A Bayesian Conundrum: From Pragmatism to Mentalism in Bayesian Decision and Game Theory

Cyril Hédoin[*]

University of Reims Champagne-Ardenne, France

Version 1.2: 30/04/2017

Abstract: This paper discusses the implications for Bayesian game theory of the behaviorism-versus-mentalism debate regarding the understanding of foundational notions of decision theory. I argue that actually the dominant view among decision theorists and economists is neither mentalism nor behaviorism, but rather pragmatism. Pragmatism takes preferences as primitives and builds on three claims: i) preferences and choices are analytically distinguishable, ii) qualitative attitudes have priority over quantitative attitudes and iii) practical reason has priority over theoretical reason. Crucially, the plausibility of pragmatism depends on the availability of the representation theorems of Bayesian decision theory. As an extension of decision-theoretic principles to the study of strategic interactions, Bayesian game theory also essentially endorses the pragmatist view. However, I claim that the fact that representation theorems are not available in games makes this view implausible. Moreover, I argue that pragmatism cannot properly account for the the generation of belief hierarchies in games. If the epistemic program in game theory is to be pursued, this should probably be along mentalistic lines.

Keywords: Bayesian synthesis – Bayesian game theory – Pragmatism – Mentalism – Preferences

Word count: 13000 (references included)

[*] Professor of Economics, REGARDS (EA 6292) economics and management research center.
Contact: cyril.hedoin@univ-reims.fr
A previous version of this paper has benefited from the comments from Jean Baccelli, Mikaël Cozic and Samir Okasha. All errors and omissions are mine.

1. Introduction

In two recent papers, Samir Okasha (2016) and Franz Dietrich and Christian List (2016) have reflected over the distinction between *mentalism* and *behaviorism* in decision theory and economics. Mentalism and behaviorism broadly refer to two sets of methodological and conceptual commitments over the nature and the measure of foundational concepts like preferences, beliefs and desires. These concepts are indeed central both in decision theory and in economics. They are routinely used in a positive perspective to describe and explain people's choices and their collective consequences. They also are essential from a normative perspective as key debates concern the properties that people's preferences and beliefs *ought to* satisfy as rational beings. In economics, the distinction between mentalism and behaviorism has come under the spotlight especially due to the rise of behavioral economics over the last three decades. The latter has largely contributed to feed a growing discomfort with the broadly behaviorist methodology that has prevailed in economics since the 1930s.[1] In decision theory, the conflict between mentalism and behaviorism dates back to the pioneering work of Frank Ramsey (1926) and Leonard Savage (1954). While these mathematicians have generally understood their contributions in a more or less behaviorist perspective, philosophers have traditionally rather entertained a mentalist understanding of decision theory.

Quite strikingly, Okasha (2016) on the one hand and Dietrich and List (2016) on the other do not use the notions of mentalism and behaviorism in the same way. According to Okasha, the former "treats an agent's utility function and credence function as psychologically real, and capable of causing/explaining their preferences and choices", while the latter "regards preferences or choices as primary and utilities and credences as derivative". Dietrich and List define mentalism as the view according to which preferences, utilities and credences are "representations of real mental or psychological phenomena" while behaviorism take these notions to be "mere re-descriptions of behavioral patterns". This divergence regarding the definitions of these key notions reflect the more general conceptual opacity that continues to surround decision theory and economics. Still, I do not aim here to resolve this difficulty. My more modest objective is to highlight some of the implications of the behaviorism/mentalism debate for what is now called the epistemic program in game theory [e.g. (Brandenburger 2014); (Perea 2012)]. The epistemic program is generally characterized as the extension of *Bayesian* decision theory to the study of strategic interactions. Bayesian game theory is then the research program dedicated to the study and the characterization of the behavior of rational players who happen to have some degree of mutual knowledge/belief of their Bayesian rationality. My main point is the following one. I distinguish between three possible broad understandings of the key notions of decision theory depending on what they take to be the primitive concepts: *behaviorism* (where primitives are actual or hypothetical choices), *pragmatism* (where primitives are preferences) and *mentalism* (where primitives are the utilities and probabilities). While the three understandings can be somehow defended, I argue that pragmatism is actually what most decision theorists and economists would be ready to endorse. Crucially, the very plausibility of pragmatism depends on the availability of the various representation theorems of Bayesian decision theory. However, I claim that representation theorems cannot be directly transferred in the context of strategic interactions studied by Bayesian game theory. It follows that pragmatism *is not* a plausible option for the latter, in spite of the fact that it is the tacit understanding endorsed by most game theorists. However, I claim that the fact that

[1] For an illustration, see the various contributions in (Caplin and Schotter 2010).

representation theorems are not available in games makes this view implausible. Moreover, I argue that pragmatism cannot properly account for the generation of belief hierarchies in games. If the epistemic program in game theory is to be pursued, this should probably be along mentalistic lines.

The rest of the paper is organized as follows. Sections 2 and 3 distinguish and characterize behaviorism, pragmatism and mentalism in the context of what I call the "Bayesian synthesis" in decision theory. Section 4 argues that pragmatism is the view endorsed by many economists and decision theorists but also emphasizes that its plausibility depends on the existence of representation theorems. Section 5 highlights that the extension of representation theorems to Bayesian game theory is not straightforward and probably doubtful. Section 6 discusses the status of belief hierarchies in games and their implications for the comparative relevance of pragmatism and mentalism. Section 7 concludes.

2. The Bayesian Synthesis in Decision Theory

Modern decision theory, at least in the versions used by economists and discussed by philosophers, is essentially *Bayesian*. There is no general agreement over what are the characteristics of Bayesian decision theory. Since the mid-20th century, several variants have been developed differing over the specific conditions regarding the rationality of the decision-maker or the structure of decision problems. Economists have generally taken Savage's (1954) subjective expected utility theory as the canonic version of Bayesian decision theory without much reflection.[2] Bayesian decision theory does not reduce to Savage's version however and it might be useful to locate the whole discussion over mentalism and behaviorism in the context of what Richard Bradley (2016) characterizes as the *Bayesian synthesis*. The Bayesian synthesis refers to the conjunction of principles characterizing the structure of decision problems and the properties of what constitutes a rational choice. I will refer to these propositions as *Probabilism*, *Framing* and *Rationality*:[3]

> *Probabilism*: Rational degrees of beliefs obey the laws of probability.[4]
>
> *Framing*: The decision-maker's choices range over *acts*, i.e. functions from states to consequences.
>
> *Rationality*: i) Rational choice consists in choosing the act with the maximum desirability; ii) the desirability of an act corresponds to its subjective expected utility according to the decision-maker.

[2] Of course, the last three decades have seen the development in economics (and psychology) of authentic non-Bayesian decision-theoretic accounts. Cumulative prospect theory (Tversky and Kahneman 1992) and case-based decision theory (Gilboa and Schmeidler 1995) are among the most prominent examples. At this time however, they have not been significantly used to study strategic interactions. Note moreover that the debate between mentalism and behaviorism essentially carries over to these non-Bayesian accounts.

[3] Bradley (2016) characterizes the Bayesian synthesis in a slightly different way but ultimately his definition encompasses the same set of works as mine.

[4] If Ψ is the space of events E, then any probability function $p(.)$ satisfies the three following laws: i) $\forall E \in \Psi$: $p(E) \geq 0$, ii) $p(\Omega) = 1$ with $\Omega = \bigcup_{i=1}^{\infty} E_i$, iii) $\forall E_1, E_2, \ldots, E_n \in \Psi$: $p(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} p(E_i)$, with $E_1, E_2, \ldots$ any sequence of mutually exclusive events.

Probabilism is essentially the received-view among decision theorists. Broadly, it implies that the credences (i.e. degrees of beliefs) of a rational decision-maker have to correspond to probability measures. The well-known Dutch book arguments are generally taken to provide a strong reason to endorse probabilism (Hacking 2001). Note that probabilism also implies that one's conditional beliefs must correspond to conditional probabilities. Framing is a more restrictive and contentious condition. It assumes that it is possible to partition the space of events (or propositions) into states and consequences (or outcomes). States correspond to events over which the decision-maker has no causal control, i.e. the probabilities ascribed to states are assumed to be stochastically independent. Consequences are events that result from the conjunction of an act and a state. Both states and consequences are taken to be maximally specific events, i.e. they correspond to propositions about everything that is relevant from the decision-maker's perspective. Savage's subjective expected utility theory is of course the first and most prominent Bayesian decision-theoretic account to have framed the decision-maker's options as acts in this sense. Note however that Savage imposed a condition of state-independence according to which the evaluation of consequences should be independent from the states in which they are located. Such a condition is dubious both from a positive and a normative perspectives but can be easily weakened to allow for *state-dependent* evaluations.

