavior. Now, it should be clear that what is known by the modeler, *i.e.* the broad theory $\mathcal{J}$, must be known by the players themselves. This is due to a technical point already emphasized in Aumann (1987): the information partitions I_i and the common prior $\mathbf{p}$ are necessarily “commonly known” by the players. This common knowledge is informal in the sense that it is not expressed in terms of the knowledge operators and results from the fact that each state w is a complete specification of whatever is relevant, including the players’ prior and information partitions. Given the fact that $\mathbf{p}$ and the partitions I_i are the same at all

²⁵ Aumann (1998) seems to argue for a similar conclusion but of course with very different arguments.

w, it must be true that they are known by everyone and that this is common knowledge.²⁶ On the Dennettian view, this is reasonable and even required: in a strategic context, the players themselves have to take the intentional stance to interpret and predict other players' behavior. Therefore, one needs the same assumption that the other players are reasoning like him at some fundamental level to use the intentional strategy. While the (common) knowledge of the CPA may be interpreted as the fact that the players know that they are symmetric reasoners as suggested above, the (common) knowledge of information partitions may be interpreted in terms of "mutual accessibility" (Gintis 2009). More generally, at a substantive level, these assumptions may reflect the basic fact that individuals (including the modeler) share a common culture that makes some states of affairs obviously observable and interpretable.

6. Common Knowledge of Rationality and the Objection to the Symmetry Hypothesis

A player is said to be Bayesian rational if i) he maximizes his expected utility given his subjective belief and ii) his subjective belief is consistent with his information in the sense of Bayes' law. Bayesian rationality is quite standard in decision theory since Savage's seminal contribution (Savage 1954). In particular, Savage showed that if an individual's choices over a set of uncertain prospects (probability distributions over outcomes) satisfy a set of axioms (especially a transitivity and an independence axioms), then these choices may be *represented* as the maximization of some expectational (*i.e.* linear in probabilities) utility function.²⁷ Moreover, Savage showed how both preferences over outcomes and subjective beliefs could be determined simultaneously on this basis. As I said above, the epistemic program in game theory has essentially consisted in reintroducing Bayesian decision theory in the study of strategic interactions. To say that the players are Bayesian rational is thus to assume that the utility functions u_i are expectational and that the posterior probabilities $p_{i,w}$ are determined according to Bayes' law. If the players are Bayesian rational at all states w in a given epistemic game Γ_w , then this is necessarily common knowledge as it corresponds to an event $E = \Omega$.²⁸

The Dennettian view provides a strong support to the Bayesian rationality assumption for very similar reasons than in the case of revealed preference theory extensively discussed by Ross (2005). Ross convincingly shows that the consistency axioms of revealed preference theory are problem-solving devices used to uncover the economic agents' preferences. These axioms are not substantive propositions regarding individual rationality but rather tools to interpret and predict agents' behavior. In a Dennettian perspective, the preferences revealed through

²⁶ Bacharach (1985) shows that it is possible to give this statement a formal expression. Note moreover that even if the prior and the information partitions were variable across Ω , it would still be possible to define an augmented state space where they would be both invariant.

²⁷ Expectational utility functions are unique up to a positive affine transformation. Note that it is controversial whether these functions measure *cardinal* or *ordinal* utility, *e.g.* Weymark (2005). This point does not need to be settled here.

²⁸ Following Aumann and Dreze (2008), I have assumed throughout this paper that the players' preferences over outcomes (*i.e.* strategy profiles) as represented by the utility functions u_i are invariant across Ω , which implies that they are commonly known. As Bayesian rationality is also common knowledge, the only missing piece for a player to accurately predict others' choices is a knowledge of their belief. Thus, the whole epistemic apparatus used in this paper points to the importance of second-order beliefs (beliefs over beliefs) in the characterization of rational expectations. Note that this point is almost always ignored in the discussions of the rational expectation hypothesis in macroeconomics.

this explanatory and predictive strategy are nonetheless “real” as they correspond to behavioral patterns resulting from the interaction between the agents’ latent cognitive processes and their institutional environment (e.g. market prices). A similar defense can be made for Bayesian decision theory (more precisely, expected utility theory), even though there are empirical reasons to doubt that individuals’ behaviors satisfy Savage’s axioms. The point is that Savage’s decision theory is clearly an instance of “holistic logical behaviorism” where intentional states are ascribed to a system as a whole. It is virtually always possible to rationalize an agent’s behavior as a maximization of expected utility provided that we use the “right” description of the decision problem. This is nothing but a form of externalism: an agent does not have inner preferences and beliefs (and thus an intrinsic expectational utility function) corresponding to neural processes; these preferences and beliefs are ascribed by an observer using the intentional stance given the functional relationship between the agent’s intentional states and the relevant environment. Now, as long as these preferences and beliefs are reflected in behavioral patterns that can be tracked by scientific methods (such as through the construction of utility functions or of games more generally), they are real in a well-defined sense.

