INTERSUBJECTIVE CONSISTENCY OF BELIEFS AND THE LOGIC OF COMMON BELIEF

Giacomo Bonanno
and
Klaus Nehring

Working Paper Series #95-08

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

INTERSUBJECTIVE CONSISTENCY OF BELIEFS AND THE LOGIC OF COMMON BELIEF

Giacomo Bonanno
and
Klaus Nehring

Working Paper Series No. 95-08
February 1995

Abstract

'We characterize the class of n-person belief systems for which common belief has the properties of the strongest logic of belief, KD45. The characterizing condition states that individuals are not too mistaken in their beliefs about common beliefs. It is shown to be considerably weaker than the consistency condition on interpersonal beliefs implied by the common knowledge assumption: it allows individuals to "agree to disagree" and to be quite incorrect about others' beliefs.

1. Introduction

The concepts of common knowledge and common belief have been discussed extensively in the literature, both syntactically and semantically.[1] At the *individual* level the difference between knowledge and belief is usually identified with the presence or absence of the *Truth Axiom* T: $\Box A \to A$, which is interpreted as "if individual i believes that A, then A". In such a case the individual is often said to *know* that A (thus it is possible for an individual to believe a false proposition but she cannot know a false proposition). Going to the interpersonal level, the literature then distinguishes between *common* knowledge and common belief on the basis of whether or not the Truth Axiom is postulated at the individual level. However, while at the individual level the Truth Axiom captures merely a relationship between the individual's beliefs and the external world, at the interpersonal level it has very strong implications. For example, the following is a consequence of the Truth Axiom: $\Box\Box A \to \Box A$, that is, if individual i believes that individual j believes that A, then individual i herself believes that A.[2] Thus, in contrast to the other axioms, the Truth Axiom does not merely reflect individual agents' "logic of belief'. (The reason why the Truth Axiom is much stronger in an interpersonal context than appears at first

---

[1] Aumann (1976), Bacharach (1985), Binmore and Brandenburger (1990), Bonanno (1994), Colombetti (1993), Geanakoplos (1992), Geanakoplos and Polemarchakis (1982), Halpern (1986), Halpern and Moses (1992), Kaneko and Nagashima (1993), Lewis (1969), Lismont (1993), Lismont and. Mongin (1994), Milgrom (1981), Monderer and Samet (1989), Rubinstein and Wolinsky (1990), Samet (1990), Shin (1993), Tan and Werlang (1985).

[2] This can be seen as follows. First of all, any axiomatization of belief will include the so called K axiom: $(\Box_i (A \to B) \wedge \Box_i A) \to \Box_i B$ and the so-called rule of Necessitation: from A to infer $\Box_i A$. Now, from the Truth Axiom for individual j, $\Box_j A \to A$, and the rule of Necessitation we obtain the following theorem $\Box_i (\Box_j A \to A)$. Hence, by Propositional Logic, $\Box_i \Box_j A \to \Box_i \Box_j A \wedge \Box_i (\Box_j A \to A)$. By axiom K the consequent of this last formula implies $\Box A$.

glance is that it amounts to assuming that agreement of any individual's belief with the truth is common knowledge). Given its logical force, it is not surprising to find that it has strong implications for the logic of common knowledge. In particular, if each individual's beliefs satisfy the strongest logic of knowledge (namely SS or KT5), the associated common knowledge operator satisfies this logic too (for technical details on this point and the following see the beginning of Section 2). Such is not the case for belief bereft of the Truth Axiom, even the strongest logic for individual belief (KD45) is insufficient to ensure the satisfaction of the "Negative Introspection" axiom for common belief $\neg\boxed{*}\,\mathbf{A} \to \boxed{*}\,\neg\boxed{*}\,\mathbf{A}$ (where $\boxed{*}$ denotes the common belief operator; see Colombetti, 1993, and Lismont and Mongin, 1994). That is to say, it can happen that neither is **A** commonly believed nor is it common belief that **A** is not commonly believed.

Negative Introspection has been variously challenged based on arguments of bounded rationality (see, for example, Samet, 1990). By contrast, we maintain full rationality of individuals and investigate under what conditions "full rationality" – that is, Negative Introspection – holds for common beliefs. We note in Section 2 that Negative Introspection for common belief (from now on, we shall refer to it as "axiom 5^*") amounts to common beliefs being "publicly known". The main result, Theorem 1, shows that common belief satisfies axiom 5^* if and only if individuals are not too mistaken about common beliefs. Formally, this condition is expressed as follows:

$$C^* \qquad \boxed{i}\,\boxed{*}\,A \to \neg\boxed{j}\,\neg A$$

(recall that $\boxed{i}$ denotes the belief operator of individual **i** and $\boxed{*}$ the common belief operator). Thus C* has the following interpretation: if individual **i** believes that it is common belief that **A** then it is not the case that individual j believes that not **A.** We call C* the *-Compatibility Axiom. C* is much weaker than the Truth Axiom **($\boxed{i}$ A → A).** It allows individuals to "agree to disagree" and individuals' beliefs about others' beliefs can be quite incorrect. In other words, the strong logic of common belief (KD45) turns out to be quite robust, if not completely so.

2. Compatibility of belief systems

Semantically, the notion of common knowledge is represented by the meet of the information partitions of the individuals. This is a partition itself and, therefore, it validates the same axioms that are postulated for the individuals. Requiring information partitions at the individual level amounts to postulating the following axiom schemata for every individual **i** (we use the notation and names that are standard in modal logic: see, for example, Chellas, 1980):

K. $\boxed{i}(A \to B) \wedge \boxed{i}A \to \boxed{i}B$

T. $\boxed{i}A \to A$

5. $\neg\boxed{i}A \to \boxed{i}\neg\boxed{i}A$

as well as the rule of inference of *Necessitation:* from *A* to infer $\boxed{i}$ **A.** Axiom schema 5 **is** sometimes referred to as the *Negative Introspection* axiom: if the individual does not know that **A** then she knows that she does not know that **A.** Since the notion of common knowledge is captured by the meet of the information partitions, the common knowledge operator will also satisfy axioms K, T and 5; in particular, it will be true that if a proposition is not common knowledge then it is common knowledge that it is not common knowledge.

Moving from knowledge to belief implies dropping the Truth Axiom T. The strongest axiomatization of belief at the individual level will then be represented by the following axiom schemata (as well as the inference rule of Necessitation):[3]

K. $\boxed{i}(A \to B) \wedge \boxed{i}A \to \boxed{i}B$

D. $\boxed{i}A \to \neg\boxed{i}\neg A$

4. $\boxed{i}A \to \boxed{i}\boxed{i}A$

5. $\neg\boxed{i}A \to \boxed{i}\neg\boxed{i}A.$

Axiom schema D is the *Consistency* axiom: it says that an individual cannot believe that **A** and at the same time believe that not **A.** Axiom schema 4 is often referred to as the *Positive Introspection* axiom: if the individual knows that **A** then she knows that she knows that **A.** Semantically, the above axiom schemata correspond to the following properties of the accessibility relation (cf. Chellas, 1980, pp. 76-80):[4]

---

[3] It is well known (see Chellas, 1980) that axioms D and 4 are theorems of the KT5 (or S5) logic.

[4] Economists often use *information functions* rather than accessibility relations. The two notions, however, are equivalent. An information function is a function $I : W \to 2^W$, where W is a set of "states" or "possible worlds". Given such a function one can define the corresponding accessibility

AXIOM SCHEMA	PROPERTY OF ACCESSIBILITY RELATION
K. $\boxed{i}(A \to B) \wedge \boxed{i}A \to \boxed{i}B$	no restrictions
T. $\boxed{i}A \to A$	**Reflexivity:** $\forall \alpha,\ \alpha R_i \alpha$
D. $\boxed{i}A \to \neg\boxed{i}\neg A$	**Seriality:** $\forall \alpha,\ \exists \beta : \alpha R_i \beta$
4. $\boxed{i}A \to \boxed{i}\boxed{i}A$	**Transitivity:** $\forall \alpha, \forall \beta, \forall \gamma$, if $\alpha R_i \beta$ and $\beta R_i \gamma$ then $\alpha R_i \gamma$
5. $\neg\boxed{i}A \to \boxed{i}\neg\boxed{i}A$	**Euclideanness:** $\forall \alpha, \forall \beta, \forall \gamma$, if $\alpha R_i \beta$ and $\alpha R_i \gamma$ then $\beta R_i \gamma$

For a syntactic axiomatization of the concept of *common* ***belief*** see Halpern and Moses (1992), Lismont (1993) and Lismont and Mongin (1994). We review it in Appendix 1. Semantically, the notion of common belief is captured by the transitive closure of the union of the accessibility relations of the individuals.[5] It is easy to see from this that if the individuals' belief operators satisfy axiom D (respectively, T) then the common belief operator also satisfies axiom D (respectively, T). Furthermore, the common belief operator will always satisfy axiom 4.