The Rationality condition is the conjunction of two more fundamental claims. First, it establishes a direct relationship between the evaluations of acts and the rationality of choice: a rational choice is to pick out (one of) the most desirable option(s) according to one's evaluation. Second, the desirability of an act corresponds to the expectation of its utility, measured on the basis of a unique probability function $p(.)$ and a cardinal utility function $u(.)$ defined according to some interval scale. Denoting $\{S_i\}$ and $\{C_{\alpha i}\}$ the partitions of events corresponding to the set S of states and the set C of consequences respectively, the desirability of an act α: $S \rightarrow C$ is then (assuming that S is finite)

(1) $$Des\ \alpha = \sum_{i=1}^{n} p(S_i).u(\alpha(S_i) \cap S_i) = \sum_{i=1}^{n} p(S_i).u(C_{\alpha i} \cap S_i).$$

This characterization of the Bayesian synthesis encompasses a broad range of variants of decision theory, including Ramsey's (1926), Anscombe and Aumann's (1963) as well as various formulations of the so-called "causal decision theory" [e.g. (Joyce 1999); (Lewis 1981)]. Note however that it excludes some versions that have traditionally been characterized as Bayesian, the most prominent being Richard Jeffrey's (1990) "evidential" decision theory.[5]

3. Behaviorism, Pragmatism and Mentalism

This section characterizes the different ways the Bayesian synthesis can be interpreted on the basis of a conjunction of conceptual and methodological claims. As it will appear, the different interpretations depend on which foundational notions in decision theory are regarded as primitives. The utility function $u(.)$ and the probability function $p(.)$, denoting degrees of desires and beliefs respectively, are part of these foundational notions. Another one is the notion of

[5] Jeffrey's decision theory endorses Probabilism and the first part of Rationality. It does not satisfy Framing since it allows probabilistic dependence between states and the decision-maker's choices (thus rejecting the states/consequences distinction). As a result, Jeffrey's evidential account evaluates the desirability of an action according to its "news value", i.e. its value conditional on the fact that it is actually chosen.

preference which is formalized through the binary weak relation $\succcurlyeq$. It is possible to characterize $\succcurlyeq$ through a formal equivalence with the pair of utility and probability functions:

(P) For any pair of acts α, β: $\alpha \succcurlyeq \beta$ if and only if *Des* α ≥ *Des* β.

According to (P), an act is weakly preferred to another one if and only if it is at least as desirable.[6] Moreover, Rationality states that a rational choice consists in choosing the act with the maximum desirability. Denote A any set of available acts and $D(A) \subseteq A$ a *choice function* picking at least one act among the available ones. Then, Rationality entails

(D) $D(A) = \{\alpha \mid Des\ \alpha \geq Des\ \beta$ for any $\beta \in A\}$.

Obviously, $D(.)$ and the preference relation $\succcurlyeq$ can also be mutually expressed through a formal equivalence. We have thus the following formal equivalence between the utilities/probabilities pair < $u(.)$, $p(.)$ >, the preference relation $\succcurlyeq$ and the choice function:

For any set A of available acts and all α, β ∈ A,

$$\alpha \in D(A) \Leftrightarrow \alpha \succcurlyeq \beta \Leftrightarrow Des\ \alpha \geq Des\ \beta$$

This formal equivalence establishes that *in principle*, the conditions constitutive of the Bayesian synthesis can be expressed indifferently in terms of choices, preferences or pairs of utilities and probabilities. However, it is clear that a formal equivalence does not imply a conceptual identity and as a result the semantics of the Bayesian synthesis (i.e. its substantive meaning) partly depends on which of these foundational notions is regarded as the primitive one. Okasha (2016) and Dietrich and List (2016) precisely characterize the distinction between mentalism and behaviorism on this basis but interestingly do not retain the same partition.

Okasha (2016) characterizes behaviorism as the interpretation that takes either choices or preferences as its primitives.[7] Dietrich and List (2016) restricts behaviorism to the interpretation taking choices as primitives and do not seriously consider the possibility of giving the utilities/probabilities pairs this status, except maybe through what they take to be a misguided form of psychological reductionism. Okasha's and Dietrich and List's partitions are nothing more than typologies. As such, they are neither true nor false but rather more or less useful. In my view, both are insufficiently fine-grained: Okasha's characterization of behaviorism misses the fact that taking choices or preferences as primitives does indeed make a difference; Dietrich and List's is insufficiently specific regarding what is implied by taking utilities/probabilities pairs rather than preferences as primitives. Indeed, combining these two partitions reveals that there are at least three rather than two ways to interpret the Bayesian synthesis: *behaviorism* takes choice functions as primitives, *pragmatism* takes preference relations as primitives, *mentalism* (or strong mentalism) takes utilities/probabilities pairs as primitives. Each of these

[6] Strict preference and indifference are defined accordingly, in the usual way.

[7] As I understand it, the notion of "primitive" as it is widely used in the philosophical literature on decision theory refers to the concepts in terms of which other concepts are defined. As such, to focus on the choice of a (set of) primitive(s) is obviously not sufficient to characterize any scientific account. In particular, primitive notions are not necessarily the most foundational ones given an underlying ontology. For instance, while the binary relation "at least as long as" is arguably the primitive notion of any plausible scientific account of length, it does not seem that it should be regarded as more foundational than the corresponding quantitative measure of length. This highlights the fact that the choice of primitives is only one part of the story and that this choice is meaningful only alongside with corresponding methodological, conceptual and metaphysical commitments. I thank Mikaël Cozic for having pointed out this to me.

interpretations is grounded on methodological and conceptual claims regarding both the meaning of the various concepts and the ways various kinds of information can be gathered to give an empirical content to the theory.

I take behaviorism to be the conjunction of two claims. The first, conceptual claim is that preferences, utilities and probabilities have no independent meaning and are merely formal representations of choices. This is of course the interpretation favored by a significant number of economists who have endorsed the "revealed-preference" approach pioneered by Paul Samuelson (1938). The second, methodological claim is that behavioral evidence is the only legitimate one and that the empirical information feeding the theory should be restricted to actual behavior. Obviously, this excludes any kind of "subjective" and/or unobservable information derived by means like introspection or questionnaires.

Pragmatism refers to a broad set of views taking preferences as the fundamental notions to derive both choices and utilities/probabilities pairs. Though neither Okasha nor Dietrich and List identify this approach as an alternative to mentalism and behaviorism, it is routinely characterized and used by decision theorists.[8] Contrary to behaviorism, it is hard to identify pragmatism with specific methodological claims regarding the relevant "informational basis": some variants of pragmatism retain the behaviorist's commitment to the exclusive use of choice data, while others allow for other kinds of more subjective information. However, I submit that three conceptual or theoretical claims are common to all versions of pragmatism:

(1) Choices and preferences are two conceptually separate and independent notions.
(2) Primacy of qualitative attitudes over quantitative attitudes.
(3) Primary of practical reason over theoretical reason.

The first claim is essential in the departure of pragmatism from behaviorism. It is reminiscent of Amartya Sen's (1973) argument that preferences have to be conceptually distinguished from observable behavior for the rationality axioms to make sense at all. Accordingly, on the basis of this first claim, the *formal equivalence* between the choice function *D*(.) and the preference relation $\succcurlyeq$ that I have stated above should not be mistaken as an *ontological identity*. A significant implication is that this allows for the possibility of "counter-preferential" choices.[9] The two other claims concern the relationship between the preference relation and the utilities/probabilities pairs. The second claim can be interpreted both as a methodological or a conceptual one. If we retain the latter, then it implies that the cardinal properties of utilities and probabilities does not necessarily carry over to preferences, i.e. in spite of the fact that probabilities and utilities are cardinal measures, pragmatists consider that preferences are merely *ordinal*. In turn, that means that the rational properties of quantitative attitudes are derived from the rational properties of qualitative attitudes: "it is the fact that the quantitative

[8] See for instance Joyce (1999) or Bradley (2016).

[9] It is worth pointing out that the formal independence between the choice function and the preference relation is not due to an insufficiently fine-grained framing of the decision problem. To understand this point, consider a partition $\{B_j\}$ of "background conditions" in which any decision problem can take place. Any B_j captures the features that may causally affect one's choices and preferences but that are not taken into account in the original formalization of the decision problem. Then, one may argue that it is always possible to recover the formal equivalence between choices and preferences by indexing the choice function and the preference relation to the background conditions, i.e. $\alpha \in D_{B_j}(A) \Leftrightarrow \alpha \succcurlyeq_{B_j} \beta$ for all α, β ∊ *A*. Pragmatism rejects this possibility however on the ground that if events are maximally specific, then a non-indexed preference relation $\succcurlyeq$ must hold between any pair of "refined acts" of the kind $\alpha \cap B_j$. Claim (1) then states that preferences defined over such refined acts are not necessarily revealed by a non-indexed choice function.

attitudes are rationally required to have certain relational properties that explains why the quantitative attitudes are rationally required to have corresponding numerical ones" (Bradley 2016: 62). This leads to a key feature of pragmatism (a feature that Okasha attributes to behaviorism as a whole): utilities and probabilities are merely numerical *representations* of preferences. I will return on this point in the next section when I discuss the role or representation theorems.