Therefore, I would argue that on the Dennettian view, Bayesian rationality is essentially unproblematic as a scientific assumption. Things are different however regarding *common knowledge* of Bayesian rationality. Of course, this assumption appears to follow from the tacit symmetry hypothesis discussed above: whatever is known by the modeler (the broad theory $\mathcal{N}$) should be known by the players as the intentional stance is used by everyone. If Bayesian rationality follows from the intentional stance, then the players must know that others are Bayesian rational. But as soon as one realizes that others are also using the intentional stance, he must deduce that others know that he is Bayesian rational. Incidentally, this derivation of the CKBRA is very similar to the one that results from Muth’s (1961, 316) classic statement that rational expectations are “essentially the same as the predictions of the relevant economic theory”. As soon as the “relevant economic theory” includes expected utility maximization, knowing the theory implies knowing that everyone maximizes expected utility. But as it is *de facto* also part of the relevant theory, everyone knows that, and so on (Aumann and Dreze 2008, 81). However, while the symmetry hypothesis was mostly unproblematic in the case of the CPA and IPA, the same cannot be said here essentially because of well-known epistemic paradoxes the CKBRA gives rise to.

Epistemic paradoxes in game theory have been particularly studied in the case of finite extensive-form games with perfect information. In these games, the assumption that rationality is common knowledge licenses the use of the backward induction algorithm to find the solution. However, a well-known difficulty with this approach is that it relies on the use by the players of a reasoning process that takes into account impossible counterfactuals [(Bicchieri 1993); (Binmore 1987)]: the backward induction algorithm depends on assigning a choice at nodes which each player knows that they cannot be attained given common knowledge of rationality. However, if one of those nodes is indeed attained, then the player’s broad theory of the game is obviously false as that there cannot be common knowledge of rationality. The players are then left without any alternative theory and cannot formally make any choice. According to Bicchieri (1993, 134, my emphasis), this is a strong argument against the symmetry hypothesis: “Backward induction as a *reductio* proof is a proof given *outside the game by an external observer*. If we instead want to model *how the players*

themselves reason to an equilibrium, we have to model how they come to decide that a given action is optimal for them”. Bicchieri’s point is that while backward induction may be a good algorithm for the game theorist to find the solution of a game, it cannot be used by the players themselves and therefore it cannot reflect their actual reasoning. This indicates that common knowledge of rationality is neither sufficient nor necessary for players to make a choice in an extensive-form game and that actually it can lead to logical inconsistencies in case of deviations.²⁹

Since we are exclusively concerned with normal-form games, it can be thought that the above analysis does not apply. However, I think this is mistaken as the state space in any epistemic game Γ_w is built to encompass all the counterfactual scenarios that may have happened. Though what happens in a game situation is entirely specified by the actual state w , the corresponding epistemic game also indicates what would have happened at any other state w . Now, we are perfectly entitled to suppose that the interaction at any state w corresponds to some strategic path in an extensive-form game. In other words, each state w corresponds to a possible path in such an extensive-form game and the epistemic game Γ_w is therefore a complete description of what each player would have done in each counterfactual situation.³⁰ By assuming that the players are Bayesian rational over the whole state space Ω , it is therefore clear that we could in principle recover the epistemic paradoxes of the preceding paragraph.