---

relation as follows: $\alpha R \beta$ if and only if $\beta \in I(\alpha)$. Conversely, given an accessibility relation R on W one can define the corresponding information function as follows: $I(\alpha) = \{\beta : \alpha R \beta\}$. For example, if $I_i(\cdot)$ denotes the information function of individual i, then reflexivity of i's accessibility relation corresponds to the following property of the information function: for every α, $\alpha \in I_i(\alpha)$.

[5] Let $R_1, \ldots, R_n$ be binary relations on a set W and let R_* denote the transitive closure of the union of these relations. Then R_* is a binary relation on W defined as follows: for all $\alpha, \beta \in W$, $\alpha R_* \beta$ if and only if there is a sequence $i_1, \ldots i_m$ in (1, ..., n) and a sequence $\eta_0, \eta_1, \ldots, \eta_m$ in W such that: (i) $\eta_0 = \alpha$, (ii) $\eta_m = \beta$ and (iii) for every $k = 0, \ldots, m-1$, $\eta_k R_{i_{k+1}} \eta_{k+1}$.

Therefore, the only property at issue is the Negative Introspection axiom at the "interpersonal level", which we will denote by 5^*: $\neg\boxed{*}A \rightarrow \square\neg\boxed{*}A$.

It has been noted (Colombetti, 1993, Lismont and Mongin, 1994) that, even if one imposes axioms K, D, 4 and 5 at the individual level, axiom 5^* need not hold, that is, it is possible that A is not common belief and yet it is not the case that it is common belief that A is not common belief An example of this is given below (Figure 1). Note that 5^* implies the following:

$$P^*. \qquad \neg\boxed{*}A \rightarrow \boxed{i}\neg\boxed{*}A.$$

Since the complementary axiom:

$$PR. \qquad \boxed{*}A \rightarrow \boxed{i}\boxed{*}A$$

is part of the axiomatization of common belief (see Appendix 1), P* amounts to saying that there is shared knowledge about common belief. Note also that 5^* is, in fact, equivalent to P* (a proof of this claim is given in Appendix 2).

The following example shows that, even if individual beliefs satisfy KD45, the common belief operator need not satisfy 5*. Let there be two individuals, 1 and 2 and two worlds, α and β. Let p be a sentence which is true at α and false at β. Let the accessibility relations R_1 and R_2 be as illustrated in Figure 1 and let $Tr\, R_1 \cup R_2$ be the transitive closure of the union of R_1 and R_2 (which, in this case, coincides with $R_1 \cup R_2$: see Figure 1).

Insert Figure 1

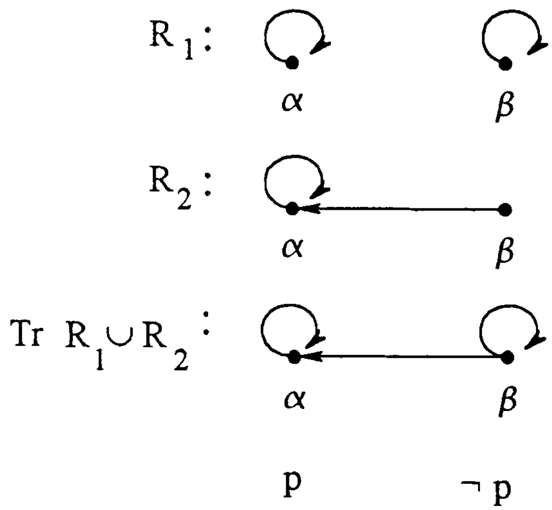

FIGURE 1

Note that R_1 is an equivalence relation and R_2 is serial, transitive and euclidean. Hence they both validate K, D, 4 and 5. Denote Tr $R_1 \cup R_2$ by R_*. Since p is false at β and $\beta R_* \beta$, p is not common belief at β, that is, the formula $\neg \boxed{*}(p)$ is true at β.[6] On the other hand, p *is* common belief at α, that is, the formula $\boxed{*}(p)$ is true at α. It follows that, since $\beta R_* \alpha$, it is not common belief at β that p is *not* common belief, that is, the following formula is ***false*** at β: $\boxed{*} \neg \boxed{*}$ **(p).** Thus we can conclude that at β the formula $(\neg \boxed{*}\, p \rightarrow \boxed{*} \neg \boxed{*}\, p)$ which is an instance of axiom schema 5*, is false. To put it differently, the above example shows that the transitive closure of n relations, each of which is serial, transitive and euclidean, is not necessarily euclidean.

---

[6] Recall that. for every formula **A,** $\Box$**A** is true at world δ if and only if **A** is true at every world γ such that $\delta R_i \gamma$. Similarly for $\boxed{*}$ **A.** See Appendix 1 for more details on this.

The example of Figure 1 has the following feature: at world β individual 1 believes that not p while individual **2** believes that p. In other words their beliefs are completely ***incompatible*** (the two individuals are "worlds apart"). This situation cannot arise when the Truth Axiom is imposed at the individual level, because at any world a there will be a world, namely a itself, which everybody can access, hence complete disagreement is ruled out. In this paper we explore the implications of merely requiring interpersonal compatibility of beliefs, in various forms, while avoiding the extra strong implications of the Truth Axiom.

We begin with a simple, but rather strong, axiom which we call the ***Compatibility*** Axiom and denote by C:

$$\text{C.} \qquad \boxed{i}\, A \rightarrow \neg \boxed{j}\, \neg A.$$

Axiom C is the interpersonal counterpart of the consistency axiom D: it says that it is not possible for one individual to believe that A and, at the same time, for another individual to believe that not **A.** Thus if individual i believes that A, then individual j must allow for the possibility that **A.** Note that by choosing i = j we obtain the consistency axiom D at the individual level. Thus C implies D for every individual. The following proposition gives the semantic counterpart of axiom C. To prove Proposition 1 we need to make use of notation and definitions from modal logic and therefore we postpone the proof to Appendix 2.

PROPOSITION 1. Axiom C is characterized by the following property of the ***set*** $\{R,, ..., R_n\}$ (where R_i is the accessibility relation of individual i, i = 1, ..., n):

Compatibility: $\forall i, \forall j, Va, 3\beta : \alpha R_i \beta$ and $\alpha R_j \beta$.

That is, every model where the set {R,, ..., R_n} satisfies the Compatibility property validates axiom C and, conversely, given a set of relations {R,, ..., R_n} that violates the Compatibility property, there is a model based on it and an instance of axiom C which is falsified at some world in the model.

In Section **3** it will be shown that Compatibility is sufficient to yield 5*; however, it is far from necessary, as the example of Figure 2 below shows.

Insert Figure 2

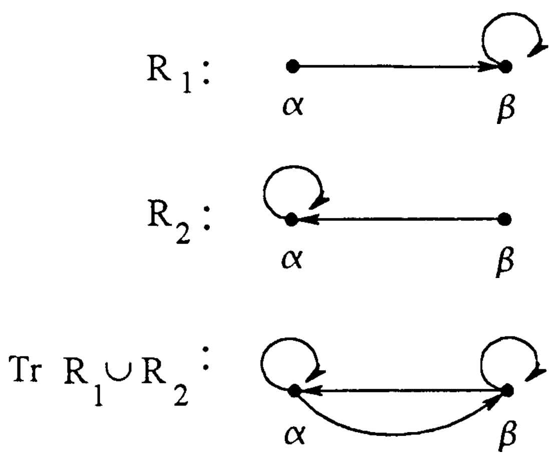

FIGURE 2

In the above example, Compatibility is violated (hence, by Proposition 1, there is a model based on this frame that falsifies axiom C). On the other hand, Tr $R_1 \cup R_2$ is euclidean (in fact it is an equivalence relation), hence 5^* is valid in this frame.

Necessary conditions for 5^* generally involve the common belief operator in their statement (on this point see the example of Figure 7 in Section 3). As theorem 1 below shows, a necessary and sufficient condition for 5^* is that individuals be correct in their beliefs that something is commonly believed:

$$\text{TN}^*. \qquad \boxed{i}\boxed{*}A \rightarrow \boxed{*}A.$$

Note that: it follows from the definition of common belief that (if individual beliefs are consistent) individuals must be correct in their belief that something is not common belief (for a proof see Appendix 2):

$$\boxed{i}\neg\boxed{*}A \rightarrow \neg\boxed{*}A.$$

Thus TN^* amounts to requiring that individual beliefs *about* common beliefs be correct. Since 5^* is also equivalent to P* ($\neg\boxed{*}A \rightarrow \boxed{i}\neg\boxed{*}A$), it is equivalent to the property that common beliefs be "publicly known". While interesting (and non-trivial), TN^* is somewhat lame as a characterization of 5*, since it involves restrictions on common beliefs themselves. As a result,, there is no straightforward way to infer sufficiency of C from that of TN^*. This problem is overcome by the following condition C^*:

$$C^*. \qquad \boxed{i}\boxed{*}A \rightarrow \neg\boxed{j}\neg A$$

We will call C* the *-Compatibility Axiom. It says that if individual i believes that it is common belief that A then it is not the case that individual j believes that not A. Thus C* requires individuals' beliefs about what is commonly believed to be "not too far from the truth".

