INTERSUBJECTIVE CONSISTENCY OF KNOWLEDGE AND BELIEF

Giacomo Bonanno

and

Klaus Nehring

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

e-mail: gfbonanno@ucdavis.edu
kdnehring@ucdavis.edu

February, 1998

Abstract

We consider interactive epistemic models where individuals are described by both their "knowledge" and their "beliefs". Three intersubjective consistency conditions are examined: Intersubjective Caution (if an individual believes something to be common belief then he knows it to be common belief), Truth of Common Belief (only true facts are commonly believed) and Qualitative Agreement. These conditions are employed in characterizations of the following properties which describe either the extent of intersubjective truth and/or the logic of common belief: common belief in no error, common knowledge of common belief, negative introspection of common belief, coincidence of common knowledge and common belief, and collapse of individual belief and knowledge. We also discuss to what extent the three fundamental conditions can be viewed as intersubjective rationality conditions.

1. Introduction

We consider interactive epistemic models where individuals are described by both their "knowledge" and their "beliefs". Knowledge is distinguished from belief by its higher degree of subjective "certainty" (whatever is known is also believed, but not vice versa) and of veridicality. In an intersubjective context, imposing the truth axiom on knowledge is a strong assumption: not only are individuals never mistaken in what they know, but also this very fact is common knowledge among them; this implies, in particular, that agent i must know event E whenever he knows that some other agent j knows E.[1] Such "common knowledge of no error of knowledge" is a prominent example of an *intersubjective consistency condition* on epistemic states, and plays a prominent role in the foundations of game theory (cf. Ben Porath, 1997, Morris, 1994, Stalnaker, 1994, 1996, Stuart, 1997, Bonanno and Nehring, 1996, 1997ab). While this assumption might be considered plausible for beliefs with the highest epistemic commitment (knowledge), the distinguishing feature of the notion of belief proper is precisely the possibility of error. In particular, individuals may come to believe other individuals to have mistaken beliefs. In this paper, we will investigate various weaker intersubjective consistency properties and relate "macro" properties of the interactive system (properties of the common belief and common knowledge operators) to "micro" properties of individual belief hierarchies. It will be shown that the assumption of common belief in no error of *belief* can be derived from intersubjective consistency conditions which involve belief, knowledge and their interplay. In Section 5 we will also offer an evaluation of the extent to which the various conditions can be justified by some appeal to "intersubjective rationality".

Integrated epistemic systems that jointly consider knowledge and belief have been studied in philosophy (Hintikka, 1962, Lentzen, 1978), artificial intelligence and computer science (Halpern, 1991, van der Hoek, 1993, van der Hoek and Meyer, 1995, Kraus and

[1] In the technical part of the paper, we will in fact make use of a much weaker axiom, namely that individuals are correct in what they know to be common knowledge. For simplicity of exposition, however, in the introduction we shall assume that individual knowledge satisfies the Truth Axiom.

Lehmann, 1978), economics and game theory (Battigalli and Bonanno, 1997, Dekel and Gul, 1997, Geanakoplos, 1994). The philosophy and artificial intelligence literature has dealt mainly with single-agent systems and the focus has been on the tendency of belief to collapse into knowledge as the result of plausible-looking axioms. In game theory a study of systems of knowledge and belief arises naturally in the context of extensive form games from the attempt to model beliefs after counterfactual histories of play (at "unreached" information sets). Our work ties in with both literatures: as in the former, there is the possibility of a somewhat surprising collapse of belief into knowledge; the link to the latter is established by the above-mentioned fine-grained analysis of the assumption of common belief in no error (of belief) which plays a crucial role in the "construction" of the Common Prior Assumption under Incomplete Information (Bonanno and Nehring, 1996) and in the justification of backward induction for interesting classes of perfect information games (Ben Porath, 1997, Stalnaker, 1996, Stuart, 1997).

Emphasizing a "macro" rather than a "truth" perspective, one can also read this paper as an account of how and when common belief differs qualitatively from common knowledge. This is of interest particularly since common belief, in contrast to common knowledge, may exhibit epistemically counterintuive properties. The next section provides a road map of the paper by describing the specific questions that are asked and the results obtained (a visual summary is given in Figure 3). By focusing on very simple yet qualitatively contrasting examples, it is hoped that this section serves also the purpose of fleshing out the notions of common belief and common knowledge to readers only minimally acquainted with the growing literature on interactive epistemology.

