INTRODUCTION TO THE SEMANTICS OF BELIEF AND COMMON BELIEF

GIACOMO BONANNO
KLAUS NEHRING

Working Paper Series #97-19

Department of Economics
University of California
Davis, California 95616-8578

Introduction to the Semantics of Belief and Common Belief

Giacomo Bonanno and Klaus Nehring

Working Paper Series No. 97-19
June, 1997

/

Paper prepared for an invited lecture at the Workshop on Bounded Rationality and Economic Modeling, Firenze, July 1 and 2, 1997 (Interuniversity Centre for Game Theory and Applications and Centre for Ethics, Law and Economics)

Abstract

We provide an introduction to interactive belief systems from a qualitative and semantic point of view. Properties of belief hierarchies are formulated locally. Among the properties considered are "Common belief in no error" (which has been shown to have important game theoretic applications), "Negative introspection of common belief" (which plays a role in the epistemic foundations of correlated equilibrium), "Truth of common belief" and "Truth about common belief". The relationship between these properties is studied.

1. Introduction

The structures that are most often used in the economics and computer science literature to discuss interactive beliefs/knowledge are partition structures[1]. Partition structures embody the S5 logic for individual beliefs, in particular the Truth Axiom, that is, the assumption that it is a *necessary* truth (true in all possible worlds of the model) that no one has any false beliefs. While at the individual level the Truth Axiom merely establishes an objective requirement of compatibility between the individual's beliefs and the external world, at the intersubjective level it has strong implications:

> "The assumption that Alice believes (with probability one) that Bert believes (with probability one) that the cat ate the canary tells us nothing about what Alice believes about the cat and the canary themselves. But if we assume instead that Alice ***knows*** that Bert knows that the cat ate the canary, it follows, not only that the cat in fact ate the canary, but that Alice knows it, and therefore believes it as well" (Stalnaker, 1996, p. 153).

As Stalnaker points out (1994, 1996) there is an important conceptual difference between a theory that builds S5 into the concept of knowledge (which – Stalnaker argues – is based on equivocating between knowledge and belief) and a theory that describes epistemic conditions under which knowledge and belief coincide, and then considers the consequences of assuming those conditions.

[1] See, for example, Aumann (1976,1987), Geanakoplos (1992), Fagin *et al* (1995).

In this paper we discuss the semantic approach to interactive beliefs and discuss the main issues that arise when properties of beliefs are defined locally, that is, with respect to the true or actual belief hierarchies.

2. Interactive belief frames

DEFINITION 1. A *KD45 frame* for interactive *beliefs* (or frame, for short) is a tuple

$$\mathcal{F} = (N, \Omega, \tau, \{P_i\}_{i\in N})$$

where

- $N = \{1, ..., n\}$ is a finite set of individuals.
- Ω is a finite set of states (or possible worlds). The subsets of Ω are called events.
- $\tau \in 52$ is the "true" or "actual" state.
- for every individual $i \in N$, $P_i : 52 \to 2^{\Omega} \setminus \varnothing$ (where 2^{Q} denotes the set of subsets of Ω) is i's *possibility correspondence* satisfying the following properties (whose interpretation is given in Remark 3 below): $\forall a, (3 \in \Omega$,

Transitivity:	if $\beta \in P_i(\alpha)$ then $P_i(\beta) \subseteq P_i(\alpha)$,
Euclideanness:	if $(3 \in P(\alpha)$ then $P_i(\alpha) \subseteq P_i(\beta)$.

For every $\alpha \in \Omega$, $P_i(\alpha)$ represents the set of states that individual i considers possible at a.

REMARK 1 (Graphical representation). A non-empty-valued and transitive possibility correspondence $P : \Omega \rightarrow 2^{\Omega} \backslash \varnothing$ can be uniquely represented (see Figures 1-5) as an asymmetric directed graph[2] whose vertex set consists of disjoint events (called cells and represented as rounded rectangles) and states, and each arrow goes from, or points to, either a cell or a state that does not belong to a cell. In such a directed graph, $\omega' \in P(\omega)$ if and only if either ω and ω' belong to the same cell or there is an arrow from ω, or the cell containing ω, to ω', or the cell containing ω'. Conversely;-given a transitive directed graph in the above class such that each state either belongs to a cell or has an arrow out of it, there exists a unique non-empty-valued, transitive possibility correspondence which is represented by the directed graph.

The possibility correspondence is euclidean if and only if all arrows connect states to cells and no state is connected by an arrow to more than one cell (for an example of a non-euclidean possibility correspondence see the common possibility correspondence P_* of Figure 3 below).

Finally, if – in addition – the possibility correspondence is reflexive ($\omega \in P(\omega)$, $\forall \omega \in \Omega$), then one obtains a partition model where each state is contained in a cell and there are no arrows between cells.

A directed graph is *asymmetric* if, whenever there is an arrow from vertex v to vertex **v'** then there is no arrow from v' to v.

EXAMPLE 1. Figure 1 represents the following frame using the convention established in Remark 1: $N = \{1, 2\}$, $\Omega = \{\tau, \beta, \gamma\}$, $P_1(\tau) = \{r\}$, $P_1(\beta) = P_1(\gamma) = \{\beta\}$, $P_2(r) = P_2(\gamma) = \{\tau, \gamma\}$, $P_2(\beta) = \{\beta\}$.