Proposition 2, which is proved in Appendix 2, gives a characterization of' C* in terms of a property of the ***set*** of accessibility relations.

PROPOSITION 2. Axiom C^* is characterized by the following property of the set $\mathcal{R} = \{R,, \ldots, R_n\}$ (where R is the accessibility relation of individual i, i = 1, ..., n). Let $R_* = Tr \bigcup \mathcal{R}$ be the transitive closure of the union of the individual accessibility relations.

$$\textbf{*-Compatibility:} \qquad \forall i, \forall j, \forall \alpha, \exists \beta, \exists \gamma : \alpha R_i \beta \text{ and } \beta R_* \gamma \text{ and } \alpha R_j \gamma.$$

Our main result is contained in Theorem 1 below.

THEOREM 1. Assume that, for every individual i, the belief operator $\boxed{i}$ satisfies axioms D and 5 (as well as K and the rule of Necessitation). Then the following axioms are equivalent.

(i) 5^* $(\neg \boxed{*} A \to \boxed{*} \neg \boxed{*} A)$,

(ii) TN^* $(\boxed{i} \boxed{*} A \to \Box A)$,

(iii) C^* $(\boxed{i}\, q\, A \to \neg \boxed{j} \neg A)$.

Proof. (i) $\Rightarrow$ (ii). The proof is as follows (PL stands for "Propositional Logic"):

1.	$\neg \boxed{*} A \to \boxed{*} \neg \boxed{*} A$	(5^*)
2.	$\boxed{*} \neg \boxed{*} A \to \boxed{i} \neg \boxed{*} A$	(SB: see .Appendix 1)
3.	$\boxed{i} \neg \boxed{*} A \to \neg \boxed{i} \boxed{*} A$	(D for i)

4. $\neg\boxed{*}A \rightarrow \neg\boxed{i}\boxed{*}A$ (1, 2, 3, PL)

5. $\boxed{i}\boxed{*}A \rightarrow \boxed{*}A$ (4, PL).

(ii) $\Rightarrow$ (iii). The proof is as follows:

1. $\boxed{i}\boxed{*}A \rightarrow \boxed{*}A$ (TN^*)

2. $\boxed{*}A \rightarrow \boxed{j}A$ (SB: see Appendix 1)

3. $\boxed{j}A \rightarrow \neg\boxed{j}\neg A$ (D for j)

4. $\boxed{i}\boxed{*}A \rightarrow \neg\boxed{j}\neg A$ (1, 2, 3, PL).

The more difficult part is to prove that (iii) $\Rightarrow$ (i). In order to do this we need to introduce a new property of relations (which we call "quasi-euclideanness") and prove two lemmas.

DEFINITION. A binary relation R on a set W is called *quasi-euclidean* if it satisfies the following property:

$$\forall \alpha,\beta,\gamma \in W, \text{ if } \alpha R\beta \text{ and } \alpha R\gamma, \text{ then there exists an integer } k \geq 1 \text{ such that } \beta R^k \gamma$$

where $\beta R^k \gamma$ means that there is a path of length k from β to γ (that is, there is a sequence $\delta_0, \ldots, \delta_k$ in W such that: (i) $\delta_0 = \beta$, (ii) $\delta_k = \gamma$ and (iii) for every $j = 0, \ldots, k-1$, $\delta_j R \delta_{j+1}$).

(Thus quasi-euclideanness is a weakening of euclideanness, since the latter corresponds to the case where $k = 1$).

LEMMA **1.** Let R be a relation on the set W and let Tr R be its transitive closure (that is, the smallest transitive relation containing R). Then R is quasi-euclidean *if and only* if Tr R is euclidean.

Proof. ($\Rightarrow$) Let R be a quasi-euclidean relation and, to simplify the notation, denote the transitive closure of R by R,. We want to prove that R, is euclidean. The proof is illustrated in Figure **3**. Fix arbitrary α, β and y such that $\alpha R_* \beta$ and $\alpha R_* \gamma$. We need to show that $\beta R_* \gamma$. Since $\alpha R_* \beta$, there is sequence β_0, β_1, ..., β_m in W such that (i) β_0 = a, (ii) $\beta_m = \beta$ and (iii) for every $k = 0, ..., m-1$, $\beta_k R \beta_{k+1}$. Similarly, since $\alpha R_* \gamma$, there is a sequence γ_0, γ_1, ..., γ_s in W such that (i) γ_0 = a, (ii) γ_s = y and (iii) for every $t = 0, ..., s-1$, $\gamma_t R \gamma_{t+1}$. Since $\alpha R \beta_1$ and $\alpha R \gamma_1$ and R is quasi-euclidean, there is a positive integer ℓ and an R-path of length ℓ from β_1 to γ_1. Let 6 be the first node on this path. We want to show that for every $k \geq 1$ there is an R-path fiom β_k to 6. For $k = 1$ we have already proved it. By quasi-euclideanness of R, since $\beta_1 R \delta$ and $\beta_1 R \beta_2$, there is an R-path fiom β_2 to 6. Let E be the first node on this path. Then, since $\beta_2 R \varepsilon$ and $\beta_2 R \beta_3$, by euclideanness of R there is an R-path from β_3 to E. Joining this path with the path from E to 6 we obtain an R-path fiom β_3 to 6. By repeating this argument m times we obtain an R-path from β_m to δ, that is, a path **from** β to 6 (**since** $\beta_m = \beta$). Joining this path with the path from δ to γ_1 and then with the path fiom γ_1 to γ_t = y, we obtain an R-path from β to y. Hence, since R_* is the transitive closure of R, we have that $\beta R_* \gamma$.

(⊂)et R be a binary relation on the set W, whose transitive closure, denoted by R_*, is euclidean. We want to show that R is quasi-euclidean. Fix arbitrary a, β and y such that $\alpha R \beta$ and

$\alpha R\gamma$. Then we also have that $\alpha R_* \beta$ and $\alpha R_* \gamma$. Since R_* is euclidean, it follows that $\beta R_* \gamma$. By definition oftransitive closure, this means that there exists a sequence δ_0, 6,, ..., δ_m in W such that (i) 6, $= \beta$, (ii) δ_m = y and (iii) for every k = 0, ..., $m-1$, $\delta_k R \delta_{k+1}$. Hence R is quasi-euclidean. □

Insert Figure 3

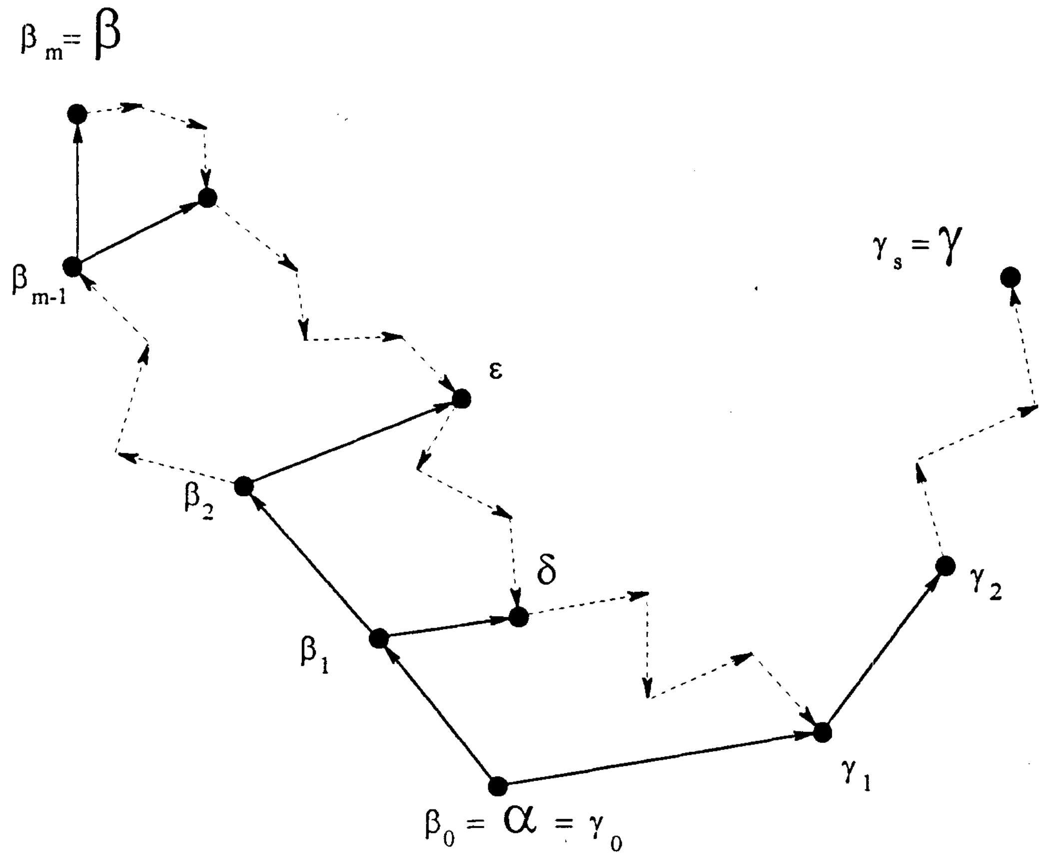

FIGURE 3

LEMMA 2. Let 22 = {R,, R_2, ..., R_n} be a set of binary relations on the set W and let $\cup\mathcal{R}$ be their union. If

(1) for all i = 1, ..., n, R_i is quasi-euclidean, and

(2) the set {R,, R,, ..., R_n} satisfies the *-Compatibility property,

then $\bigcup\mathcal{R}$ is quasi-euclidean.