The third and last claim can also be given a methodological interpretation according to which the determination of beliefs is done through the determination of preferences. But it is its conceptual interpretation that is the most interesting here. Joyce (1999: 90) characterizes it in the following way: "The laws of rational belief are underwritten by the laws of rational desires. To call any belief rational or irrational is always to say something about the rationality of the believer's desires". Beyond these three claims, pragmatists differ in the way they interpret preferences. Two broad sets of interpretations are to be found (Bradley 2016: 63-7): "choice-theoretic" and "judgmentalist". The former takes preferences not as actual or hypothetical choices but as *dispositions* to choose. This interpretation maintains the conceptual distinction between choices and preferences because this is not a necessity that actual choices are generated by specific choice dispositions. Indeed, choice dispositions will lead to corresponding actual choices only if the appropriate circumstances hold. Still, because this interpretation maintains a conceptual link between choices and preferences, it is appropriate to label it "weak behaviorism". The judgmentalist account interprets preferences as mental attitudes. As I will spend most of the next section on it, I shall not give more details. Note however that it is natural to characterize this variant as "weak mentalism". As I argue below, pragmatism seems to be the view that currently prevails in decision theory and among a majority of game theorists and economists. It is also worth noting that, because of claim (3), both variants of pragmatism favor a functionalist view of mental states, especially beliefs. On a functionalist reading, beliefs are characterized through their constitutive or causal role in the generation of action and/or preference judgments. This view is in principle compatible with either a realist (i.e. naturalistic) or non-realist (i.e. instrumentalist) ontological account of scientific theories and models. However, it seems that the choice-theoretic brand of pragmatism ("weak behaviorism") can separate itself from behaviorism only by endorsing a realist stance as otherwise both views essentially conflate in spite of taking different notions as primitives.[10]

Mentalism (or "strong mentalism" to avoid confusion with the judgmentalist variant of pragmatism) is thus the view that takes utilities/probabilities pairs as primitives. Therefore, it rejects the methodological and conceptual claims (2) and (3) without taking a definite stance regarding the relationship between choices and preferences: on the one hand, quantitative attitudes are regarded as more fundamental than qualitative ones; on the other hand, practical reason and theoretical reason are viewed on an equal footing. In particular, while the issue of the rationality of desires/preferences and actions is regarded as pertaining to decision theory properly speaking, issues related to the rationality of beliefs are taken to belong to epistemology. Mentalism is rarely explicitly endorsed under this characterization either in economics or in the decision-theoretic literature. However, as suggested by Okasha (2016), it seems to be the received view among a dominant fraction of philosophers. Many philosophical

[10] See Dietrich and List's (2016) related distinction between "radical behaviorism" and "epistemic behaviorism". On the deep affinity between functionalism and what I characterize here as pragmatism, see for instance Zynda (2000), Christensen (2006) and Elliott (2017).

discussions related to decision theory are indeed conducted assuming pre-existing utility and probability functions. An illustration is provided by the early discussions over the so-called "causal decision theory" [e.g. (Lewis 1981); (Skyrms 1982)]. Perhaps more significantly, there has been recent attempts to vindicate different versions of the expected utility maximization criterion grounded on clearly mentalistic perspectives. A first example is Joyce (1999) who provides a representation theorem for causal decision theory. He explicitly rejects pragmatism and particularly claim (3). Instead, his representation theorem builds upon *two* separate binary relations, one corresponding to rational preferences and the other to rational comparative likelihood attitudes. Moreover, the latter are assumed to be representable by a probability function.[11] Both relations are articulated through a "coherence" condition formally similar to those found in other representation theorems but crucially it is not given a pragmatist interpretation. A second and more extreme example is given by Martin Peterson's (2002) defense of the expected utility maximization criterion in terms of "transformative decision rules". Peterson assumes that the decision-maker has well-defined utility and probability functions and proceed to show that expected utility maximization reduces to the choice of the best option in a uncertainty-free decision problem, where the latter is obtained through a sequence of "normatively reasonable" transformations of the original decision problem. I cannot delve into the details but the point is that we have an authentic mentalistic account of Bayesian decision theory.[12]

Outside the Bayesian synthesis, the recent developments in both positive and normative behavioral economics emphasize the distinction between "decision utility" and "experienced utility": while the former corresponds to the usual utility-as-representation-of-preferences, the latter rather refers to a hedonic state of mind that can be accessed by non-behavioral ways. As a whole, this distinction is part of the "back to Bentham" approach that has been pioneered by Daniel Kahneman and others (e.g. Kahneman, Wakker, and Sarin 1997), arguably in the context of non-Bayesian decision theory. The research program of neuroeconomics can also be viewed as defending a mentalistic account. In particular, the work of Paul Glimcher (2010) is illustrative of recent attempts to provide neural foundations for the concept of utility as a quantitative measure.[13] However, although behavioral and neural approaches are gaining momentum, mentalism remains marginal in economics.

The table below provides a tentative characterization of the three major understandings of the Bayesian synthesis in terms of their primitives, what they take to be the relevant informational basis and their ontological commitments toward mental attitudes and the concepts used to capture them:

[11] Not that "representable" does not mean "must be represented by". Joyce is not rejecting claim (2) of pragmatism.

[12] Peterson (2008) characterizes his account as non-Bayesian. However, this is because he restricts Bayesian decision theory to the use of representation theorems for vindicating the expected utility maximization criterion. Peterson's account satisfies all the constitutive features of the Bayesian synthesis though.

[13] See Fumagalli (2013) for methodological and theoretical reflections over this search for "true" utility.

Table 1

	Behaviorism	Pragmatism	Mentalism
Primitive notions	Choice functions	Preferences (either as choice dispositions or as judgments)	Utilities and probabilities (or the corresponding pair of binary relations)
Informational basis	Narrow (choice behavior)	Either narrow (weak behaviorism) or large (weak mentalism)	Large (introspection, surveys, neural data…)
Ontological commitments toward mental states and scientific concepts	Eliminativism/anti-realism toward mental states; scientific instrumentalism	Functionalist view of mental states; scientific realism (though instrumentalism is also plausible)	Realism toward mental states (desires and beliefs are causally relevant); scientific realism

4. Representation Theorems and Pragmatism

This section and the next will concentrate on the pragmatist view of Bayesian decision theory and more particularly on the weak mentalism variant. Indeed, as I suggest above, this is not far from being the dominant view among decision theorists and economists. Regarding the former, though the founding fathers of Bayesian decision theory are generally thought to have endorsed a form of behaviorism, they were actually less radical regarding the conceptual and methodological primacy of choices and choice data. For instance, Savage's (1954) definition of preferences was in terms of *conditional* choices. Perhaps more strikingly, Ramsey's approach to decision theory seemed to be compatible with the use of introspection and of some form of linguistic communication to elicit people's preferences (Bradley 2004). Though properly speaking not belonging to the Bayesian synthesis, Jeffrey's (1990) evidential decision theory also was not grounded on a conflation of choices and preferences and even allows for the anti-behaviorist notion of "preference among preferences" (Jeffrey 1974).[14]

The dominance of pragmatism in economics is less transparent due to the loudness of some revealed-preference theorists [e.g. (Binmore 2009); (Gul and Pesendorfer 2008)]. As characterized by Wade Hands (2013), contemporary revealed preference theory corresponds to the conjunction of three related claims. Two of them are just statements of the conceptual primacy of choices over preferences and utilities/probabilities pairs on the one hand, and of the methodological primacy of choice data on the other hand. The third claim is an imperialistic one making revealed preference theory the methodological core of all choice-theoretic economics. However, as Dietrich and List (2016) and others show, this last claim is unjustifiable from a philosophy of science perspective. Unsurprisingly, in their actual practice, most standard economists seem rather to endorse a more pluralist and flexible view. For instance, a plausible account of the economist's use of the concept of preferences is offered by

[14] Perhaps more controversially, I would argue that Donald Davidson's (2004) "unified theory of action and meaning" is also an instance of pragmatism, since it is grounded on the claim that the norms of rationality one is imposing on preferences are essential to make others' behavior intelligible.

Daniel Hausman (2011). Hausman criticizes the behaviorist/revealed-preference view and defends instead what he calls a “consequentialist” approach.[15] According to Hausman’s consequentialism, preferences should (and are *actually*) be taken as all-things-considered subjective and total comparative rankings of alternatives. Preferences over acts then summarize all the decision-maker’s practical and theoretical reasons for ranking an alternative over another one. These reasons are provided by the agent’s “basic preferences” (i.e. desires) and beliefs. Hausman’s account belongs to the judgmentalist side of pragmatism (i.e. weak mentalism) because while it obviously rejects behaviorism, it does not endorse strong mentalism as characterized above. It is not established that all economists would endorse Hausman’s consequentialist understanding of preferences. Still, it provides a plausible rationalization of their practices.