Insert Figure 1

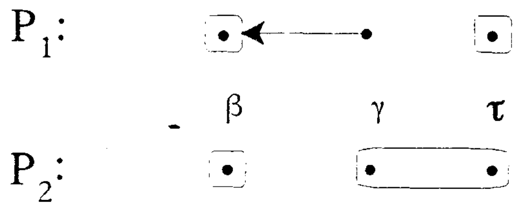

Figure 1

Given a frame and an individual i, i's *belief operator* $B_i : 2^\Omega \to 2^\Omega$ is defined as follows: $\forall E \subseteq \Omega$, $B_i E = \{\omega \in \Omega : P_i(\omega) \subseteq E\}$. $B_i E$ can be interpreted as the event that (i.e. the set of states at which) individual i *believes* that event *E* has occurred.

DEFINITION 2. Beliefs pertain to propositions. Events, that is, subsets of Ω should be thought of as representing propositions. In order to establish the interpretation of events as propositions we need to introduce the notion of a model based on a frame.

- We consider a language with n modal operators $\square_1, \square_2, \ldots, \square$, one for each individual. The intended interpretation of $\square_i \phi$ is "individual i believes that ϕ". The alphabet of the language consists of: (1) a finite or countable set Π of *sentence letters* (representing atomic propositions), (2) the *connectives* $\neg$ (for "not"), $\vee$ (for "or"), and, for every $i \in N$, $\square_i$, (3) the *bracket symbols* (and). The set Φ of formulae is obtained from the sentence letters by closing with respect to negation, disjunction and the operators $\Box_i$.[3] As is customary, we shall often omit the outermost brackets [e.g. we shall write $\phi \vee I)$ instead of $(\phi \vee I))$] and use the following (metalinguistic) abbreviations: $\phi \wedge \psi$ for $\neg(\neg\phi \vee \neg\psi)$ (the symbol $\wedge$ stands for "and"), $\phi \rightarrow \psi$ for $(\neg\phi) \vee \psi$ (the symbol $\rightarrow$ stands for "if... then ...") and $\phi \leftrightarrow \psi$ for $(\phi \rightarrow I)) \wedge (\psi \rightarrow \phi)$ (the symbol $\leftrightarrow$ stands for "if anti only if").

Given a frame $\mathcal{F}$ one obtains a *model* $\mathcal{M}$ *based on it* by adding a function $f : \Pi \rightarrow 2^{\Omega}$ that associates with every sentence letter π the set of states at which π is true. For every formula $\phi \in \Phi$, the *truth set* of ϕ in $\mathcal{M}$, denoted by $\|\phi\|^{\mathcal{M}}$, is defined recursively as follows:

(1) If $\phi = (\pi)$ where π is a sentence letter, then $\|\phi\|^{\mathcal{M}} = f(\pi)$,

(2) $\|\neg\phi\| = \|\phi\|^{\mathcal{M}}$ (with a slight abuse of notation, the symbol '$\neg$' is also used to denote complement: $\neg E = \Omega \setminus E$)

(3) $\|\phi \vee \psi\|^{\mathcal{M}} = \|\phi\|^{\mathcal{M}} \cup \|\psi\|^{\mathcal{M}}$,

(4) For all $i \in N$

$$\|\Box_i\phi\|^{\mathcal{M}} = \{\omega \in \Omega : P_i(\omega) \subseteq \|\phi\|^{\mathcal{M}}\}.$$

[3] Thus Φ is obtained recursively as follows: (i) for every sentence letter π, $(\pi) \in \Phi$, (ii) if $\phi, \psi \in \Phi$ then $(\neg\phi) \in \Phi$, $(\phi \vee \psi) \in \Phi$ and, for every $i \in N$, $(\Box_i\phi) \in \Phi$.

If $\omega \in \|\phi\|^{\mathcal{M}}$ we say that ϕ *is true at state* w *in model* $\mathcal{M}$ (an alternative notation for $\omega \in \|\phi\|^{\mathcal{M}}$ is $\models^{\mathcal{M}}_{\omega} \phi$ and an alternative notation for $w \notin \|\phi\|^{\mathcal{M}}$ is $\not\models^{\mathcal{M}}_{\omega} \phi$). A formula ϕ is *valid in model* $\mathcal{M}$ if and only if $\|\phi\|^{\mathcal{M}} = \Omega$.

Let $\mathcal{F}$ be a frame, $E \subseteq \Omega$ an event and $\mathcal{M}$ a model based on $\mathcal{F}$ where E is the truth set of some formula ϕ, that is, $E = \|\phi\|^{\mathcal{M}}$. Let $B_i : 2^{\Omega} \to 2^{\Omega}$ be the belief operator of individual i (cf. Definition 2). Then B_iE is the truth set of the formula $\Box_i\phi$, that is, $B_iE = \|\Box_i\phi\|^{\mathcal{M}}$. Hence the interpretation of B_iE as the event that individual i believes E (or, more precisely, the proposition represented by event E).