2. Overview

§1. While common knowledge obeys the formal logic of agents' knowledge (semantically, common knowledge inherits the partitional structure from individual agents' knowledge[2]), common belief may fail to obey the formal logic of belief. In particular, the event that E is *not* commonly believed need not be itself commonly believed (technically, common belief may violate the axiom of Negative Introspection; cf. Colombetti, 1993, Lismont and Mongin, 1994, 1995), and – even more strikingly – whereas individuals always know what they believe (Transparency[3]), this is not necessarily so at the "common" level: it may well be that the agents don't commonly know what they commonly believe (we will say that Common Transparency is violated). To see this, consider the following example.

EXAMPLE 1. Individual 1 is a philosopher who knows the correct spelling of his name (Rabinowi*cz*). Individual 2 mistakenly believes that the spelling is Rabinowi*zc*. She even believes this spelling to be common belief between them. These beliefs are represented by state α in Figure 1, where the (thick) rectangles represent the knowledge partitions and the (thin) ellipses represent the belief sets.[4]

Insert Figure 1

[2] The notion of common knowledge was introduced by Lewis (1969) and Aumann (1976).

[3] Our terminology; there does not seem to be an established one.

[4] For more details on the graphical representation of knowledge and belief see the next section.

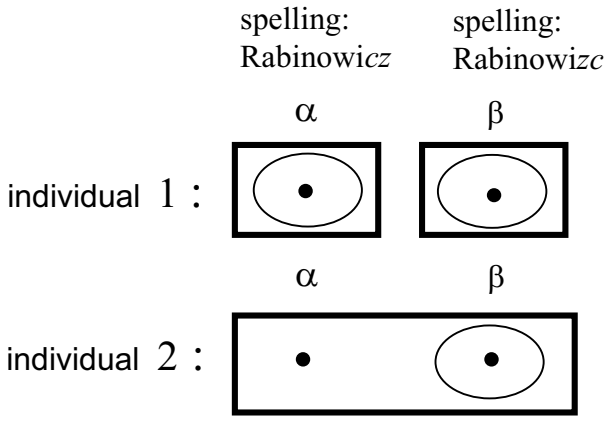

Figure 1

Let E be the event that represents the proposition "the spelling is Rabinowi*zc*", that is, E = {β}. Then, at state α, although E is not commonly believed (because individual 1 believes the spelling to be Rabinowi*cz*), it is not common belief that E is not commonly believed (because of individual 2's belief that E *is* common belief). Furthermore, at state β, E is commonly believed, but it is not common knowledge that it is commonly believed (because individual 2's knowledge set at β contains state α where E is *not* commonly believed).

Can one understand the properties of the common belief operator in terms of properties of individual beliefs? This question is answered by the first main result of the paper which provides a simultaneous characterization of the *public* (= commonly known) versions of these conditions in terms of a condition on individual belief hierarchies called *Intersubjective Caution* (Theorem 1). An agent is "intersubjectively cautious" (at some state) if, for any event E, he only *believes* E to be commonly believed if he in fact *knows* E to be commonly believed, that is, he does not open himself to the epistemic risk of being mistaken about what is commonly believed. Note that in Example 1, individual 2 fails to be intersubjectively cautious at either state.

EXAMPLE 2. In the modification of Example 1 illustrated in Figure 2, at state α individual 2 still mistakenly believes the spelling to be Rabinowi*zc*, but no longer believes this to be commonly believed: at state α she considers it possible (according to her beliefs) that the true state is β where individual 1 believes that the spelling is Rabinowi*cz*.