An alternative account of the economists’ use of decision-theoretic concepts is provided by Don Ross (2014). Ross defends an “anti-behaviorist-anti-behavioralist” account that he labels “neo-Samuelsonian”. According to this neo-Samuelsonian account, economic agency is constituted by choice patterns either at the individual or collective level. Though preferences are ultimately identified as actual or hypothetical choices (in particular on the basis of the use of the choice functions apparatus and axioms of revealed preference), it is recognized that latent cognitive functions and their interactions with the institutional environment play a key role in the determination of the agent’s behavior. Ross’s account is a sophisticated one, depending on an externalist understanding of intentional states according to which the semantic content of states like desires or beliefs is partially determined by the context in which the agent is embedded. Such externalism seems hardly compatible with any form of strong mentalism. At the same time, building on Daniel Dennett’s (1989) “intentional-stance functionalism”, Ross suggests that these intentional states are nonetheless real in the sense that they are explanatory of people’s behavior. Whether or not economists would agree with this characterization of modern choice-theoretic economics is unclear. Nevertheless, it provides a plausible and pragmatist-compatible rationalization of the economists’ practices.

Arguably, even though the distinction between pragmatism (in its weak mentalism variant) and strong mentalism is clearly delineated by the disagreement over the choice of primitives, it may look somewhat ambiguous at times. This is due to the fact that if preferences are interpreted as judgmental (and hence mental) attitudes, then they seem to be of the same nature than beliefs and desires. As I note above, it is true that preferences are generally taken to be qualitative attitudes while beliefs and desires are more naturally seen as quantitative attitudes. But this is not decisive for at least two reasons: first, one may easily define comparative and qualitative variants of beliefs and desires (e.g. I consider that *x* is more likely than *y*; I take *z* to be more desirable than *w*). Correspondingly, it is perfectly possible to construct a *quantitative* (cardinal) notion of preference, for instance through a quaternary preference relation.[16] From this perspective, there is no clear difference between desires understood as a folk psychological

[15] Hausman’s terminology is a bit confusing as it may mistakenly be taken to refer to *consequentialism* in the sense of Hammond (1988). Hammond’s consequentialism is an account about the proper way to evaluate alternatives in terms of their consequences. Though it is more general than Bayesian decision theory, the satisfaction of more or less restrictive axioms generates a formal equivalence with subjective expected utility theory. It is agnostic however regarding what notions are taken as primitives. For a critical discussion of Hammond’s consequentialism, see Levi (1997: chap. 4).

[16] A quaternary preference $\succcurlyeq^d$ relation is a defined over the domain A^2xA^2 rather than *A*x*A*. For any α, β, γ, δ ∈ *A*, a plausible reading of (α, β) $\succcurlyeq^d$ (γ, δ) is “α is more intensively preferred to β than γ to δ”.

notion and preferences. The primacy of practical reason over theoretical reason constitutive of pragmatism then seems in need of a justification. Second, though preferences may be primary both from a methodological and a conceptual point of view, it remains that in a causal perspective desires and beliefs are more fundamental. This is indeed what Hausman's consequentialism seems to suggest. Preferences are all-things-considered evaluations that may transpire either behaviorally or verbally (a kind of behavior admittedly) that are *caused* and *explained* by various reasons, starting with desires and beliefs, and possibly other factors (such as social norms). This is indeed perfectly in line with the standard understanding of folk psychology and the underlying idea that mental events are both causes of and explanations for actions. I see no reason to reject this possibility, though the notion of mental causation is notoriously controversial among philosophers of mind. But then, what is the basis for the three claims constitutive of pragmatism that I have highlighted in the preceding section? I submit that the only plausible answer lies in the role played by the various *representation theorems* to be found in the work of decision theorists.

Representation theorems in decision theory almost invariably consist into two related parts: an *existence theorem* establishes a formal correspondence between two abstract structures (e.g. a preference ordering and the expectation of a utility function) under a set of sufficient and necessary conditions; a *uniqueness theorem* states the extension of this correspondence (e.g. the class of utility functions whose expectation corresponds to a given preference ordering). The representation theorems of interest here are those that establish the isomorphism between a binary preference relation $\succcurlyeq$ satisfying a set of axioms (e.g. ordering, independence or sure-thing principle…) and a numerical representation of this relation in terms of a utility function $u(.)$ and a probability function $p(.)$. The formal relationship between these two abstract structures therefore goes in two directions: from $< u(.), p(.) >$ to $\succcurlyeq$, and from $\succcurlyeq$ to $< u(.), p(.) >$. The latter states the sufficient conditions for the preference relation to be represented by a pair of utility and probability functions. It therefore shows that the expectational representation (i.e. linear in probabilities) is justified in virtue of a set of "local" axioms satisfied by the preference relation. The former indicates what is demanded to preferences if a decision-maker is to act as if maximizing the expectation of a utility function. Representations theorem therefore fulfill at least three roles [e.g. (Bradley 2004); (Cozic and Hill 2015); (Meacham and Weisberg 2011)]. First, they provide a normative justification for two constitutive characteristics of the Bayesian synthesis: *Probabilism* and *Rationality*. Indeed, the left-to-right relationship in representation theorems establishes that if one's preferences satisfies some axioms, then it is as if one were maximizing the expectation of a class of utility functions. Thus, instead of asking if one ought to maximize expected utility, representation theorems allow to conduct the normative analysis on specific axioms which rationality can be more directly assessed. Correspondingly, representation theorems permit the characterization of any choice model in terms of preferences and the related rationality axioms. The second role concerns measurement: representation theorems provide an account of how to measure utilities and probabilities from an informational basis which may range from observed behavior to verbal statements. These first two roles are essential both from a positive perspective and a normative perspective. The third role played by representation theorems is semantic: they provide a resource to determine the meaning of foundational decision-theoretic notions. In particular, as I suggest in the preceding section, representation theorems may help to sustain the functionalist view of mental states that characterizes pragmatism. According to the latter, the nature and content of desires and beliefs is entirely determined by the causal or constitutive dependence they entertain with preferences.

The three roles are essential for sustaining the distinction between pragmatism and mentalism.[17] They indeed establish that utilities/probabilities pairs can be *defined* and *measured* on the basis of preferences. Moreover, they indicate that the justification and the evaluation of the expected utility maximization criterion entirely depend on the properties of the preference relation. Now, the fact that utilities and preferences are derivative of preferences does not mean that they do not "exist" or are not "real". Under a naturalistic ontology of the kind defended by Dietrich and List (2016), theoretical terms can be regarded as real as soon as they are part of the formal structure of some empirically successful scientific theory or model.[18] These terms constitute the "ontological commitments" of the theory, i.e. the objects, properties and relations that the theory needs to assume. In this perspective, even "non-observable" entities can be assumed to be real as long as they are constitutive of the theory and/or of its models:

> "we should (at least provisionally) accept that the entities, properties, and relations to which the theory is committed correspond to real entities, properties and relations… For example, if our best physical theories tell us that there are quarks, leptons and bosons, we have every reason to believe in this particles' existence, regardless of their unobservable status." (Dietrich and List 2016).

Though pragmatists do not take utilities and probabilities to be directly observable, they nonetheless are committed to regard these as "real" in the sense that these are somehow implied by their theory. Whether or not this scientific realism is ultimately compatible with a functionalist account of mental states is a difficult issue.[19] The fact that qualitative attitudes (i.e. ordinal preferences) are nevertheless regarded as more fundamental and quantitative attitudes as derivatives can be justified in several ways, with at least some of them depending on the endorsement of a form of scientific realism. First, the informational basis needed to define and measure qualitative attitudes is smaller and/or coarser than for quantitative attitudes. In other words, the amount and precision of information required are smaller as far as qualitative attitudes are concerned, regardless of the nature of this information (behavioral, verbal or introspective). Second, while expectational utility functions are indeed valid representations of preferences relations satisfying some set of axioms, they are not the only ones. This is indeed a key point which has been somewhat muddled in the literature: representation theorems state that a binary relations satisfying the appropriate axioms *can be* represented by a class of expectational utility functions unique up to any positive affine transformation, but they *do not establish that they cannot be represented by other kinds of mathematical structures* (Weymark 2005).[20] In particular, representation theorems do not imply the interval-scale cardinality of the

[17] Behaviorism can also make use of representation theorems. Of course, if preferences and choices are taken to be co-extensive, then the representation theorems provide a way to directly represent choices in terms of utility and probability functions.

[18] There are competing accounts of what is a scientific theory. In a syntactic perspective, a theory refers to a set of sentences or formulae expressed in some language. These formulae are generated by a set of atomic sentences combined by a set of axioms and are closed under implications. In a semantic perspective, a theory rather refers to a class of models, which are itself abstract structures ascribing the truth value "true" to all sentences constitutive of the theory.

[19] See Dennett (1991) for a positive answer. Christensen (2001) and Guala (2017) are more skeptical. As I note in the preceding section, it is not clear what would be the real difference between behaviorism and the preferences-as-choice-dispositions brand of pragmatism (weak behaviorism) if functionalism and scientific realism are noncompatible. Also, it is hard to make sense of the preferences-as-judgments brand of pragmatism (weak mentalism) in an instrumentalist perspective.