EXAMPLE 2. Consider the frame of Figure 1 and a model based on it where π is an atomic proposition (e.g. "the universe is expanding") which is true at states γ and τ: $\|\pi\| = \{\gamma, \tau\}$. The model is illustrated in Figure 2. Here the truth set of $\Box_1\pi$ is $\{\tau\}$, while the truth set of $\Box_2\pi$ is $\{\gamma, \tau\}$. Thus the truth set of $\Box_1\Box_2\pi$ is $\{\tau\}$. The true state τ describes a world where in fact the universe is expanding and both individuals correctly believe that it is expanding; however, while individual 1 believes that individual 2 believes that the universe is expanding, individual 2 is uncertain as to whether 1 (correctly) believes that it is expanding or 1 incorrectly believes that it is not expanding ($\|\Box_1\neg\pi\| = \{\beta, \gamma\}$) and incorrectly attributes the same belief to individual 2 ($\|\Box_1\Box_2\neg\pi\| = \{\beta, \gamma\}$.

Insert Figure 2

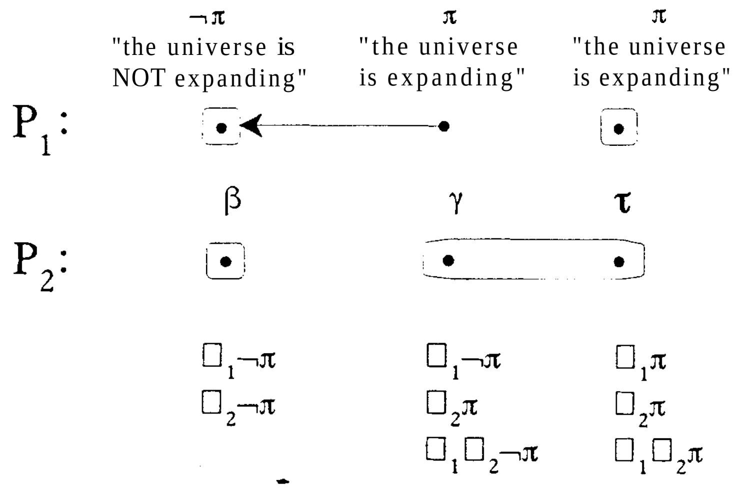

Figure *2*

REMARK 2. Let $P : \Omega \to 2^{\Omega}$ be a possibility correspondence and $B : 2^{\Omega} \to 2^{\Omega}$ the corresponding belief operator (that is, $\forall E \subseteq \Omega$, $BE = \{\omega \in \Omega : I(\omega) \subseteq E\}$). Then B satisfies the following properties: $\forall E, F \subseteq \Omega$,

Necessity: $B\Omega = \Omega$

Conjunction: $B(E \cap F) = BE \cap BF$

Monotonicity: if $E \subseteq F$ then $BE \subseteq BF$

An operator $B : 2^{\Omega} \to 2^{\Omega}$ that satisfies the above properties is called *normal*. Thus the operator that. is obtained from a possibility correspondence is always normal. Instead of taking possibility correspondences as primitives, one could start with a normal belief operator $B_i : 2^{\Omega} \to 2^{\Omega}$ for each individual i and obtain from it i's possibility correspondence as follows: $\forall \alpha \in \Omega$, $P_i(\alpha) =$

$\{\omega \in \Omega : \alpha \in \neg B_i \neg \{\omega\}\}$. The two approaches are equivalent, in the sense the two mappings are one the inverse of the other.[4]

REMARK 3. Fix a frame $\mathcal{F}$. The following is well known (see Chellas, 1984, p. 164):

1. Non-empty valuedness of P_i corresponds to *consistency* of i's beliefs: the following are equivalent.

   (i) $\forall \omega \in \Omega,\ P_i(\omega) \neq \emptyset$,

   (ii) $\forall E \subseteq \Omega,\ B_i E \subseteq \neg B_i \neg E$,

   (iii) for every model $\mathcal{M}$ based on $\mathcal{F}$ and for every formula ϕ, the formula $\Box_i \phi \rightarrow \neg \Box_i \neg \phi$ is valid in $\mathcal{M}$, that is, $\| \Box_i \phi \rightarrow \neg \Box_i \neg \phi \|^{\mathcal{M}} = \Omega$ (individual i cannot simultaneously believe ϕ and not ϕ).

2. Transitivity of P_i corresponds to *posih've introspection* of beliefs: the following are equivalent

   (i) $\forall \alpha, \beta \in \Omega$, if $\beta \in P_i(\alpha)$ then $P_i(\beta) \subseteq P_i(\alpha)$,

   (ii) $\forall E \subseteq \Omega,\ B_i E \subseteq B_i B_i E$,

---

[4] Let $P : \Omega \rightarrow 2^{\Omega}$ be a possibility correspondence, $B : 2^{\Omega} \rightarrow 2^{\Omega}$ the associated belief operator ($\forall E \subseteq \Omega$, $BE = \{\omega \in \Omega : P(\omega) \subseteq E\}$) and $P' : \Omega \rightarrow 2^{\Omega}$ the possibility correspondence obtained from B ($\forall \alpha \in \Omega$, $P'(\alpha) = \{\omega \in \Omega : \alpha \in \neg B \neg \{\omega\}\}$). Then P' = P. Conversely, let B be a belief operator, P the possibility correspondence obtained from B and B' the belief operator obtained from P. Then B = B'.