Proof: The proof is illustrated in Figure 4. Fix α, 6 and ε such that a $\bigcup\mathcal{R}\,\delta$ and $\alpha \bigcup\mathcal{R}\,\varepsilon$. Then there exist i and j such that $\alpha R_i \delta$ and $\alpha R_j \varepsilon$. By *-Compatibility there exist β and y such that $\alpha R_i \beta$, $\beta R_* \gamma$ and $\alpha R_j \gamma$, where R_* denotes the transitive closure of $\bigcup\mathcal{R}$. By quasi-euclideanness of R, there exists an R_i-path from 6 to β. By definition of transitive closure, there exists an $\bigcup\mathcal{R}$-path from β to γ. Finally, by euclideanness of R_j, there exists an R_j-path from γ to ε. Hence there is an $\bigcup\mathcal{R}$-path from 6 to E, that is, $\bigcup\mathcal{R}$ is quasi-euclidean. □

Insert Figure 4

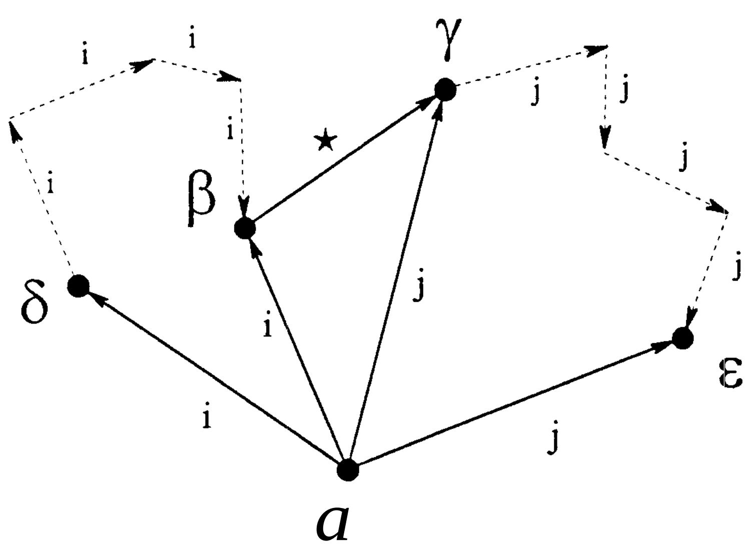

FIGURE 4

We can now complete the proof of Theorem 1. By Proposition 2, C* is valid in all frames where $\{R_1, ..., R_n\}$ satisfies *-Compatibility, in particular, in the class of such frames where every R_i is euclidean. By Lemma 2, in this class of frames $\bigcup\mathcal{R}$ is quasi-euclidean and, by

Lemma 1, the transitive closure R_* of $\bigcup\mathcal{R}$ is euclidean. Since 5^* is valid in the class of frames where R_* is euclidean, by the completeness theorem for the logic of common belief (see Halpern and Moses, 1992, Lismont, 1993, and Lismont and Mongin, 1994; see also Appendix 1), 5^* is a theorem of every normal logic of common belief where the individuals' belief operators satisfy axiom **5** and, furthermore, axiom C* holds at the "interpersonal level" (recall that axiom D for every individual is a consequence of c*). ■

Thus Theorem 1 says that a (normal) logic where the individual belief operators satisfy axioms D (Consistency) and 5 (Negative Introspection) and the *common* belief operator satisfies axiom **5** is equivalent to a (normal) logic where the individual belief operators satisfy axiom 5 and, at the interpersonal level, axiom **C*** is satisfied. (It is worth noting that axiom 4 – Positive Introspection – plays no role whatsoever in all the results proved in this section.)

3. Intersubjective consistency of beliefs

In this concluding section we shall discuss the relative strength of axioms T ($\boxed{i}\, A \rightarrow A$), **C** ($\boxed{i}\, A \rightarrow \neg\boxed{j}\,\neg A$) and C^* ($\boxed{i}\,\Box\, A \rightarrow \neg\boxed{j}\,\neg A$). The connections among these axioms and their internal structure are much clarified by relating them to four conditions on intersubjective beliefs implied by the Truth Axiom T (at the individual level); these conditions might also prove quite valuable in future research on related matters. T and C prove to be significantly stronger than $\mathbf{C}^*$; in particular, they imply some "agreement" among individuals, which plays no role in C^*

(respectively 5^*). The four implications of the Truth Axiom T for intersubjective beliefs are the following:[7]

TN. $\quad [i][j]A \rightarrow [j]A$

TP. $\quad [i]\langle j\rangle A \rightarrow \langle j\rangle A$

IN. $\quad [i][j]A \rightarrow [i]A$

IP $\quad \langle i\rangle [j]A \rightarrow \langle i\rangle A.$

The axioms come in two natural pairs, TN – TP and IN – IP. The T-axioms are simply instances of the Truth Axiom T: truth conditions on individuals' beliefs ***about others' beliefs.*** The I-axioms, on the other hand, are "internal" conditions on individual belief systems relating beliefs about the world to beliefs about other agents' beliefs. IN, for instance, forbids agents to knowingly disagree. IN and IP say, essentially, that individuals take ***others*** to know something whenever they believe it. This interpretation corresponds to the syntactical fact that the I-axioms derive from T not simply as instances but as implications based on the inference rule of Necessitation (as well as axiom schema K). Thus the I-axioms reflect not so much T "per se", but individuals' shared/common knowledge that T.

---

[7] We use the notation $\langle i\rangle$ **A** as a short-hand for $\neg[i]\neg$**A**. Recall that, semantically, $\Box$ **A** is true at world a if and only if *for **all*** β such that $\alpha R_i \beta$, **A** is true at β. It is easy to see that, on the other hand, $\langle i\rangle$ **A** is true at a if and only if *there* exists a β such that $\alpha R_i \beta$ and **A** is true at β.

Figure 5 below gives a complete picture of the implication relation among the seven presented axioms considered in isolation. The arcs are labeled by the assumptions on the individual belief operators necessary to establish a particular implication; if an arc is absent, no implication holds, even with the strongest logic for individual belief, namely KD45.

We also have included a single implication of one axiom by others, namely that of C by the conjunction of IP and TP. In the transitive case (that is, when axiom 4 is satisfied), this yields in fact a characterization of C.