[20] This point is crucial in the debates over the axiomatic defense of utilitarianism proposed by John Harsanyi (1955) on the basis of expected utility theory. Harsanyi contended that one's endorsement of the axioms of expected utility theory both for individual and social preferences plus the satisfaction of a Pareto condition commits one to

utility function $u(.)$. A supplementary assumption stating that to prefer α over β *is to* ascribe a higher expected utility to the former than to the latter is needed. This is in this spirit that John Broome (1991) defends what he calls "Bernoulli's hypothesis". The latter is stated in terms of goodness and betterness rather in terms of utilities and preferences and indicates that an act α is better than β to any person if and only if its "expectational goodness" is higher. Thus Bernoulli's hypothesis entails risk-neutrality regarding goodness. What is the justification for Bernoulli's hypothesis? To argue that this a possible implication of representation theorems is to beg the question since it is precisely what has to be established. Broome (2008: 231) offers an argument related to the natural salience of the expectational measure: "there is something to be said for the expectational concept as opposed to these others. The use of probabilities provides a natural analogue of a pair of scales for measuring the strength – analogous to weight – of preferences… The rival concepts are less natural… The expectational concept of degree is the most natural, but it not forced on us by preference alone. Preferences by themselves do not determine a concept of degree. The expectational concept is derived from preferences together with an idea of naturalness". Broome's argument encapsulates what I take to be the essence of pragmatism: qualitative attitudes like preferences are more fundamental and do not force us to adopt any specific view regarding the nature of quantitative attitudes. Still, it is natural to capture the latter in terms of the expectational concept of degree and this provides a strong reason to see utilities and probabilities as real-though-derived entities.

The same is true regarding probabilism and the status of the probability function $p(.)$. It is well-established that representation theorems establish probabilism only in a weak sense: while there is a (unique) probability function that represent the agent's credences over events, the same qualitative relation can be represented by other functions conflicting with probability axioms such as additivity (e.g. Zynda 2000). Realism about beliefs-as-probabilities necessitates additional assumptions according to which, if one's credences can be represented by a probability function, then one's credences *really are* probabilities. This is not the place here to settle this quite complicated issue. What matters however is that representation theorems are a key part of the pragmatist understanding of the Bayesian synthesis, alongside with a functionalist account of mental states and possibly a form of scientific realism.

5. The Implausibility of Pragmatism in Bayesian Game Theory

Given the pervasiveness of pragmatism in decision theory and economics, one can also expect to be the dominant view in Bayesian game theory. Bayesian game theory is simply defined as the extension of Bayesian decision theory (as captured by the Bayesian synthesis) to the study of strategic interactions. It is closely related to the so-called epistemic program whose main purpose is to characterize the meaning and the implications of various degrees of mutual knowledge/belief of rationality in games. In this endeavor, Bayesian game theory makes a heavy use of mental attitudes such as beliefs and preferences, including attitudes about attitudes (e.g. one's beliefs about others' beliefs and preferences). It is useful here to contrast the approach retained by Bayesian game theory with the classical game theory's one. The latter is fundamentally "top-down". It proceeds on the basis of the adoption of a given solution concept

weighted utilitarianism. But this is true if and only if we restrict the class of the corresponding utility functions to the expectational ones. Representation theorems *do not* license this kind of restrictions. See Weymark (2005) for considerations on this point on the basis of measurement theory.

such as Nash equilibrium or subgame perfection. For a given game (defined as a set of two or more players, each endowed with a set of pure strategies and a utility function whose domain is the set of strategy profiles), it is then asked which strategy profiles (i.e. a vector of strategies, one per player) satisfy the relevant solution concept. Bayesian game theory proceeds the other around, in a "bottom-up" fashion. Starting with assumptions about the players' knowledge and beliefs about others' rationality and strategy choices, we determine which are the admissible strategy choices for each player and thus the admissible strategy profiles. It is then possible to define a set of sufficient conditions for any given solution concept to be implemented in a game.

Robert Aumann's (1987) well-known characterization of the correlated equilibrium (CE) solution concept in terms of (common knowledge of) Bayesian rationality provides a representative illustration of the methodological strategy of Bayesian game theory.[21] To understand Aumann's result, I need to introduce some further terminology and formalism. Define any game as a triple $G = < N, \{A_i, u_i\}_{i \in N} >$ with N the set of $n \geq 2$ players, A_i the set of player i's strategy and u_i: $S \rightarrow \Re$ player i's utility function mapping any strategy profile $s \in \prod_i A_i$ onto some real number. Each player is assumed to be Bayesian rational and thus the utility functions satisfy the expectational property. We call a *model* of G a semantic structure $M = < W, \boldsymbol{w}, \{I_i, D_i, p_i\}_{i \in N} >$. W denotes the set of states or possible world w, with $\boldsymbol{w}$ the "actual" world. Each possible world is a complete specification of what happens in the game, including the players' strategy choices. I_i defines a partition of the state space W for player i. Each cell $I_i(w)$ in the partition specifies player i's knowledge (defined as probability 1 true beliefs).[22] In particular, for any event (i.e. set of possible worlds) E, i knows that E at w if and only if $I_i(w) \subseteq E$. D_i: $W \rightarrow A_i$ is player i's decision function indicating i's strategy choice in each possible world. Finally, while knowledge and full beliefs are defined by the cells of the partitions I_i, partial beliefs are generated by the probability functions p_i: $2^W \rightarrow [0, 1]$. Player i's beliefs at any possible world w is simply obtained by applying Bayes's rule, i.e. i's belief that E at w is given by

$$(2) \qquad p_{i,w}(E) = \frac{p_i(E \cap I_i(w))}{p_i(I_i(w))}$$

In his article, Aumann establishes two related results: 1) in any game G, for each CE there exists a semantic model M where (a) the players are Bayesian rational in all $w \in W$ and (b) share a common prior $\boldsymbol{p} = p_1 = \ldots = p_n$; 2) if the semantic model M of G satisfies (a) and (b), the players implements a CE in G. Aumann (1987) claims that this result shows that Bayesian rationality entails an equilibrium play (though weaker than Nash equilibrium).[23] To understand the significance of this claim, one should contrast it with previous ones arguing that Bayesian rationality has no theoretical implications regarding the way games are (or should be) played (e.g. Kadane and Larkey 1982).

Since Aumann's paper, the literature on Bayesian decision theory has kept growing at a steady pace. Its conceptual framework and formal tools is nowadays not only used to deal with formal

[21] See (Brandenburger 2014) for several other examples. Stalnaker (1994) provides what I consider to be the clearest and most lucid statement of the general approach adopted by Bayesian game theory. Basically, the objective is to characterize each solution concept in terms of a *class* of semantic epistemic models of games, i.e. structures representing what the players' know and believe about others' rationality and behavior.

[22] Since I_i is a partition, for any $w, w' \in W$, we either have $I_i(w) = I_i(w')$ or $I_i(w) \cap I_i(w') = \varnothing$.

[23] This is not the place to evaluate the relevance of this claim. For contrasting views, see (Gintis 2009) and (Levi 1997).

issues strictly related to game theory, but also to advance on other topics in fields like social ontology and the economics of institutions [e.g. (Gintis 2009); (Guala 2016); (Hédoin 2017)]. As the brief formal overview proposed above indicates, Bayesian game theory gives a central place to mental states, especially beliefs. The conditions characterizing the Bayesian synthesis (Probabilism, Framing and Rationality) are generally taken for granted by game theorists working in this perspective. Though there is scarcely any explicit statement about this point, it seems that the notions of preferences and beliefs are mostly understood in a pragmatist fashion. In particular, the players' strategy choices are thought to be essentially equivalent to Savage's acts in one-person decision problems.[24] The identification of utilities and probabilities is then assumed to proceed along the standard lines of representation theorems, with a conceptual priority given to practical reason (preferences) over theoretical reason (beliefs).

Compared with Bayesian decision theory, it is important to note however that Bayesian game theory has to deal with far more complex structures capturing nested qualitative and quantitative attitudes.[25] Indeed, as each state specifies what the players are doing (through the decision functions $D_i(.)$) and knowing/believing (through the partitions $I_i(.)$ and the probabilities $p_i(.)$) about everything that is relevant, the players entertain beliefs about others' strategy choices but also about others' beliefs about others' strategy choices and so on. These *belief hierarchies* are indeed an essential conceptual object whose study is at the core of the whole epistemic program in game theory (Perea 2012). The importance of belief hierarchies leads to two related issues in the context of the extension of Bayesian decision theory to the study of strategic interactions: first, is it still appropriate to conceive strategy choices as acts in spite of the fact that the players' probabilistic assessments are made over events regarding players' choices and beliefs? Second, what is the status of mental attitudes (beliefs and preferences) that enter as propositional content into the players' belief hierarchies?