(iii) for every model $\mathcal{M}$ based on $\mathcal{F}$ and for every formula ϕ, the formula $\Box_i\phi \to \Box_i\Box_i\phi$ is valid in $\mathcal{M}$ (if the individual believes E then she believes that she believes E).

3. Euclideanness of P_i corresponds to *negative introspection* of beliefs: the following are equivalent.

(i) $\forall \alpha,\beta \in \Omega$, if $\beta \in P_i(\alpha)$ then $P_i(\alpha) \subseteq P_i(\beta)$,

(ii) $\forall E \subseteq \Omega$, $\neg B_iE \subseteq B_i\neg B_iE$,

(iii) For every model $\mathcal{M}$ based on $\mathcal{F}$ and for every formula ϕ, the formula $\neg\Box_i\phi \to \Box_i\neg\Box_i\phi$ is valid in $\mathcal{M}$ (if the individual does not believe E, then she believes that she does not believe E).

Notice that we have allowed for false beliefs by not assuming reflexivity of the possibility correspondences [$\forall \omega \in \Omega, \omega \in P_i(\omega)$], which – as is well known (Chellas, 1984, p. 164) – is equivalent to the *Truth Axiom*: $\forall E \subseteq \Omega$, $B_iE \subseteq E$ (if the individual believes E then E is indeed true).

The *common belief operator* B_* is defined as follows. First, for every $E \subseteq \Omega$, let $B_eE = \bigcap_{i \in N} B_iE$, that is, B_eE is the event that everybody believes E. The event that E is commonly believed is defined as the infinite intersection:

$$B_*E = B_eE \cap B_eB_eE \cap B_eB_eB_eE \cap \ldots$$

The corresponding *common possibility correspondence* $P_*: \Omega \to 2^{\Omega}\setminus\emptyset$ is given by: for every $\alpha \in \Omega$, $P_*(\alpha) = \{\omega \in \Omega : \alpha \in \neg B_*\neg\{\omega\}\}$. It is well known[5] that P_* can be characterized as the *transitive closure* of $\bigcup_{i \in N} P_i$, that is,

$\forall \alpha, \beta \in \Omega$, $\beta \in P_*(\alpha)$ if and only if there is a sequence $\langle i_1, \ldots i_m \rangle$ in N and a sequence $\langle \eta_0, \eta_1, \ldots, \eta_m \rangle$ in Ω such that: (i) $\eta_0 = \alpha$, (ii) $\eta_m = \beta$ and (iii) for every $k = 0, \ldots, m-1$, $\eta_{k+1} \in P_{i_{k+1}}(\eta_k)$.

REMARK 4. Note that, although P_* is always non-empty-valued and transitive, in general it need not be euclidean (despite the fact that the individual possibility correspondences are: for an example see Figure 3; recall that − cf. Remark 3 − P_* is euclidean if and only if B_* satisfies Negative Introspection: $\forall E \subseteq \Omega$, $\neg B_* E \subseteq B_*\neg B_* E$)

REMARK 5. In order to capture the notion of common belief in a model, one needs to extend the language by adding another operator $\Box_*$. If ϕ is a formula, the intended interpretation of $\Box_*\phi$ is "it is common belief that ϕ" and the truth set of $\Box_*\phi$ is given by $B_* E = \{\omega \in \Omega : P_*(\omega) \subseteq E\}$, where E is the truth set of ϕ.

EXAMPLE 3. Consider again the frame of Figures 1 and 2. The common possibility correspondence is given by $P_*(\beta) = \{\beta\}$ and $P_*(\gamma) = P_*(\tau) = \{\beta, \gamma, \tau\}$. Figure 3 illustrates P_* and the model of Figure 2 with the extended language that includes the common belief operator $\Box_*$.

---

[5] See, for example, Bonanno (1996), Fagin et al (1995). Halpern and Moses (1992), Lismont and Mongin (1994, 1996). These authors also show that the common belief operator can be alternatively defined by means of a finite

At state γ individual 1 wrongly believes that it is common belief that the universe is not expanding; hence, since $\gamma \in P_2(\tau)$, at state τ individual 2 considers it possible that individual 1 has such incorrect beliefs ($\neg\Box_2\neg\Box_1\Box_*\neg\pi$ is true at τ).