I do not intend to mean that the CKBRA is necessarily false or problematic. There are cases where such common knowledge can indeed obtain as part of the intentional strategy. As documented by Chwe (2003), public events play an important role in the organization of human societies. By definition, these events are common knowledge in the relevant population.³¹ Formally, the fact that player i is Bayesian rational is itself an event. Can Bayesian rationality be a public event? There is no reason to think that this cannot be so and thus in some cases the CKBRA is warranted. However, while the IPA and the CPA can be argued to be constitutive parts of the intentional strategy under a game-theoretic formalism, I cannot see why this should be the case for the CKBRA. Beyond the problem of the epistemic paradoxes surveyed above, the point is that such an assumption will generally be unnecessary to predict an entity’s behavior through the intentional stance. From a game-theoretic point of view, though CKBRA (together with IPA and CPA) is sufficient to predict that a correlated equilibrium will be played in some game, it is not a necessary condition. The same is true on the intentional stance: ascribing a common knowledge of rationality may help to predict a system’s behavior but most of the time this will not be required. This is consistent with the symmetry hypothesis, as the above is true both for the modeler and for the players. It follows

²⁹ Bicchieri (1993, 142-155) establishes that for any extensive-form game with perfect information the sufficient number of levels of knowledge to infer the backward induction solution is the same than the necessary number. The point is that if the actual number of levels of knowledge is inferior or superior to this required number, then the backward induction solution cannot be worked out.

³⁰ For a similar interpretation, see Stalnaker (1998, 40, emphasis in original): “The original idea of the early developers of game theory, as I understand it, was that all the strategically relevant features of the dynamic interaction could be represented in the dispositions that the players had at the beginning of the game. At least if players are fully rational, then what they will decide to do when a certain situation arises can be assumed to be the same as what they *would* do if that situation *were* to arise... We can interpret a strategy choice, not as an instantaneous commitment, but as a representation of what the player will and would do in the course of the playing of the game”.

³¹ Formally, a public event PE is an event E such that K^*E for all $w \in E$.

that on the Dennettian view, the rational expectation hypothesis as formalized by Aumann and Dreze in a game-theoretic framework may ultimately be considered as a not fully warranted hypothesis.³²

7. Conclusion

Aumann and Dreze's (2008) characterization of the rational expectation hypothesis in a strategic context is grounded on three assumptions: IPA, CPA and CKBRA. Then, a player's rational expectation in a given game situation is his conditional payoff to a correlated equilibrium defined by a common prior over some state space. Though the rational expectation hypothesis in macroeconomics is sometimes criticized for its lack of realism and empirical plausibility, I have built on the postulate that in a strategic context its relevance depends on what is considered as the appropriate theory of intentionality. I have thus evaluated the three assumptions according to what I have called the Dennettian view of intentionality, a specific variant of externalism and functionalism, two dominant paradigms in the philosophy of mind.

The result is somewhat mixed. On the one hand, I have argued that the Dennettian view provides a strong support to the IPA and the CPA. Regarding the latter however, the Dennettian view calls for a reinterpretation of the "prior" notion as a formal characterization of the players' modes of reasoning rather than as a subjective belief. On the other hand, the CKBRA cannot be fully defended along the lines of Dennett's intentional-stance functionalism. This assumption is simply unnecessary from the intentional stance most of the time. As a consequence, the rational expectation hypothesis in strategic contexts is not fully warranted by the Dennettian view though it cannot be wholly rejected.

References

- Aumann, Robert J. 1976. "Agreeing to Disagree." *The Annals of Statistics* 4 (6): 1236–39.
- . 1987. "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica* 55 (1): 1–18.
- . 1998. "Common Priors: A Reply to Gul." *Econometrica* 66 (4): 929–38.
- . 1999. "Interactive Epistemology I: Knowledge." *International Journal of Game Theory* 28 (3): 263–300.
- Aumann, Robert J., and Jacques H. Dreze. 2008. "Rational Expectations in Games." *The American Economic Review* 98 (1): 72–86.
- Bacharach, Michael. 1985. "Some Extensions of a Claim of Aumann in an Axiomatic Model of Knowledge." *Journal of Economic Theory* 37 (1): 167–90.

³² Note that the argument of the preceding section where I have assumed common knowledge of the information function η to reinterpret the CPA cannot be used here. One of the reason is that under this interpretation, we no longer assume that the players have a probability measure over Ω . As a consequence, the players no longer maximize *expected* utility. Another reason is that even though each player may infer the common knowledge of rationality from the common knowledge of the function η and the common knowledge of the players' preferences, this is not an assumption (*i.e.* an *axiom*) but an *implication* (a theorem) of the broad theory $\mathcal{J}$.