The first issue is a foundational one. I have argued in the preceding section that the meaningfulness of pragmatism depends on the availability of the representation theorems found in decision theory. Game theorists mostly implicitly assume that the Bayesian concept of acts, and thus the Framing condition, applies in a game-theoretic context (e.g. Myerson 1991). This assumption is motivated on the ground that there is no significant difference between the kinds of events the Bayesian decision-maker has to deal with in one-person decision problem and in games. Obviously, there is indeed a difference: in one-person decision problems, events are assumed to be exogenous to the decision-maker's choices; in games, the decision-maker has to ascribe probabilities to events that involve other players' choices, as well as their beliefs about everyone else's choices. In short, events range over possible belief hierarchies.[26] Bayesian game

[24] I discuss the possibility of viewing preferences and beliefs in game theory in a mentalistic perspective in the next section. Some game theorists have suggested that the players' preferences represented by the utility functions should be understood in a revealed-preference/behaviorist sense. See especially Binmore (1994). This position has few adherents however. The most significant difficulty is the same as the one faced by pragmatism: as representation theorems are unavailable, there is no way to identify (revealed) preferences independently of beliefs (Hausman 2011). Since the converse is also true, a vicious circularity is lurking. Moreover, by completely downplaying the role played by mental attitudes, behaviorism is unable to account for the way each player reasons *about* others' behavior and mental attitudes. For a more detailed criticism on behaviorism in game theory, see (Lehtinen 2011).

[25] It may be added that a full account of the players' reasoning process, especially in sequential games, should also capture *conditional attitudes* over events, including null events (i.e. events one initially ascribes probability 0). See (Stalnaker 1996).

[26] Note that this point is not the same than the one about the stochastic independence between acts and states. Indeed, in the Bayesian synthesis, stochastic independence must be assumed if and only if there is *causal*

theorists simply assume (rather than argue) that this difference is not relevant. Representation theorems are thus also assumed to hold in the study of strategic interactions. This is not so straightforward however. First, it should be noted that in semantic epistemic models of games of the kind introduced above, possible worlds include the strategy choice of *each* player, including the decision maker's. It follows that one's *own* strategy choice is itself an event on which one formally ascribes a probability. This is essentially a formal convenience: to excludes one's strategy choice from the description of a possible world would imply that each player faces a different state space. This can be regarded as a problem however as it has been argued that ascribing probabilities to one's acts is nonsensical in a Bayesian perspective (Levi 1997).[27] A second, more fundamental problem, is highlighted by Mariotti (1995) who shows that Savage's axioms of rationality (actually only a complete ordering axiom and a dominance axiom) are inconsistent with a set of plausible rationality principles in games:[28]

> **G1:** The elimination of a dominated strategy should not change the ordering of the remaining strategies.
>
> **G2:** In a perfect-information extensive-form game, strategies which are not part of subgame perfect equilibrium cannot be dominant.
>
> **G3:** If the consequence of some strategy profile (i.e. the conjunction of an act and an event regarding others' choices in the perspective of the decision-maker) in a game *G* is another game *G'*, any strategy that is dominant in *G* should not be dominated in the resulting extensive-form game *G''*.
>
> **G4:** If the strategy profile $\boldsymbol{s}$ is a non-Pareto-dominated strict Nash equilibrium in *G'* and offers player i a gain of $u_i(\boldsymbol{s})$, then it should be possible for i to strictly prefer any sure consequence $u_i(c) > u_i(\boldsymbol{s})$ in *G* than playing *G'*.

All four rationality requirements are arguably reasonable. However, Mariotti (1995) establishes an impossibility result showing that Savage's rationality axioms and principles **G1**-**G4** may be inconsistent, i.e. there are games where no preference ordering satisfies them all. Beyond this formal result, Mariotti points out the conceptual difficulties emerging when strategies are conceived as acts, at least in Savage's framework. Savage's representation theorem indeed builds on a "structural" axiom according to which *all* acts, including constant acts (i.e. acts leading to a sure consequence), be available and ordered. More precisely, what Savage was assuming is that for all consequences C_i there is an act α_{C_i} such that $\alpha_{C_i}(S_j) = C_i$ for all S_j in the partition. The set of acts is then closed under mixing of constant acts.[29] Even in one-person decision problems, this structural axiom is regarded as problematic, especially when combined with an axiom of completeness of preferences on acts. It makes however even less sense in the context of games where it is hard to see why a player should rank all his strategy choices, including those that are not in fact available. In particular, which strategies/acts are actually available in a game for a player must matter *to all players* due to the possibility of iterative

independence between acts and states. It is arguable that such independence should also be assumed in (normal-form) games. Still, in games, causal independence does not imply stochastic independence, as Aumann's account of CE illustrates. See (Stalnaker 1996).

[27] This claim is not consensual though. For a critique, see (Rabinowicz 2002).

[28] Mariotti's argument is developed in the context of Savage's version of the Bayesian synthesis. However, it seems to apply to all of its versions extended to strategic interactions.

[29] This "richness axiom" is generally not stated as one of Savage's seven "official" axioms. But it is actually required for the other axioms to do the work in the proof of Savage's representation theorem. See Joyce (1999).

elimination procedures of strategies. It then seems that the derivation of the players' utilities and probabilities through a Bayesian representation theorem should be made on the basis of some "universal game" played by all players and where all constant acts are available. This is clearly untenable. Alternatively, one may assume the existence of "hypothetical acts" outside the games but on which the players could take side-bets (Mariotti 1997). An obvious consequence is that the game played is no longer the same. If hypothetical acts belong to players' private knowledge, the game structure is no longer common knowledge. Arguably, this conceptual difficulty can be regarded as peculiar to Savage's version of Bayesian decision theory and there are indeed ways to define acts otherwise than functions from states to consequences. There are also representation theorems in decision theory that build on weaker structural axioms regarding the richness of the domain of alternatives. But this takes us away from the Bayesian synthesis and these are not those that are implicitly assumed in Bayesian game theory.[30]

It is worth noting that the difficulty to directly transpose the representation theorems of Bayesian decision theorem to strategic interactions affects both the issues of the *measure* and of the *(rational) properties* of utilities and probabilities. Regarding the former, we have seen above that one function of representation theorems is to provide a method to measure quantitative attitudes. Probabilities are defined as betting rates (which themselves can be seen as behavioral dispositions to buy and sell bets) and utilities as preferences between sure and mixed consequences. However, it seems that it is not possible to measure quantitative attitudes in games this way. Regarding the latter, the unavailability of representation theorems makes impossible to ground the assumed properties of quantitative attitudes on local and more easily evaluable axioms. If this argument holds, pragmatism is no longer a viable option for Bayesian game theorists.

6. Belief Hierarchies in Bayesian Game Theory

I focus in this section on the existence of belief hierarchies in games and the way pragmatism and mentalism can account for them. I shall argue that both pragmatism and mentalism can deal with this object but in quite different ways. In particular, I conclude that only mentalism seems able to account for the mechanisms generating belief hierarchies.

Consider first how pragmatism can deal with belief hierarchies. Without representation theorems, it is not clear how pragmatists (endorsing a functionalist view of mental states) can derive the players' probabilistic beliefs. Since belief hierarchies are nested iterative chains of (probabilistic) beliefs over (probabilistic) beliefs, the argument of the unavailability of representation theorems for decision-making in strategic interactions carries over to them. However, *even under the supposition* that the Bayesian game theorists could rely on a representation theorem to define and measure beliefs, it appears that another difficulty may surface. Indeed, belief hierarchies are far more complex objects than simple probabilistic

[30] A recent paper by Aumann and Dreze (2009) seems to provide a partial solution to Mariotti's impossibility result. Relying on the preference concept as the primitive (2009: 3), they show how to derive a class of expectational utility functions representing the preference ordering on the basis of the available acts solely. But Aumann and Dreze's account does not solve the problem discussed in the text completely for several reasons: i) their theorem does not establish the uniqueness of the probability measure (though all admissible probability functions yield the same expected utility), ii) the theorem provides a method to elicit the players' beliefs about others' choices *but not the belief hierarchies* (2009: 11).

beliefs. This is not an objection per se of course, but the point is that even if we have the formal and axiomatic apparatus to represent the players' beliefs over states of nature, that does not imply that this formal apparatus is sufficient to derive players' belief hierarchies.[31] To clarify this issue, take the simplest case of strategic interactions, i.e. a two-player-two-strategy normal form game. The two players are respectively P1 and P2 and I denote the two pure strategies as A1 and B1 and A2 and B2 for P1 and P2 respectively. In a pragmatist perspective, P2's beliefs over the states of nature (i.e. P1's strategy choice) are derived from P2's preferences over the pair of acts/strategies {A2, B2}. The same is true for P1. A Savage's like representation theorem, if available, should make possible to determine P1's and P2's beliefs over the other player's strategy choice. Now, consider how we can derive P1's beliefs *over P2's beliefs over P1's choice over* {A1, B1}. We can imagine the following algorithmic procedure. First, refine P1's state space in the following way: a state of nature is now a pair of P2's strategy choice (A2 or B2) *along with* some set of P2's beliefs over P1's strategy choices, which we can simply denote p_2. Then, P1's state space over which she forms beliefs is {(S2, p_2)} with S2 = A2, B2 and p_2 any probabilistic distribution over {A1, B1}. Do the same thing for P2 and use Savage's like representation theorem to derive both players' beliefs over these refined state spaces. Denote p_1^2 and p_2^2 the resulting second-order beliefs (i.e. beliefs over beliefs) for P1 and P2 respectively. Second, refine again P1's state space, this time by forming triples made of a strategy choice, first-order beliefs and second-order beliefs. The resulting state space is the set {(S2, p_2, p_2^2)}. Do the same for P2 and use a Savage's like representation theorem to derive the players third-order beliefs and denote p_1^3 and p_2^3 the corresponding measures. Finally, repeat the procedure for any arbitrary *k* number of steps to derive the players' *k*+1-order beliefs p_1^{k+1} and p_2^{k+1}. We can derive this way belief hierarchies of any arbitrary length.