Insert Figure 3

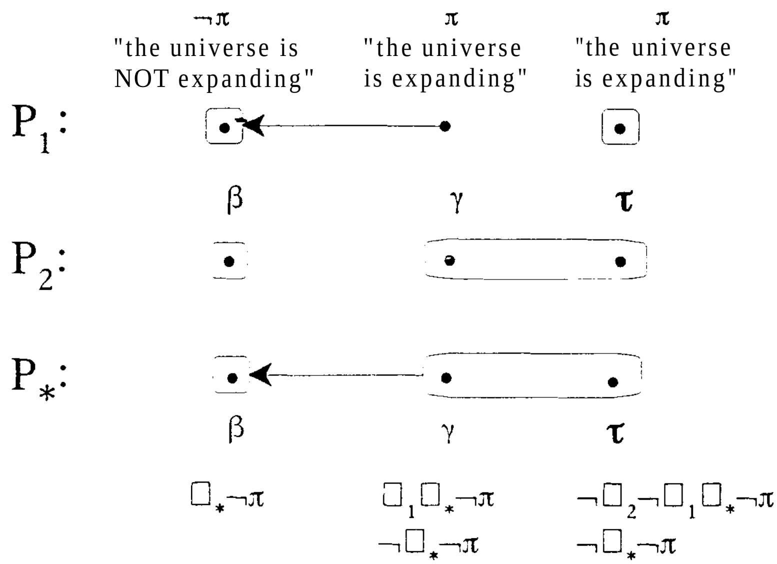

Figure 3

REMARK 6. A proposition is commonly believed if and only if everybody believes that it is commonly believed: for every $E \subseteq \Omega$, $B_*E = \bigcap_{i\in N} B_i B_* E$.

list of axioms, rather than as an infinite conjunction.

3. Properties of beliefs: (1) common belief in no error

Properties of interactive beliefs are to be defined locally, i.e. with respect to the true state τ. An equivalent, and mathematically more elegant, alternative is to define a property as an event, i.e. a set of states; the property is then satisfied at the true state τ if and only if τ belongs to that event. A characterization result will correspondingly be stated as the equality of two events.

Let $\mathbf{T}_i$ (for Truth of i's beliefs) be the following event:

$$\mathbf{T}_i = \bigcap_{E \in 2^{\Omega}} \neg(B_i E \cap \neg E)$$

Thus, for every $\alpha \in \Omega$, $\alpha \in \mathbf{T}_i$ if and only if individual i is correct in everything she believes (for every $E \subseteq \Omega$, if $\alpha \in B_i E$ then $\alpha \in E$). It is well known that $\alpha \in \mathbf{T}_i$ if and only if $\alpha \in P_i(\alpha)$ (for example, in the frame of Figure 1, $\mathbf{T}_1 = \{\beta, \tau\}$, while $\mathbf{T}_2 = \Omega$). It follows that $\alpha \in B_i \mathbf{T}_i$ if and only if, for all $\beta \in P_i(\alpha)$, $\beta \in P_i(\beta)$. By negative introspection of i's beliefs (euclideanness of P_i : see Remark 3) this property is satisfied at every state, that is, for every individual i, $B_i \mathbf{T}_i = \Omega$

Negative introspection prevents an individual from considering it possible that her own beliefs are false. On the other hand, there is nothing in the definition of frame that prevents an individual from, attributing false beliefs to *another* individual. For example, in the model of Figure 2 $\gamma \in P_2(\gamma) \cap \neg \mathbf{T}_1$: at state γ individual 2 considers it possible that individual 1 has the wrong belief that the universe is not expanding.

Let T (for Truth) be the following event:

$$\mathbf{T} = \bigcap_{i \in N} \mathbf{T}_i$$

Thus, for every $\alpha \in \Omega$, $\alpha \in T$ if and only if no individual has any false beliefs at α [5]. For example, in the frame of Figure 1, $T = \{\beta, \tau\}$ and, therefore, $B_*T = \{\beta\}$. We call B_*T the event that there is *common belief in no error*. This property has recently been shown to have important implications in the epistemic foundation of solution concepts in game theory (see, for example, Ben Porath, 1992, Stalnaker, 1994, 1996, Stuart, 1996). Proposition 1 below highlights some of the intersubjective implications of common belief in no error.

Given two individuals, i and j, and a state α, we say that *i* is *like-minded with j at* α if and only if i shares all the beliefs that she attributes to j, that is, for every event E, if $\alpha \in B_iB_jE$ then $\alpha \in B_iE$. Let $\mathbf{L}_{ij}$ be the event that i is like minded with j:

$$\mathbf{L}_{ij} = \bigcap_{E \in 2^{\Omega}} \neg\,(B_iB_jE \cap \neg B_iE).$$

Let L be the event that every individual is like-minded with every other individual:

$$\mathbf{L} = \bigcap_{i \in N} \bigcap_{j \in N} \mathbf{L}_{ij}\,.$$

[5] $\alpha \in T$ if and only if $\alpha \in \bigcap_{i \in N} P_i(\alpha)$. It follows that $\alpha \in B_*T$ if and only if, for all $\beta \in P_*(\alpha)$, $\beta \in \bigcap_{i \in N} I_i(\beta)$

Note that, in general, like-mindedness and correctness of beliefs are unrelated properties, that is, in general $T \not\subseteq L$ and $L \not\subseteq T$[6]. However, it is a consequence of negative introspection of individual beliefs that *public* like-mindedness and public correctness of beliefs coincide (for a proof of Proposition 1 see Bonanno and Nehring, 1997b).

PROPOSITION 1. $B_*L = B_*T$.

Thus common belief in no error is equivalent to common belief that every individual shares all the beliefs that she attributes to other individuals.

When the S5 logic is postulated for individual beliefs, then one obtains a partitional frame where $T = B_*T = \Omega$. One can capture the Truth Axiom ($\forall E \subseteq \Omega, \forall i \in N, B_iE \subseteq E$) as a *local* property of beliefs as follows.