Insert Figure 5

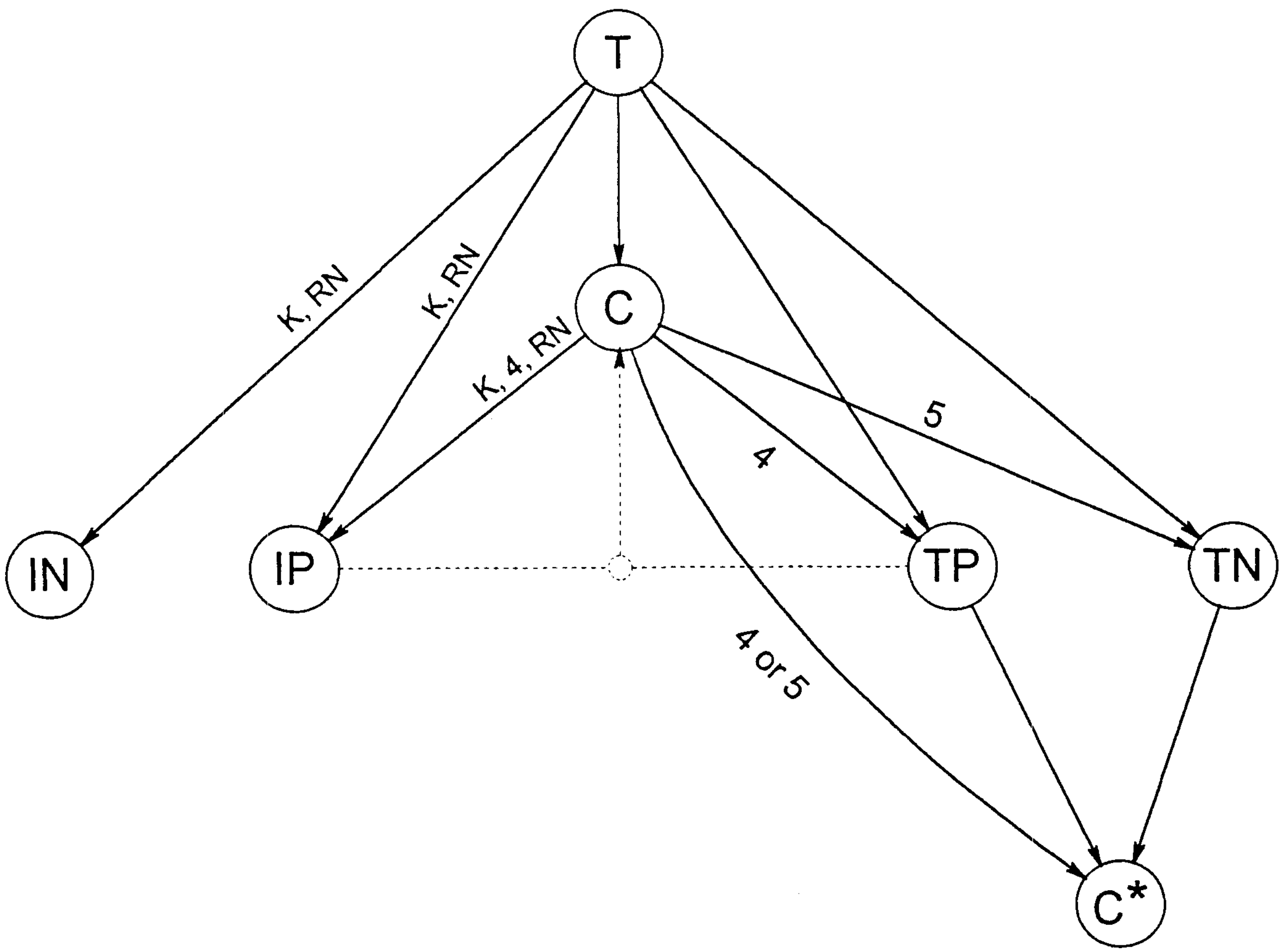

FIGURE 5

PROPOSITION 3. The following holds:

(i) IP and TP together imply C,

(ii) C and 4 together imply IP and TP,

(iii) C and 5 together imply TN.

The proof of Proposition **3** is given in Appendix 2.

Thus C has strong implications for truth of intersubjective beliefs; moreover, it crucially involves some intersubjective agreement. This is underlined by the example of Figure 2 above, which shows that intersubjective truth alone (TP and TN) fails to guarantee Compatibility since it fails to imply any intersubjective agreement (IP or IN). Note that the frame of Figure 2 satisfies the property of Proposition 2 and therefore it validates axiom C* (it also validates 5*, since the transitive closure of $R_1 \cup R_2$ is euclidean). In order to expand on the example of Figure 2 we need the following lemma, which is proved in Appendix 2.[8]

LEMMA 3. Axiom schemata TN and TP are valid in the class of frames that satisfy the following properties:

(1) R is serial for every i,

(2) $\forall i, \forall j, \forall \alpha, \forall \beta, \forall \gamma, \quad \alpha R_i \beta \ \& \ \beta R_j \gamma \Rightarrow \alpha R_j \gamma$.

[8] Note that Lemma **3** gives only a soundness result, not a characterization result.

The frame of Figure 2 satisfies the properties of Lemma **3** and therefore it validates both TP and TN. On the other hand, it falsifies both IP and IN. This can be verified easily by considering a model based on that frame where a sentence p is true at a and false at β. ***Indeed, one then has that it is common belief that I and 2 disagree (they "agree to disagree"):*** the formula $\boxed{*}\left(\boxed{1}\neg(p) \text{ A } \boxed{2}(p)\right)$ is true at every world.

Conversely, the example of Figure 6 shows that intersubjective agreement has nothing to do with intersubjective truth, hence, ***a fortiori,*** it has nothing to do with C^* (respectively, 5^*).

Insert Figure 6

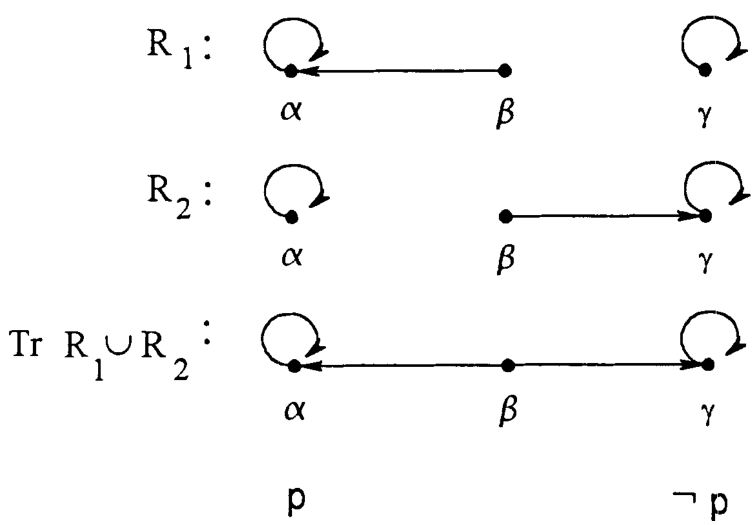

FIGURE *6*

Figure 6 can be used to show that it is possible for IP and IN to be valid while, at the same time, TN, TP, C* and 5^* are falsified. Consider a model based on the frame of Figure *6* where a given sentence p is true at α and false at β. Then at α both $\boxed{2}(p)$ and $\langle 2 \rangle(p)$ are true. Hence at β both $\boxed{1}\boxed{2}(p)$ and $\boxed{1}\langle 2 \rangle(p)$ are true. On the other hand, at β $\boxed{2}\neg(p)$ is true. Hence both the following formulas are *false* at β : $\boxed{1}\boxed{2}(p) \rightarrow \boxed{2}(p)$ (which is an instance of TN) and $\boxed{1}\langle 2 \rangle(p) \rightarrow \langle 2 \rangle(p)$ (which is an instance of TP). Furthermore, since $\boxed{*}(p)$ is true at α, at β the following formula is true: $\boxed{1}\boxed{*}(p)$. Since at β $\boxed{2}\neg(p)$, we have that at β the formula $\boxed{1}\boxed{*}(p) \rightarrow \neg\boxed{2}\neg(p)$ (which is an instance of C^*) is *false*. Finally, at β the formula $\neg\boxed{*}(p)$ is true (since at γ it is true that $\boxed{*}\neg(p)$). On the other hand, at β the formula $\boxed{*}\neg\boxed{*}(p)$ is false (since at α it is true that $\boxed{*}(p)$). It follows that at β the following formula (which is an instance of 5^*) is *false*: $\neg\boxed{*}(p) \rightarrow \boxed{*}\neg\boxed{*}(p)$. It only remains to show that IN and IP are valid in the frame of Figure 6. This is established in the following lemma, which is proved in Appendix 2. (The frame of Figure 6 satisfies the property of Lemma 4.)[9]

LEMMA 4. Axiom schemata IN and IP are valid in the class of frames that satisfy the following property:

$$\forall i,\ \forall j,\ \forall \alpha,\ \exists \beta \text{ such that } \left(\alpha R_i \beta \text{ and, } \forall \gamma,\ \text{if } \alpha R_i \gamma \text{ then } \beta R_j \gamma\right)$$

[9] Note that Lemma 4 gives only a soundness result, not a characterization result.

C^* then turns out to weaken C in three ways. Firstly, it disposes of all internal requirements. Secondly, it weakens truth of beliefs to "limited falsehood" ($\boxed{i}\,\boxed{j}\, A \rightarrow \Diamond_{i}$ A). Thirdly it allows intersubjective beliefs to be still further from the truth by imposing restrictions only on beliefs about others' common beliefs. How much weaker C^* is than C is illustrated in the example of Figure 7. We want to show, with the aid of Figure 7, that it is possible for C^* and 5^* to be valid, while at the same time TN, TP, IN and IP are simultaneously falsified. First of all, the transitive closure of $R_1 \cup R_2 \cup R_3$ is euclidean and satisfies the property of Proposition 2 (indeed it is the universal relation on $\{\alpha,\beta,\gamma\}$). Thus the frame of Figure 7 validates both 5^* and C^*. Now we show that IP, IN, TP and TN can be simultaneously falsified. Consider a model based on the frame of Figure 7 in which a sentence p is true at both α and β and false at γ. Then the following facts are easy to verify:

(1) $\boxed{1}\,\boxed{2}\,\neg(p) \rightarrow \boxed{2}\,\neg(p)$ (which is an instance of TN) is *false* at α,

(2) $\boxed{1}\,\Diamond_{2}\,\neg(p) \rightarrow \Diamond_{2}\,\neg(p)$ (which is an instance of TP) is *false* at α,

(3) $\boxed{1}\,\boxed{2}\,\neg(p) \rightarrow \boxed{1}\,\neg(p)$ (which is an instance of IN) is *false* at α,

(3) $\Diamond_{1}\,\boxed{2}\,\neg(p) \rightarrow \Diamond_{1}\,\neg(p)$ (which is an instance of IP) is *false* at α.