It is now crucial to acknowledge on which basis these belief hierarchies have been derived. To start the derivation for P1, two things are needed: first, P1's preferences over the pair {A1, B1}; second, P2's beliefs over the same pair {A1, B1}. Of course, the latter are themselves obtain on the basis of P2's preferences and P1's beliefs over {A2, B2}. There are only two possible ways to interpret the basis on which the belief hierarchies are generated, and only one of them is plausible under a functionalist view of mental states. The first possibility is to consider that belief hierarchies are indeed generated through the above algorithm. But this leads to the troublesome implication that P1's beliefs over P2's beliefs are entirely determined by the *conjunction* of P1's and P2's preferences over their respective sets of strategies. This is deeply counterintuitive, especially if one endorses a functionalist account of beliefs: functionalism holds that one's beliefs are determined by their constitutive or causal role in one's action or preference judgments, *but not* by their constitutive or causal role in the *conjunction of everyone's action or preference judgment*. The second possibility is to assume that the "states of nature" over which players have beliefs *do not* correspond to others' strategy choices as it is generally informally stated in game theory textbooks. Rather, the players' beliefs should range over others' *preferences* over acts. Since other players' preferences over acts can be represented by a pair of utility and probability functions, this implies then that each player can be represented has having beliefs over others' beliefs, leading to the generation of the belief hierarchies. This seems perfectly compatible with a functionalist view of mental states, especially as a functionalist does not need to assume that agents are "really" able to represent

[31] One should be cautious to avoid confusion between *states of the world* and *states of nature*. The former are constitutive of semantic models of games and include *all* players' strategy choices. The latter are Savage's probabilistically act-independent objects over which the decision-maker has beliefs.

others' mental states, including their preferences. But it is worth insisting that this is not how the extension of Bayesian decision theory to games is generally informally interpreted.

The importance of belief hierarchies in Bayesian game theory emphasizes another issue related to the mechanism through which these hierarchies are generated. This is of course deeply related to the status of the mental states that are constitutive of them. What behaviorism, pragmatism and mentalism all offer is a way to interpret a set of foundational concepts (including concepts of mental states) in one-person decision problems. It is important to note that in each case, the proposed interpretation is addressed *to the decision theorist*, i.e. there is absolutely no assumption about how the decision-maker herself understands these concepts. This is not a problem, at least if one does not aim at understanding how the decision-maker is "really" framing the decision problem. But matters are quite different in games because the players' mental attitudes must have as their propositional content others' mental attitudes. Whatever the game theorist's own position regarding the nature, properties and measurement of these attitudes is, she has to account for the way each player conceives others' mental attitudes.

Consider the following extreme example. Game-theorist-Bob is a hardcore revealed-preference theorist who not only takes choices to be the primitives in decision and game theory, but also consider that there is nothing more than choices in a scientific perspective, i.e. mental attitudes like beliefs or desires are mere illusions. On this basis, he may rightfully argue that preferences should be defined as observed or hypothetical choices and that utility functions are just useful mathematical devices to represent choices. Suppose however that game-player-Ann (whose behavior is studied by game-theorist-Bob), to predict the behavior of game-player-Chris, adopts the following simulation reasoning:[32] "if I were Chris in the current situation, I would believe that Ann would do *x*, therefore, given what I know about Chris's preferences and rationality, I expect him to do *y*". This kind of simulation thinking seems to be perfectly plausible and widely used in strategic interactions. Lewis (1969: 27) for instance suggested that coordination is achieved by forming concordant expectations that we acquire "by putting ourselves in the other fellow's shoes, to the best of our ability". This ability precisely builds on an assumption of "symmetric reasoning", i.e. we routinely assume that everyone else is reasoning like us and that whatever we may infer from a given situation, others will infer too (and thus that they will also assume that everyone else makes the same inference). People may assume symmetric reasoning for all sorts of reasons[33] but what matter here is that the kind of simulation thinking just described builds on our ability to ascribe mental attitudes to other persons. Now, in spite of being a hardcore revealed-preference theorist, game-theorist-Bob would probably be wrong to ignore this ability to explain and predict Ann's and Chris's coordination.

My point is that this example illustrates how a mentalistic approach could account for the generation of the belief hierarchies in a quite different way than pragmatism. In essence, pragmatism deals with belief hierarchies in a fully static way: a given preferences pattern in a game can be represented by a profile of belief hierarchies (alongside with a profile of utility functions) but we are not given any indication regarding how the players have obtained these beliefs. Arguably, this is no different than in Bayesian decision theory, where the factors responsible for the decision-maker having specific beliefs are generally not investigated. How

[32] On the importance of simulation thinking, see for instance (Morton 2005) and (Guala 2016: chap. 7).

[33] For instance, symmetric reasoning may be grounded on community-membership (Hédoin 2016a).

beliefs are formed is however an important issue in strategic interactions because one of the goals of epistemic game theory is to study how rational agents reach coordination. In other words, we want to know how rational players are able to form *convergent* belief hierarchies. Lewis's notion of symmetric reasoning corresponds to one of the mechanisms leading to this convergence. It is not clear whether such a notion is meaningful under a functionalist view of mental states. A player engaging in symmetric or simulation reasoning is more likely to start from her own desires and beliefs and then to proceed by assuming that in the same circumstances others will have the same mental states. This is at least the dominant account of meta-cognition as developed in the so-called "theory of mind" field (Morton 2005). On this basis, it can be easily shown that a player having some belief *b* and who is assuming symmetric reasoning from others will also have belief *b'* that everyone has belief *b*, belief *b''* that everyone has belief *b'* that everyone has belief *b*, and so on. Symmetric reasoning (if shared in the population) thus leads to the generation not only of belief hierarchies, but of *convergent* belief hierarchies. Mentalism thus seems particularly appealing to explain equilibrium play in games.[34]

An obvious objection is that under a mentalistic account, belief hierarchies are far too cognitively demanding mental states that human agent cannot reasonably possess. Indeed, it seems rare if not impossible for someone to entertain more than third-order beliefs in normal circumstances. This standard objection made against the usual notions of "common belief" and "common knowledge" in game theory is irrelevant in a pragmatist perspective because the latter does not require the assumption that the belief hierarchies are "real" in the sense of being intrinsic to the players. A way however to preserve the mentalistic interpretation is to supplement the Bayesian apparatus with so-called "awareness structures", as suggested by Sillari (2005) in his discussion of Lewis's theory of convention. In this case, a player "really" has belief *b* if and only if *b* is permissible given a set of assumptions (i.e. *b* conforms to Bayesian axioms, can be generated from a symmetric reasoning assumption, …) *and* the player is aware of the state/object the belief refers to. The notion of awareness here refers to some kind of mental state that can also be interpreted along mentalistic lines.

Where does this leave us? Given the difficulties surveyed in the preceding section, one may argue that the prospects of Bayesian game theory are doomed. The solution would then be simply to recover the program of classical game theory, as I have characterized it above. This is indeed the solution Mariotti (1995: 1108) entertains in light of his impossibility result:

> "a divorce is required between game theory and individual decision theory. Too often have game theorists felt the need to seek a 'decision theoretic justification' for their solution concepts. We submit that strategic decision principles may be radically different from individual decision theoretic principles".

Of course, if accepted such a conclusion would put the epistemic program in game theory to a halt since its main aim is to give "decision theoretic justifications" to solution concepts. This "back to Nash" strategy is however unattractive considering that the epistemic program is also committed to the study of the way players reason in strategic interactions. A conceptual apparatus and theoretical principles to account for the way players form beliefs and preferences

[34] Under some plausible assumptions, the standard common prior assumption and a hypothesis of symmetric reasoning may indeed be seen as formally equivalent (Hédoin 2016b).

are thus needed. An empirical content is also needed in this perspective, and it could be provided by the field of the so-called "theory of mind", as suggested by Brandenburger (2014: xxii). This leads to the second solution which would thus be to keep up at least with this aspect of the epistemic program but to take a mentalistic orientation. In the mentalistic view, the unavailability of representation theorems in games is not a problem. Okasha (2016) suggests that mentalism is hard to defend from the normative decision-theoretic perspective but this is not really relevant as the epistemic program pursue an essentially positive endeavor. In this case, utilities and probabilities would simply constitute the starting point for characterizing the players' mental attitudes. Utilities could then simply refer to "material payoffs" and beliefs elicited through non-pragmatist-non-behaviorist means. The standard assumption of expected utility maximization in games could then be defended on the basis of a method similar to Peterson's (2002) in one-person decision problem. However, if this approach is taken, it must be acknowledged that the relevance of the Bayesian apparatus might be disputed in a positive perspective. For instance, the players' beliefs may simply not correspond to a quantitative concept and even less to a probabilistic one. In particular, the players' ability to infer what others will do on the basis of symmetric reasoning and simulation thinking may be seen either as an empirically grounded substitute or complement to Bayesian reasoning.