DEFINITION 3. For every $\alpha \in \Omega$, the Truth *Axiom holds at* α if and only if

$$\alpha \in T \cap B_*T.$$

The above definition is justified by the following observation. Given a frame $\langle N, \Omega, \tau, \{P_i\}_{i \in N} \rangle$,

[6] Example 1: consider the following frame $N = \{1,2\}$, $\Omega = \{\tau, \beta\}$, $P_1(\tau) = P_1(\beta) = \{\tau\}$, $P_2(\beta) = P_2(\tau) = \{\tau, \beta\}$. Then $T = \{\tau\}$ while $L = \emptyset$. Example 2: in the frame $N = \{1,2\}$, $\Omega = \{\tau, \beta\}$, $P_i(\omega) = \{\beta\}$ for all $i \in N$ and $\omega \in \Omega$, $T = \{\beta\}$, while $L = \Omega$.

define the τ-*reduced frame* as the frame $(N, \Omega', \tau, \{P_i'\}_{i\in N})$ where 52' $= P_*(\tau)$ u $\{\tau\}$ and P_i' is the restriction of P_i to Q'. Let B_i' be the corresponding belief operator of individual i and P_*' the corresponding common possibility corresponclence. Then P_*' is the restriction of P_* to 52' [in particular, $P_*'(\tau) = P_*(\tau)$] and for every $E' \subseteq \Omega'$ $B_i'E' = BE' \cap \Omega'$. If $(N, \Omega, \tau, \{P_i\}_{i\in N})$ is a frame where $\tau \in$ T n B_***T**, then in the τ-reduced frame the following is true: $\forall i \in N$, $\forall E' \subseteq \Omega'$, $B_i' E' \subseteq E'$ (note, however, that in the originall frame in general it is not true that $\forall i \in N$, $\forall E \subseteq \Omega$, $B_i E \subseteq E$: see Figure 4a). Thus the τ-reduced frame is a partitional frame (unlike the original frame, in general). Figure 4b shows the τ-reduced frame corresponding to the frame of Figure 4a.

Insert Figure 4

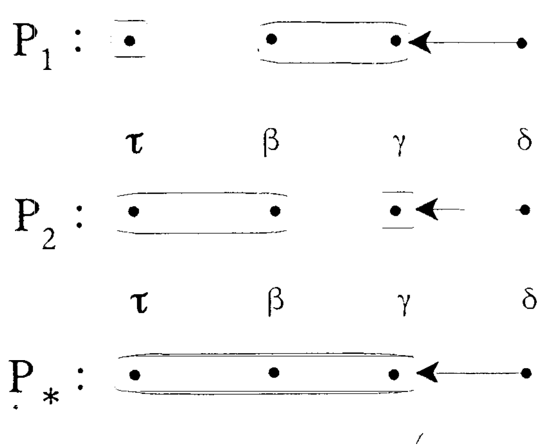

Figure 4a

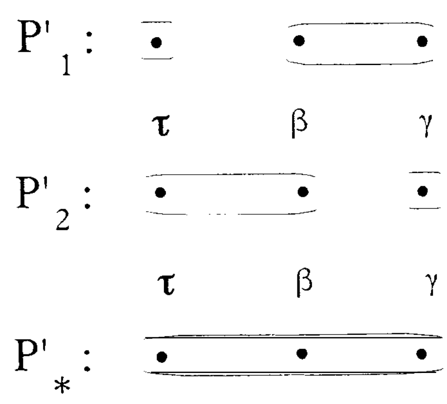

Figure 4b

A weaker property than common belief in no error is Agreement, defined as the common ***possibility*** of public like-mindedness and denoted by **A:**

$$\mathbf{A} = \neg B_* \neg B_* \mathbf{L} = \neg B_* \neg B_* \mathbf{T}.$$

The term "Agreement" is justified by the fact that this weaker property is equivalent to the impossibility of "agreeing to disagree" about qualitative belief indices (for a proof see Bonanno and Nehring, 1997b).

To gain further insight into the property of common belief in no error and the Truth Axiom we introduce two more properties that, together with Agreement, provide a decomposition of the Truth Axiom.

Let $\mathbf{T}_{CB}$ (for Truth about common belief) and T^* (for Truth of common belief) be the following events

$$\mathbf{T}_{CB} = \bigcap_{i \in N} \bigcap_{E \in 2^\Omega} \neg(B_i B_* E \cap \neg B_* E)$$

$$\mathbf{T}^* = \bigcap_{E \in 2^\Omega} \neg(B_* E \cap \neg E) .$$

$\mathbf{T}_{CB}$ captures the notion that individuals are correct in their beliefs about what is commonly believed: $a \in \mathbf{T}_{CB}$ if and only if, for every event E and individual i. if, at a, individual i believes that E is commonly believed, then, at a, E is indeed commonly believed (if $\alpha \in B_i B_* E$ then $\alpha \in B_* E$). On the other hand, $\alpha \in T^*$ if and only if at a whatever is commonly believed is true (for every event E, if $\alpha \in B_* E$ then $\alpha \in E$)[7]. Clearly, Truth of common belief is qualitatively weaker

---

[7] It is straightforward that $\alpha \in \mathbf{T}^*$ if and only if, $\alpha \in P_*(\alpha)$.

than Truth; given that $B_* T^* = \Omega$, T^* can be viewed as Truth shorn of any intersubjective implications.