Insert Figure 7

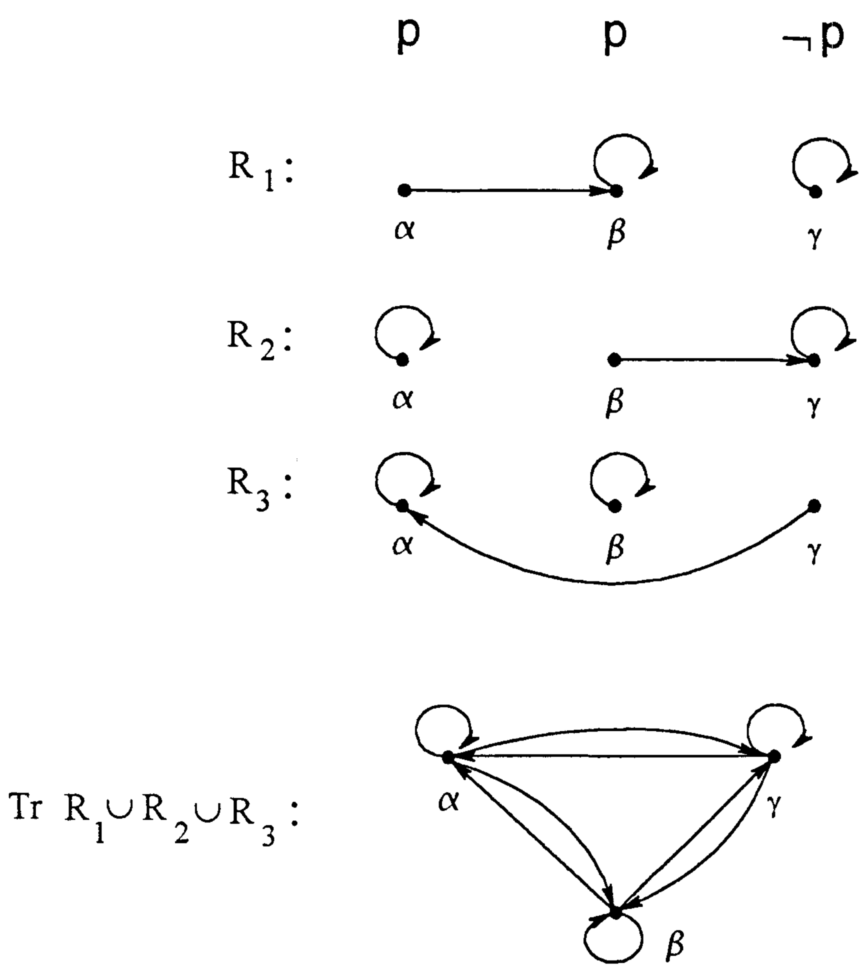

FIGURE 7

Appendix 1

In this appendix we review the axiomatic characterization of common belief (for more details see Bonanno, 1994, Halpem and Moses, 1992, Lismont, 1993, Lismont and Mongin, 1994). Given n individuais, let $\boxed{i}$ be the belief operator of individual $i = 1, \ldots, n$ and $\Box$ the common belief operator. Consider the logic, call it the *CB* logic, defined by the following axioms and rules of inference:

AXIOM SCHEMATA	RULES OF INFERENCE
(I) all the tautologies	(1) Modus Ponens: $\dfrac{A, A \to B}{B}$
(2) for every $j \in (1, \ldots, n, *\}$, axiom schema K: $\boxed{j}(A \to B) \wedge \boxed{j}A \to \boxed{j}B$	(2) Necessitation: for all $j \in \{1, \ldots, n, *\}$, $\dfrac{A}{\boxed{j}A}$
(3) for all $i \in \{1, \ldots, n\}$ axiom schema SB: $\boxed{*}A \to \boxed{i}A$	**(3)** Truism (or RJ): $\dfrac{A \to \boxed{1}A \wedge \ldots \wedge \boxed{n}A}{\boxed{1}A \wedge \ldots \wedge \boxed{n}A \to \boxed{*}A}$
(4) for all $i \in \{1, \ldots, n\}$ axiom schema PR: $\boxed{*}A \to \boxed{i}\boxed{*}A$	

The semantics of common belief is as follows. A ***standard*** *n-frame* is an (n+1)-tuple

$\langle W, R_1, \ldots, R_n \rangle$ where:

(1) W is a non-empty set whose members are called "possible worlds", or simply "worlds" and are denoted by a, β, y, ...,

(2) For every $i \in \{1,\ldots,n\}$ R_i is a (possibly empty) binary "accessibility" relation on W

A *standard n-model* is an (n+2)-tuple $\mathcal{M} = \langle W, R_1, \ldots, R_n, F\rangle$, where $\langle W, R_1, \ldots, R_n\rangle$ is a standard n-frame and $F : S \rightarrow 2^W$ is a function from the set of sentence letters S into the set of subsets of W. We say that $\mathcal{M}$ is *based on* the frame $\langle W, R_1, \ldots, R_n\rangle$.

Given a formula A and a standard n-model $\mathcal{M} = \langle W, R_1, \ldots, R_n, F\rangle$, the *truth set* of A in $\mathcal{M}$, denoted by $\|A\|^{\mathcal{M}}$, is defined recursively as follows:

(1) If A = (p) where p is a sentence letter, then $\|A\|^{\mathcal{M}} = F(p)$,

(2) $\|\neg A\|^{\mathcal{M}} = W - \|A\|^{\mathcal{M}}$ (that is, $\|\neg A\|^{\mathcal{M}}$ is the complement of $\|A\|^{\mathcal{M}}$)

(3) $\|A \vee B\|^{\mathcal{M}} = \|A\|^{\mathcal{M}} \cup \|B\|^{\mathcal{M}}$,

(4) For all i = 1, ..., n, $\|\boxed{i}A\|^{\mathcal{M}} = \{\alpha \in W : \text{for all } \beta \text{ such that } \alpha R_i \beta,\ \beta \in \|A\|^{\mathcal{M}}\}$,

(5) $\|\boxed{*}A\|^{\mathcal{M}} = \{\alpha \in W : \text{for all } \beta \text{ such that } \alpha R_* \beta,\ \beta \in \|A\|^{\mathcal{M}}\}$,

where R_* is the transitive closure of $R_1 \cup \ldots \cup R_n$ (see footnote 5). If $\alpha \in \|A\|^{\mathcal{M}}$ we say that *A is true at world α in model* $\mathcal{M}$. An alternative notation for $\alpha \in \|A\|^{\mathcal{M}}$ is $\models_{\alpha}^{\mathcal{M}} A$ and an alternative notation for $\alpha \notin \|A\|^{\mathcal{M}}$ is $\not\models_{\alpha}^{\mathcal{M}} A$.

A formula A is *valid in model* $\mathcal{M} = \langle W, R_1, \ldots, R_n, F\rangle$ if and only if $\|A\|^{\mathcal{M}} = W$, that is, if and only if $\models_{\alpha}^{\mathcal{M}} A$ for all $\alpha \in W$

Halpern and Moses (1992) and Lismont (1993) proved the following *completeness theorem:* If A is a theorem of the logic CB, then A is valid in every standard n-model; conversely, if A is a formula that is valid in every standard n-model, then **A** is a theorem of the logic CB.

It follows from the characterization of axiom 4 (Positive Introspection: see Section 2) that the following is a theorem of the logic CB: $\boxed{*}A \rightarrow \boxed{*}\boxed{*}A$.

Appendix 2

In this appendix we prove Propositions 1, 2, 3 and Lemmas 3 and 4, as well as a few extra results. First we need to recall some definitions and notation from modal logic (cf. Chellas, 1980).