7. Conclusion

The Bayesian synthesis in decision theory can be given three different interpretations depending on which foundational notions are regarded as the primitives: choices for behaviorism, preferences for pragmatism and utilities/probabilities pairs for mentalism. While all three interpretations are plausible, I have argued that pragmatism is the dominant one among decision theorists and game theorists. It states that choices and preferences are conceptually separated, that qualitative attitudes are more fundamental than quantitative attitudes and that practical reason has priority over theoretical reason. I have suggested that the plausibility of the pragmatist view entirely relies on the existence of representation theorems that characterize utilities and probabilities as representations of preferences. Bayesian game theorists working in the context of the epistemic program in game theory basically take pragmatism for granted. However, once it is recognized that representation theorems do not carry over to games and strategic interactions, this endorsement of pragmatism rests on shaky grounds. I have suggested that the prospects of Bayesian game theory in a mentalistic perspective look more promising.

References

Anscombe, F. J., and R. J. Aumann. 1963. "A Definition of Subjective Probability." *The Annals of Mathematical Statistics* 34 (1): 199–205.

Aumann, R. J., and J. H. Dreze. 2009. "Assessing Strategic Risk." *American Economic Journal: Microeconomics* 1 (1): 1–16.

Aumann, Robert J. 1987. "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica* 55 (1): 1–18.

Binmore, K. G. 1994. *Game Theory and the Social Contract: Playing Fair*. MIT Press.

———. 2009. *Rational Decisions*. Princeton University Press.

Bradley, Richard. 2004. “Ramsey’s Representation Theorem.” *Dialectica* 58 (4): 483–97.

———. 2016. *Decision Theory with a Human Face*. Mimeo, London School of Economics.

Brandenburger, Adam. 2014. *The Language of Game Theory: Putting Epistemics Into the Mathematics of Games*. World Scientific.

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Wiley.

———. “Can There Be a Preference-Based Utilitarianism?” In M. Fleurbaey, M. Salles and J. A. Weymark (eds.), *Justice, Political Liberalism, and Utilitarianism. Themes from Harsanyi and Rawls*. Cambridge: Cambridge University Press, 221-238.

Caplin, Andrew, and Andrew Schotter. 2010. *The Foundations of Positive and Normative Economics: A Handbook*. Oxford University Press.

Christensen, David. 2001. “Preference-Based Arguments for Probabilism.” *Philosophy of Science* 68 (3): 356–76.

Cozic, Mikaël, and Brian Hill. 2015. “Representation Theorems and the Semantics of Decision-Theoretic Concepts.” *Journal of Economic Methodology* 22 (3): 292–311.

Davidson, Donald. 2004. *Problems of Rationality*. Clarendon Press.

Dennett, Daniel. 1989. *The Intentional Stance*. MIT Press.

———. 1991. “Real Patterns.” *The Journal of Philosophy* 88 (1): 27–51.

Dietrich, Franz, and Christian List. 2016. “Mentalism Versus Behaviourism in Economics: A Philosophy-of-Science Perspective.” *Economics and Philosophy* 32 (2): 249–81.

Elliott, Edward. 2017. “Probabilism, Representation Theorems, and Whether Deliberation Crowds Out Prediction.” *Erkenntnis* 82 (2): 379–99.

Fumagalli, Roberto. 2013. “The Futile Search for True Utility.” *Economics &amp; Philosophy* 29 (3): 325–47.

Gilboa, Itzhak, and David Schmeidler. 1995. “Case-Based Decision Theory.” *The Quarterly Journal of Economics* 110 (3): 605–39.

Gintis, Herbert. 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton University Press.

Glimcher, Paul W. 2010. *Foundations of Neuroeconomic Analysis*. 1 edition. New York: Oxford University Press.

Guala, Francesco. 2016. *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton University Press.

———. 2017. “Preferences: Neither Behavioural nor Mental.” Departmental Working Paper. Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano. http://econpapers.repec.org/paper/milwpdepa/2017-05.htm.

Gul, Faruk B. and Wolfgang Pesendorfer. 2008. “The Case for Mindless Economics.” In A. Caplin & A. Schotter (eds.), *The Foundations of Positive and Normative Economics*, Oxford University Press, 3-39.

Hacking, Ian. 2001. *An Introduction to Probability and Inductive Logic*. Cambridge University Press.

Hammond, Peter J. 1988. “Consequentialist Foundations for Expected Utility.” *Theory and Decision* 25 (1): 25–78.

Hands, D. Wade. 2013. “Foundations of Contemporary Revealed Preference Theory.” *Erkenntnis* 78 (5): 1081–1108.

Harsanyi, John C. 1955. “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility.” *Journal of Political Economy* 63 (4): 309–21.

Hausman, Daniel M. 2011. *Preferences, Value, Choice, and Welfare*. Cambridge University Press.

Hédoin Cyril. 2016a. “Community-Based Reasoning in Games: Salience, Rule-Following, and Counterfactuals.” *Games* 7 (4): 36.

———. 2016b. “Bayesianism and the Common Prior Assumption in Game Theory.” *Working Paper REGARDS n° 7-2016*, University of Reims Champagne-Ardenne.

———. 2017. “Institutions, Rule-Following and Game Theory.” *Economics and Philosophy* 33 (1): 43-72.

Jeffrey, Richard C. 1974. “Preference Among Preferences.” *The Journal of Philosophy* 71 (13): 377–91.

———. 1990. *The Logic of Decision*. University of Chicago Press.

Joyce, James M. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kadane, Joseph B., and Patrick D. Larkey. 1982. “Subjective Probability and the Theory of Games.” *Management Science* 28 (2): 113–20.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. “Back to Bentham? Explorations of Experienced Utility.” *The Quarterly Journal of Economics* 112 (2): 375–406.

Lehtinen, Aki. 2011. “The Revealed-Preference Interpretation of Payoffs in Game Theory.” *Homo Oeconomicus* 28 (3): 265–96.

Levi, Isaac. 1997. *The Covenant of Reason: Rationality and the Commitments of Thought*. Cambridge University Press.

Lewis, David. 1969. *Convention: A Philosophical Study*. John Wiley and Sons.

———. 1981. “Causal Decision Theory.” *Australasian Journal of Philosophy* 59 (1): 5–30.

Mariotti, Marco. 1995. “Is Bayesian Rationality Compatible with Strategic Rationality?” *The Economic Journal* 105 (432): 1099–1109.

———. 1997. “Decisions in Games: Why There Should Be a Special Exemption from Bayesian Rationality.” *Journal of Economic Methodology* 4 (1): 43–60.

Meacham, Christopher J. G., and Jonathan Weisberg. 2011. “Representation Theorems and the Foundations of Decision Theory.” *Australasian Journal of Philosophy* 89 (4): 641–63.

Morton, Adam. 2005. *The Importance of Being Understood: Folk Psychology as Ethics*. Routledge.

Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict*. Cambridge, Mass.: Harvard University Press.

Okasha, Samir. 2016. “On The Interpretation of Decision Theory.” *Economics and Philosophy* 32 (3): 409–33.

Perea, Andrés. 2012. *Epistemic Game Theory: Reasoning and Choice*. New York: Cambridge University Press.

Peterson, Martin. 2002. "An Argument for the Principle of Maximizing Expected Utility." *Theoria* 68 (2): 112–28.

———. 2008. *Non-Bayesian Decision Theory: Beliefs and Desires as Reasons for Action*. Springer Science & Business Media.

Rabinowicz, Wlodek. 2002. "Does Practical Deliberation Crowd Out Self-Prediction?" *Erkenntnis* 57 (1): 91–122.

Ramsey, Franck. 1926. "Truth and Probability". In D.H. Mellor (ed.), *Philosophical Papers*, Cambridge: Cambridge University Press, 1990.

Ross, Don. 2014. *Philosophy of Economics*. Palgrave Macmillan.

Samuelson, P. A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Economica* 5 (17): 61–71.

Savage, Leonard J. 1954. *The Foundation of Statistics*. Courier Dover Publications.

Searle, John R. 2003. *Rationality in Action*. MIT Press.

Sen, Amartya. 1973. "Behaviour and the Concept of Preference." *Economica* 40 (159): 241–59.

Skyrms, Brian. 1982. "Causal Decision Theory." *The Journal of Philosophy* 79 (11): 695–711.

Stalnaker, Robert. 1994. "On the Evaluation of Solution Concepts." *Theory and Decision* 37 (1): 49–73.

———. 1996. "Knowledge, Belief and Counterfactual Reasoning in Games." *Economics and Philosophy* 12 (02): 133–63.

Sillari, Giacomo. 2005. "A Logical Framework for Convention." *Synthese* 147 (2): 379–400.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297–323.

Weymark, John A. 2005. "Measurement Theory and the Foundations of Utilitarianism." *Social Choice and Welfare* 25 (2–3): 527–55.

Zynda, Lyle. 2000. "Representation Theorems and Realism about Degrees of Belief." *Philosophy of Science* 67 (1): 45–69.