The following proposition gives a decomposition of the Truth Axiom in terms of quasi-coherence, Truth of common belief and (common belief of) Truth about common belief (for a proof see Bonanno and Nehring, 1997b).

PROPOSITION 2. $\mathbf{T} \cap B_* \mathbf{T} = T^* \cap B_* \mathbf{T}_{CB} \cap \mathbf{A}$.

REMARK 7. None of T^*, $\mathbf{T}_{CB}$ and $B_* \mathbf{T}_{CB}$, either individually or in conjunction with the others, has any "agreement" implications. This can be seen from Figure 5 where $T^* = \mathbf{T}_{CB} = B_* T_{CB} = \Omega$ and yet at both τ and β the individuals agree to strongly disagree, in the sense that it Is common belief that individual 2 believes E and individual 1 believes not E, where $E = \{\tau\}$: $B_*(B_1 \neg E \cap B_2 E) = \Omega$. On the other hand, as remarked before, **A** is precisely the property that rules out such phenomena.

Insert Figure 5

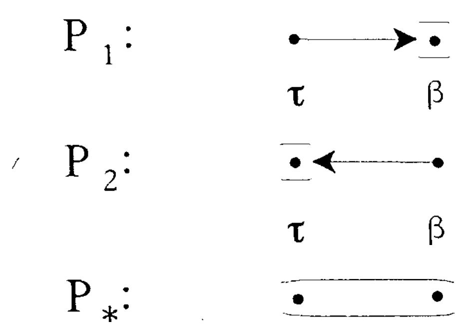

Figure 5

4. Properties of beliefs: (2) negative introspection of common belief

As noted in Remark 4, the common possibility correspondence P_* satisfies non-empty-valuedness and transitivity but not euclideanness. It follows (cf. Remark 3) that the common belief operator B_* satisfies consistency ($B_*E \subseteq \neg B_*\neg E$) and positive introspection ($B_*E \subseteq B_*B_*E$) but not necessarily negative introspection ($\neg B_*E \subseteq B_*\neg B_*E$).[8] Thus Negative Introspection of common belief implies intersubjective restrictions on beliefs. The purpose of this section is to find out what these restrictions are.

Let (NI stands for "Negative Introspection")

$$\mathbf{NI} = \bigcap_{E \in 2^{\Omega}} (B_*E \cup B_*\neg B_*E)$$

Thus $\alpha \in \mathbf{NI}$ if and only if – for every event E – whenever at α it is not common belief that E, then, at α, it is common belief that E is not commonly believed (if $\alpha \in \neg B_*E$ then $\alpha \in B_*\neg B_*E$).

REMARK 8. $\alpha \in \mathbf{M}$ if and only if, $\forall \beta, \gamma \in P_*(\alpha)$, $\gamma \in P_*(\beta)$.[9]

The following propositions are proved in Bonanno and Nehring (1997a).

---

[8] For example, in the frame of Figure 3, let $E = \{\beta\}$. Then $B_*E = \{\beta\}$. Hence $\neg B_*E = \{\gamma, \tau\}$ and $B_*\neg B_*E = \varnothing$. Thus $\neg B_*E \not\subseteq B_*\neg B_*E$.

[9] *Proof.* (i) Suppose that $\beta, \gamma \in P_*(\alpha)$ and $\gamma \notin P_*(\beta)$. Let $E = P_*(\beta)$. Since $\gamma \in P_*(\alpha) \cap \neg E$, $P_*(\alpha) \not\subseteq E$, that is, $\alpha \in \neg B_*E$. Since $P_*(\beta) = E$, $\beta \in B_*E$. Hence, since $\beta \in P_*(\alpha)$, $P_*(\alpha) \cap B_*E \neq \varnothing$, that is, $\alpha \in \neg B_*\neg B_*E$. Thus $\alpha \in \neg B_*E \cap \neg B_*\neg B_*E$. Hence $\alpha \notin$ NI. (ii) Conversely, suppose that $\alpha \notin$ NI. Then there exists an $E \subseteq \Omega$

PROPOSITION 3. $NI = T_{CB} \cap B_*T_{CB}$.

According to Proposition 3, Negative Introspection of common belief hinges on common knowledge of truth restricted to beliefs about common belief. One may wonder whether there is something qualitatively different about the truth of this very special type of beliefs. This question can be answered affirmatively, in that truth about common belief is necessary and sufficient for individuals' beliefs about common belief to ***coincide:*** we call this "Shared Worlds". (By comparison, having correct beliefs about what others believe, in general, does not imply sharing their beliefs.) Let SW be the following event:

$$SW = \bigcap_{i \in N} \bigcap_{j \in N} \bigcap_{E \in 2^{\Omega}} (\neg B_i B_* E \cup B_j B_* E)$$

SW captures the notion that individuals agree on what is commonly believed: $\alpha \in SW$ if and only if, for every event E, whenever one individual believes that it is common belief that E, then every other individual believes that too[10].