We say that an axiom schema S is *characterized* by the class $\mathcal{B}$ of standard n-frames (for a definition of standard n-frames see Appendix 1) if and only if

(1) every instance of S is valid in every model based on a Frame in $\mathcal{B}$, and

(2) if $\langle W, R_1, \ldots, R_n \rangle$ is a frame that does *not* belong to $\mathcal{B}$ then there is a model $\mathcal{M} = \langle W, R_1, \ldots, R_n, F \rangle$ based on it and an instance A of S such that A is not valid in $\mathcal{M}$ (that is, for some world α in $\mathcal{M}$, $\not\models_{\alpha}^{\mathcal{M}}$ A).

We now prove Proposition 1, which states that axiom schema C: $\boxed{i} A \rightarrow \neg \boxed{j} \neg A$ is characterized by the class of standard frames that satisfy the *Compatibility* property:

$$\forall i, \forall j, \forall \alpha, \exists \beta \text{ such that } \alpha R_i \beta \text{ and } \alpha R_j \beta.$$

PROOF OF PROPOSITION 1. (1) Let $\langle W, R_1, \ldots, R_n \rangle$ be a frame that satisfies the Compatibility property. Let $\mathcal{M}$ be a model based on this Frame. Fix arbitrary i, j and α and an arbitrary formula A. Suppose that $\models_{\alpha}^{\mathcal{M}} \boxed{i}$ **A.** Then $\models_{\gamma}^{\mathcal{M}}$ **A** for all γ such that $\alpha R_i \gamma$. By Compatibility, there exists a β such that $\alpha R_i \beta$ and $\alpha R_j \beta$. Hence $\not\models_{\alpha}^{\mathcal{M}} \boxed{j} \neg A$, that is, $\models_{\alpha}^{\mathcal{M}} \neg \boxed{j} \neg A$.

(2) Let $\langle W, R_1, ..., R_n\rangle$ be a frame that violates the Compatibility property. Then there exist i, j and α such that, for no β, $\alpha R_i\beta$ and $\alpha R_j\beta$. Three cases are possible: (2.1) there is no world which is R_j-accessible from α, (2.2) there is no world which is R_i-accessible from α, (2.3) there are worlds R_i-accessible from α and there are worlds 5-accessible from α but no world is both R_i-accessible and 5-accessible from α. In case (2.1) choose an arbitrary model $\mathcal{M}$ based on this frame. Then, for every formula A, $\models^{\mathcal{M}}_{\alpha} \boxed{j} A$ (see Chellas, p. 77). Let B be a tautology. Then $\models^{\mathcal{M}}_{\alpha} \boxed{j} \neg B$. Whether or not there are worlds that are R_i-accessible from α, it must be $\models^{\mathcal{M}}_{\alpha} \boxed{i} B$. Thus $\not\models^{\mathcal{M}}_{\alpha} (\boxed{i} B \rightarrow \neg\boxed{j} \neg B)$. Case (2.2.) is dealt with in a similar way. Finally, consider case (2.3). Let $\Gamma_i = \{\gamma \in W \mid \alpha R_i\gamma\}$ and $\Gamma_j = \{\gamma \in W \mid \alpha R_j\gamma\}$. Then $\Gamma_i \neq \emptyset, \Gamma_j \neq \emptyset$ and $\Gamma_i \cap \Gamma_j = \emptyset$. Let p be a sentence letter and $\mathcal{M}$ be a model based on this frame such that F(p) $= \|(p)\|^{\mathcal{M}} = \Gamma_i$. Then $\models^{\mathcal{M}}_{\alpha} \boxed{i}(p)$ and $\models^{\mathcal{M}}_{\alpha} \boxed{j} \neg(p)$. Thus $\not\models^{\mathcal{M}}_{\alpha} (\boxed{i}(p) \rightarrow \neg\boxed{j}\neg(p))$. ■

PROOF OF PROPOSITION 2. (1) Let $\langle W, R_1, ..., R_n\rangle$ be a frame that satisfies the $*$-Compatibility property. Let $\mathcal{M}$ be a model based on this frame. Fix arbitrary i, j and α and an arbitrary formula A. Suppose that $\models^{\mathcal{M}}_{\alpha} \boxed{i}\boxed{*} A$. Then by $*$-Compatibility there exist β and γ such that $\alpha R_i \beta$, $\beta R_* \gamma$ and $\alpha R_j \gamma$. It follows from $\alpha R_i \beta$ that $\models^{\mathcal{M}}_{\beta} \boxed{*} A$ and from the fact that $\beta R_* \gamma$ that $\models^{\mathcal{M}}_{\gamma} A$. Since $\alpha R_j \gamma$, $\models^{\mathcal{M}}_{\alpha} \neg\boxed{j}\neg A$.

(2) Let $\langle W, R_1, ..., R_n\rangle$ be a frame that violates the $*$-Compatibility property. Then there exist i, j and α such that, for all β and for all γ, if $\alpha R_i \beta$ and $\beta R_* \gamma$ then not $\alpha R_j \beta$. Let p be a

sentence letter and let $\mathcal{M}$ be a model based on this frame such that $F(p) = \|(p)\|^{\mathcal{M}} = \Gamma =$ $\{\gamma \in W : \exists \beta \in W \text{ with } \alpha R_i \beta \text{ and } \beta R_* \gamma\}$. Let $B = \{\beta \in W : \alpha R_i \beta\}$. Suppose first that $B = \mathbf{O}$. Then also $\Gamma = \varnothing$. Since $B = \varnothing$, for every formula A, $\models^{\mathcal{M}}_{\alpha} \boxed{i} A$, in particular for $A = \boxed{*}(p)$. Thus $\models^{\mathcal{M}}_{\alpha} \boxed{i}\boxed{*}(p)$. On the other hand, since $\Gamma = \mathbf{O}$, $\models^{\mathcal{M}}_{\alpha} \boxed{j}\neg(p)$. It follows that $\not\models^{\mathcal{M}}_{\alpha} (\boxed{i}\boxed{*}(p) \rightarrow \neg\boxed{j}\neg(p))$. Consider now the case where $B \neq \mathbf{O}$. Fix an arbitrary β such that

$\alpha R_i \beta$. Then $\models^{\mathcal{M}}_{\beta} \boxed{*}(p)$ (this is true, trivially, in the case where there are no worlds that are R_*-accessible from β, and, by construction, also in the case where there are worlds that are R_*-accessible from β, because every such world belongs to Γ). Hence $\models^{\mathcal{M}}_{\alpha} \boxed{i}\boxed{*}(p)$. On the other hand, by hypothesis, for every $\gamma \in \Gamma$, it is not the case that $\alpha R_j \gamma$. Hence $\models^{\mathcal{M}}_{\alpha} \boxed{j}\neg p$. It follows that $\not\models^{\mathcal{M}}_{\alpha} (\boxed{i}\boxed{*}(p) \rightarrow \neg\boxed{j}\neg(p))$. ■

PROOF OF PROPOSITION 3. (i) We want to show that C is a theorem of every system that contains axiom schemata IP and TP. The proof goes as follows (PL stands for "Propositional Logic"):

1.	$\langle j \rangle \boxed{j}\neg A \rightarrow \neg\boxed{i} A$	(IP)
2.	$\boxed{i} A \rightarrow \boxed{i}\langle j \rangle A$	(1, PL)
3.	$\boxed{i}\langle j \rangle A \rightarrow \langle j \rangle A$	(TP)

4. $\boxed{i} A \rightarrow \langle i \rangle A$ (2, 3, PL)

(ii) First we show that IP is a theorem of every normal system containing axiom 4 at the individual level (that is, for every individual) and axiom C at the interpersonal level. The proof is as follows (RN stand for "Rule of Necessitation", MP for "Modus Ponens", PL for "Propositional Logic"). The proof is as follows:

1. $\boxed{i}\neg A \rightarrow \neg\boxed{j} A$ (axiom C)
2. $\boxed{i}(\boxed{i}\neg A \rightarrow \neg\boxed{j} A)$ (1, RN for i)
3. $\boxed{i}(\boxed{i}\neg A \rightarrow \neg\boxed{j} A) \rightarrow (\boxed{i}\boxed{i}\neg A \rightarrow \boxed{i}\neg\boxed{j} A)$ (axiom K for i)
4. $\boxed{i}\boxed{i}\neg A \rightarrow \boxed{i}\neg\boxed{j} A$ (2, 3, MP)
5. $\boxed{i}\neg A \rightarrow \boxed{i}\boxed{i}\neg A$ (axiom 4 for i)
6. $\boxed{i}\neg A \rightarrow \boxed{i}\neg\boxed{j} A$ ((4, 5, PL)
7. $\langle i \rangle\boxed{j} A \rightarrow \langle i \rangle A$ (6, PL).

Next we show that TP is a theorem of every normal system containing axiom 4 at the individual level and axiom C at the interpersonal level:

1. $\boxed{j}\neg A \rightarrow \boxed{j}\boxed{j}\neg A$ (axiom **4** for j)
2. $\boxed{j}\boxed{j}\neg A \rightarrow \neg\boxed{i}\neg\boxed{j}\neg A$ (axiom C)
3. $\boxed{j}\neg A \rightarrow \neg\boxed{i}\neg\boxed{j}\neg A$ (1,2 , PL)
4. $\boxed{i}\langle j \rangle A \rightarrow \langle j \rangle A$ (3, PL).