- . 1993. “When do we have information partitions?” In M. Bacharach, M. Dempster and J. Enos (Eds.), *Mathematical Models in Economics*, University of Oxford, 1-23.
- . 1994. “The Epistemic Structure of a Theory of a Game.” *Theory and Decision* 37 (1): 7–48.
- Bicchieri, Cristina. 1993. *Rationality and Coordination*. CUP Archive.
- Binmore, Ken. 1987. “Modeling Rational Players: Part I.” *Economics and Philosophy* 3 (2): 179–214.
- Binmore, K. G. 2007. *Playing for Real: A Text on Game Theory*. Oxford University Press.
- Brandenburger, Adam. 2014. *The Language of Game Theory: Putting Epistemics Into the Mathematics of Games*. World Scientific.
- Broome, John. 1989. “Should Social Preferences Be Consistent?” *Economics and Philosophy* 5 (1): 7–18.
- Burge, Tyler. 1986. “Individualism and Psychology.” *The Philosophical Review* 95 (1): 3–45.
- Chwe, Michael Suk-Young. 2003. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton University Press.
- Crane, Tim. 2015. *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representation*. Routledge.
- Dennett, Daniel C. 1989. *The Intentional Stance*. MIT Press.
- . 1991. “Real Patterns.” *The Journal of Philosophy* 88 (1): 27–51.
- Geanakoplos, John. 1992. “Common Knowledge.” *Journal of Economic Perspectives* 6 (4): 53–82.
- Gintis, Herbert. 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton University Press.
- Gul, Faruk. 1998. “A Comment on Aumann’s Bayesian View.” *Econometrica* 66 (4): 923–27.
- Halpern, Joseph Y., and Yoram Moses. 1990. “Knowledge and Common Knowledge in a Distributed Environment.” *J. ACM* 37 (3): 549–87.
- Halpern, Joseph Y., and Riccardo Pucella. 2011. “Dealing with Logical Omniscience: Expressiveness and Pragmatics.” *Artificial Intelligence*, 175 (1): 220–35.
- Hédoin, Cyril. 2014. “A Framework For Community-Based Salience: Common Knowledge, Common Understanding and Community Membership.” *Economics and Philosophy* 30 (3): 365–95.
- Lau, Joe, and Max Deutsch. 2014. “Externalism About Mental Content.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2014. <http://plato.stanford.edu/archives/sum2014/entries/content-externalism/>.
- Lismont, Luc, and Philippe Mongin. 1994. “On the Logic of Common Belief and Common Knowledge.” *Theory and Decision* 37 (1): 75–105.
- Mongin, Philippe. 1995. “Consistent Bayesian Aggregation.” *Journal of Economic Theory* 66 (2): 313–51.
- Morris, Stephen. 1995. “The Common Prior Assumption in Economic Theory.” *Economics and Philosophy* 11 (2): 227–53.
- Muth, John F. 1961. “Rational Expectations and the Theory of Price Movements.” *Econometrica* 29 (3): 315–335.
- Putnam, Hilary. 1975. “Language and reality.” *Mind, language and reality* 2: 272-290.
- Ross, Don. 2002. “Dennettian Behavioural Explanations and the Roles of the Social Sciences.” in A. Brooks and D. Ross (eds.), *Daniel Dennett*, Cambridge University Press, 140-83.
- . 2005. *Economic Theory And Cognitive Science: Microexplanation*. MIT Press.
- . 2014. *Philosophy of Economics*. Palgrave Macmillan.

- Savage, Leonard J. 1954. *The Foundation of Statistics*. Courier Dover Publications.
- Searle, John R. 2004. *Mind: A Brief Introduction*. Oxford University Press.
- Sillari, Giacomo. 2006. "Models of awareness." In *Proceedings of the 7th conference on logic and the foundations of game and decision theory (LOFT)*, 209-218.
- . 2008. "Quantified Logic of Awareness and Impossible Possible Worlds." *The Review of Symbolic Logic* 1 (4): 514–29.
- Skyrms, Brian. 1996. *Evolution of the Social Contract*. Cambridge University Press.
- Smith, Vernon L. 2009. *Rationality in Economics: Constructivist and Ecological Forms*. Cambridge University Press.
- Stalnaker, Robert. 1998. "Belief Revision in Games: Forward and Backward induction." *Mathematical Social Sciences* 36 (1): 31–56.
- . 2006. "On Logics of Knowledge and Belief." *Philosophical Studies* 128 (1): 169–99.
- Weymark, John A. 2005. "Measurement Theory and the Foundations of Utilitarianism." *Social Choice and Welfare* 25 (2-3): 527–55.