PROPOSITION 4. $SW = T_{CB}$.

---

such that $\alpha \in \neg B_* E \cap \neg B_* \neg B_* E$. Since $\alpha \in \neg B_* \neg B_* E$, there exists a $\beta \in P_*(\alpha)$ such that $\beta \in B_* E$, that is, $P_*(\beta) \subseteq E$. Since $\alpha \notin B_* E$, there exists a $\gamma \in P_*(\alpha)$ such that $\gamma \notin E$. Hence $\gamma \not\subset P_*(\beta)$.

[10] Note that $\alpha \in SW$ requires that at α the individuals share the same "model of the world" $\Omega_i(\alpha) \equiv \bigcup_{\omega \in I_i(\alpha)} I_*(\omega)$, that is, $\alpha \in SW$ if and only if for all $i, j \in N$, $\Omega_i(\alpha) = \Omega_j(\alpha)$. Note that *common belief* in Shared Worlds rules out, by definition. even uncertainty about the others' model of the world, as the following example shows: $N = \{1,2\}$. $\Omega = \{\tau, \beta, \gamma\}$, $I_1(\tau) = \{\tau\}$, $I_1(\beta) = I_1(\gamma) = \{\beta\}$, $I_2(\tau) = I_2(\gamma) = \{\tau, \gamma\}$, $I_2(\beta) = \{\beta\}$. Thus $I_*(\tau) = I_*(\gamma) = \{\tau, \beta, \gamma\}$ and $I_*(\beta) = \{\beta\}$. Here $SW = \{\tau, \beta\}$. However, while $\tau \in SW$, $\tau \notin B_* SW = \{\beta\}$: at τ (and γ) individual 2 is uncertain as to whether 1's personal model is $\{\beta\}$ or Ω.

Finally, since **NI** can be viewed as describing the "logic" of common belief, a global (or "axiomatic") version of Proposition 3 which incorporates Proposition 4 is of some interest. It is provided in the following corollary.

COROLLARY 1. $\mathbf{NI} = \Omega$ if and only if $\mathbf{SW} = \Omega$ [11].

5. Conclusion

This paper reviewed the semantic approach to belief and common belief. Properties such as common belief in no error and Negative Introspection of common belief were examined and decomposed into further properties of individual beliefs. For the syntactic approach to belief and common belief the reader is referred to Bonanno (1996), Fagin et al (1995), Halpern and Moses (1992), Lismont and Mongin (1994, 1996).

A companion paper discusses the importance of the properties considered above for the epistemic foundations of solution concepts in game theory. In particular, Negative Introspection of common belief plays a role in the extension of Aumann's (1987) characterization of correlated equilibrium to situations of incomplete information, while common belief in no error plays a role in the justification for backward induction in an important class of extensive-form games, which includes the finitely repeated prisoners' dilemma and the centipede game. A third paper examines the intersubjective interpretation of the Common Prior Assumption in the context of incomplete information and various generalizations of the notion of "agreeing to disagree" introduced by Aumann (1976).

[11] That is, (i) and (ii) below are equivalent:

(i) $\forall E \subseteq \Omega$, $\neg B_* E \subseteq B_* \neg B_* E$,

(ii) $\forall i, j \in N$, $\forall E \subseteq \Omega$, $B B_* E \subseteq B B_* E$.

References

Aumann, R. (1976), Agreeing to disagree, *Annals of Statistics,* 4, 1236-1239.

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

Ben-Porath, E. (1992), Rationality, Nash equilibrium and backward induction in perfect information games, mimeo, Tel Aviv University.

Bonanno, G. (1996), On the logic of common belief, *Mathematical Logic Quarterly*, 42, 305-311.

Bonanno, G. and Nehring, K. (1997a), On the logic of common belief: two local characterizations of negative introspection, Working Paper, University of California, Davis.

Bonanno, G. and Nehring, K. (1997b), Assessing the Truth Axiom under incomplete information, Working Paper, University of California, Davis..

Chellas, B. (1984), *Modal logic,* Cambridge University Press, Cambridge.

Fagin, R., J. Halpern, Y. Moses and M. Vardi (1995), *Reasoning about knowledge, MIT* Press, Cambridge.

Geanakoplos, J. (1992).Common knowledge, *Journal of Economic Perspectives,* 6, 53-82.

Halpem, J. and Y. Moses (1992), A guide to completeness and complexity for modal logics of knowledge and belief, *Artificial intelligence*, 54, 319-379.

Lismont, L. and P. Mongin (1994), On the logic of common belief and common knowledge, *Theory and Decision,* 37, 75-106.

Lismont, L. and P. Mongin (1995), Belief closure: a semantics for common knowledge for modal propositional logic, *Mathematical Social Sciences,* 30, 127-153.

Stalnaker, R. (1994), On the evaluation of solution concepts, *Theory and Decision,* 37, 49-74.

Stalnaker, R. (1996), Knowledge! belief and counterfactual reasoning in games, *Economics and Philosophy,* 12, 133-163.

Stuart, H. (1996), Common belief of rationality in the finitely repeated Prisoners' Dilemma, mimeo, Harvard Business School.