(iii) We want to show that TN is a theorem of every normal system containing axiom 5 at the individual level and axiom C at the interpersonal level. The proof is as follows:

1.	$\neg\boxed{j}A \rightarrow \boxed{j}\neg\boxed{j}A$	(axiom 5 for j)
2.	$\boxed{j}\neg\boxed{j}A \rightarrow \neg\boxed{i}\boxed{j}A$	(axiom C)
3.	$\neg\boxed{j}A \rightarrow \neg\boxed{i}\boxed{j}A$	(1,2 , PL)
4.	$\boxed{i}\boxed{j}A \rightarrow \boxed{j}A$	(3, PL). ■

PROOF OF LEMMA **3.** (1) First we prove that axiom schema TN is valid in the class of frames that satisfy the two properties of Lemma **3.** Fix an arbitrary model based on a frame that satisfies those properties. Fix arbitrary i, j, a and an arbitrary formula **A.** Suppose that $\models^{m}_{a}\boxed{i}\boxed{j}$**A.**

By seriality of R_i there exists a β such that $\alpha R_i\beta$. Fix an arbitrary such β. Then $\models^{m}_{p}$ **q A. By** seriality of R„ there exists a y such that $\beta R_j\gamma$. Fix an arbitrary such y. Then $\models^{m}_{\gamma}$ **A.** By property **(2)** $\alpha R_j\gamma$. Hence $\models^{m}_{a}\boxed{j}$ **A.**

(2) Now we turn to axioms schema TP. Fix an arbitrary model based on a frame that satisfies the two properties of Lemma **3.** Fix arbitrary i, j, α and an arbitrary formula **A.** Suppose that $\models^{m}_{a}\boxed{i}\langle j\rangle$'**A.** By seriality of R_i there exists a β such that $\alpha R_i\beta$. Then $\models^{m}_{p}\langle j\rangle$ **A.** Hence there exists a γ such that $\beta R_j\gamma$. and $\models^{m}_{\gamma}$ A. By property (2) $\alpha R_j\gamma$. Hence $\models^{m}_{\alpha}\langle j\rangle$ **A.** ■

PROOF OF LEMMA 4. (1) First we prove validity of IN. Fix an arbitrary model based on a frame that satisfies the property of Lemma 4. Fix arbitrary i, j, α and an arbitrary formula **A.** Suppose that $\vDash^{\mathcal{M}}_{\alpha} \Box\Box$ A. Then, by the property, there exists a β such that $\alpha R_i \beta$. Choose an arbitrary such β. Then $\vDash^{\mathcal{M}}_{\beta} \Box$ **A.** Choose an arbitrary γ such that $\alpha R_i \gamma$. By the assumed property, $\beta R_j \gamma$. Hence (since $\vDash^{\mathcal{M}}_{\beta} \Box$ **A)** $\vDash^{\mathcal{M}}_{\gamma}$ **A.** Therefore $\vDash^{\mathcal{M}}_{\alpha} \Box$ A.

(2) Now we prove validity of IP. Fix an arbitrary model based on a frame that satisfies the property of Lemma 4. Fix arbitrary i, j, α and an arbitrary formula **A.** Suppose that $\vDash^{\mathcal{M}}_{\alpha} \Diamond\Box$ A.

Then there exists a β such that $\alpha R_i \beta$ and $\vDash^{\mathcal{M}}_{\beta} \boxed{j}$ A. By the assumed property (choosing $\gamma = \beta$), $\beta R_j \beta$. Hence $\vDash^{\mathcal{M}}_{\beta}$ A. Therefore, $\vDash^{\mathcal{M}}_{\alpha} \Diamond$ A. ■

We conclude this Appendix by proving two claims made in Section 2.

CLAIM 1. Axiom schema 5^{*} is equivalent to axiom schema P*: $\neg\boxed{*}\,\mathbf{A} \rightarrow \boxed{i}\,\neg\boxed{*}\,\mathbf{A}$.

Proof (1) First we prove that 5^{*} implies P*:

1. $\neg\boxed{*}\,A \rightarrow \boxed{*}\,\neg\boxed{*}\,A$ (5^{*})

2. $\boxed{*}\,\neg\boxed{*}\,A \rightarrow \boxed{i}\,\neg\boxed{*}\,A$ (SB: see Appendix 1)

3. $\neg\boxed{*}\,A \rightarrow \boxed{i}\,\neg\boxed{*}\,A$ (1, 2, PL).

(2) Next we prove that P* implies 5*.

1. $\neg\boxed{*}A \rightarrow \boxed{i}\neg\boxed{*}A$ (P*)

....

n. $\neg\boxed{*}A \rightarrow \boxed{n}\neg\boxed{*}A$ (P*)

n+1. $\neg\boxed{*}A \rightarrow \left(\boxed{1}\neg\boxed{*}A \text{ A } \ldots \text{ A } \boxed{n}\neg\boxed{*}A)\right)$ (1, ..., n, PL)

n+2. $\left(\boxed{1}\neg\boxed{*}A \text{ A } \ldots \text{ A } \boxed{n}\neg\boxed{*}A)\right) \rightarrow \boxed{*}\neg\boxed{*}A$ (n+1, Truism: see Appendix 1)

n+3. $\neg\boxed{*}A \rightarrow \boxed{*}\neg\boxed{*}A$ ((n+1, n+2, PL). ■

CLAIM **2.** If individual beliefs satisfy axiom D (consistency) then the following is a theorem of the logic CB: $\boxed{i}\neg\boxed{*}\mathbf{A} \rightarrow \neg\boxed{*}\mathbf{A}$ (which is equivalent to: $\boxed{*}\mathbf{A} \rightarrow \neg\boxed{i}\neg\boxed{*}\mathbf{A}$**).**

Proof.

1. $\boxed{*}A \rightarrow \boxed{i}\boxed{*}A$ (PR: see Appendix 1)

2. $\boxed{i}\boxed{*}A \rightarrow \neg\boxed{i}\neg\boxed{*}A$ (D for i)

3. $\boxed{*}A \rightarrow \neg\boxed{i}\neg\boxed{*}A$ (1, **2,** PL). ■

References

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

Bacharach, M. (1985), Some extensions of a claim of Aumann in an axiomatic model of knowledge, Journal of Economic Theory, 37, 167-190.

Binmore, K. and A. Brandenburger (1990), Common knowledge and game theory, in: K. Binmore, Essays on the foundations of game theory, Basil Blackwell, Cambridge, MA.

Bonanno, G. (1994), The logic of shared belief, public rumor, public shared belief and common belief, Working Paper #94-10, Department of Economics, University of California, Davis.

Brandenburger, A. and E. Dekel (1987), Common knowledge with probability 1, Journal of Matheniatical Economics, 16, 237-246.

Chellas, B. (1980), Modal logic, Cambridge University Press, Cambridge.

Colombetti, M. (1993), Formal semantics for mutual beliefs, *Artificial* intelligence, 62, 341-353.

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

Geanakoplos, J. and H. Polemarchakis (1982), We can't disagree forever, Journal of Economic Theory, 28, 192-200.

Halpern, J. (1986), Reasoning about knowledge: an overview, in J. Halpern (Ed.), Theoretical aspects of reasoning about *knowledge*, Morgan Kaufmann, Los Altos, CA, 1-17.

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

Hintikka, J. (1962), Knowledge *and* belief, Cornell University Press, Ithaca (NY).

Kaneko, M. and T. Nagashima (1993), Indefinability of the common knowledge concept in finitary logics, Working Paper No. 551, Institut of Socio-Economic Planning, University of Tsukuba.

Lewis, D. (1969), *Convention: a philosophical study*, Harvard University Press, Cambridge (MA).

Lismont, L. (1993), La connaissance commune en logique modale, *Mathematical Logic* Quarterly, 39, 115-130.

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

Milgrom, P. (1981), An axiomatic characterization of common knowledge, *Econometrica*, 49, 219-222.

Monderer, D. and D. Samet (1989), Approximating common knowledge with common belief, Games and Economic Behavior, 1, 170-190.

Rubinstein, .A. and A. Wolinsky (1990), On the logic of "Agreeing to Disagree" type of results, *Journal of* Economic *Theory*, 51, 184-193.

Samet, D. (I 990), Ignoring ignorance and agreeing to disagree, Journal *of* Economic Theory, 52, 190-207.

Shin, H. (1993), Logical structure of common knowledge, Journal of Economic Theory, 60, 1-13.

Tan, T. and S. Werlang (1985), On Aumann's notion of common knowledge: an alternative approach, mimeo, Department of Economics, Princeton University.