Insert Figure 2

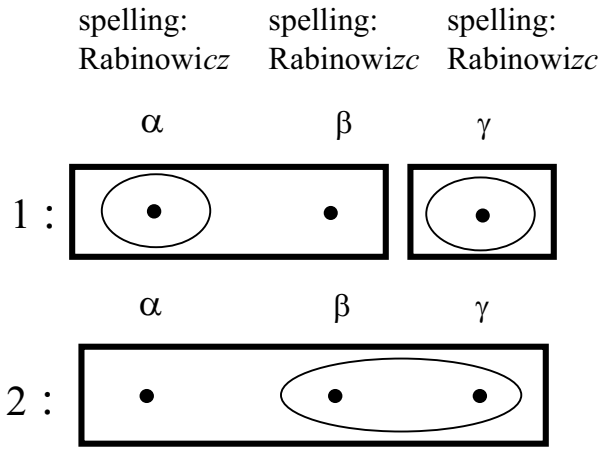

Figure 2

In Example 2 both agents are cautious at every state; as implied by Theorem 1, both Negative Introspection of Common Belief and Transparency are satisfied at every state. (Indeed, in Example 2, the only event that is commonly believed at any state is the universal event and common belief and common knowledge come to coincide.)

§2. Example 2 illustrates that it is quite easy for common knowledge and common belief to coincide even though individual knowledge and belief differ substantially. When exactly is this the case? First, such coincidence clearly requires Transparency at the common-level, because any event that is common knowledge is commonly known to be common knowledge. In addition – since any event that is commonly believed is commonly believed to be true – anything that is commonly believed must in fact be commonly *known* to be true. Intuitively, this

expresses agents' knowledge that the group can never be wrong *collectively* in their beliefs while they can be wrong individually. Building on Theorem 1, Theorem 2 asserts also the public converse, characterizing the equality of common belief and common knowledge in terms of public intersubjective caution plus public truth of common belief.[5]

§3. In Example 2, the agents "disagree" about state β: it is common belief that agent 1 believes for sure that β won't occur and that agent 2 believes that state β might occur (she is not certain of its non-occurrence). This opens the possibility of extreme forms of betting among the agents, even if they are risk-averse: agent 2 might promise agent 1 a high payoff if state β does not occur, in exchange for an extremely high payoff if it does occur; note that it is common knowledge among the two agents that such bets are acceptable. If agents' utility-functions are unbounded, then arbitrarily large gains from trade (measured in terms of expected utility) can be achieved through such bets. In Bonanno and Nehring (1997b) it has been shown that there are at most bounded gains from trade for "moderately risk-averse" individuals if and only if it is not commonly believed that there is no common belief in the truth of agents' beliefs (a property which we shall call Qualitative Agreement). In Theorem 4 we show that adding Qualitative Agreement to the assumptions of Theorem 2 leads to the collapse of belief into knowledge *for every individual*. One thus encounters an intersubjective version of the "collapse problem" known from the single-agent literature (Hintikka, 1962, Lenzen, 1978, van der Hoek, 1993, van der Hoek and Meyer, 1995). Here, it is resolved by reading Theorem 4 as follows: if agents are publicly intersubjectively cautious, and if "the group is always correct", then any gap between belief and knowledge results in disagreement. This reading suggests that the assumption of common knowledge of the truth of common belief is the least plausible.

[5] Theorem 2 is proved for the weakest possible system, where no veridicality assumptions of any sort are imposed on agents' knowledge.

§4. Finally, we explore the consequences of assuming only Qualitative Agreement and public Intersubjective Caution. Theorem 3 characterizes these as equivalent to common belief in the correctness of agents' beliefs plus Transparency. In view of the degeneracy uncovered by Theorem 4, these conditions arguably define the strongest plausible integrated intersubjective logic of knowledge and belief. Since Qualitative Agreement is motivated most compellingly via the absence of infinite gains from betting, it applies to belief rather than knowledge; as a result, Theorem 3 reveals that there is a sense in which the assumption of common belief in the correctness of agents' beliefs is easier to justify at the level of belief proper than at the level of knowledge.

Figure 3 contains a summary of the results proved in this paper.[6]

Insert Figure 3

§5. Section 5 concludes by providing an assessment of the three fundamental conditions as conditions of intersubjective rationality. In a nutshell, we argue in favor of the intersubjective rationality of Qualitative Agreement, in qualified favor for that of Intersubjective Caution, and in disfavor with respect to truth of common belief.

[6] For grater clarity some of the arrows in Figure 3 point only in one direction. However, all the results proved are full characterizations.

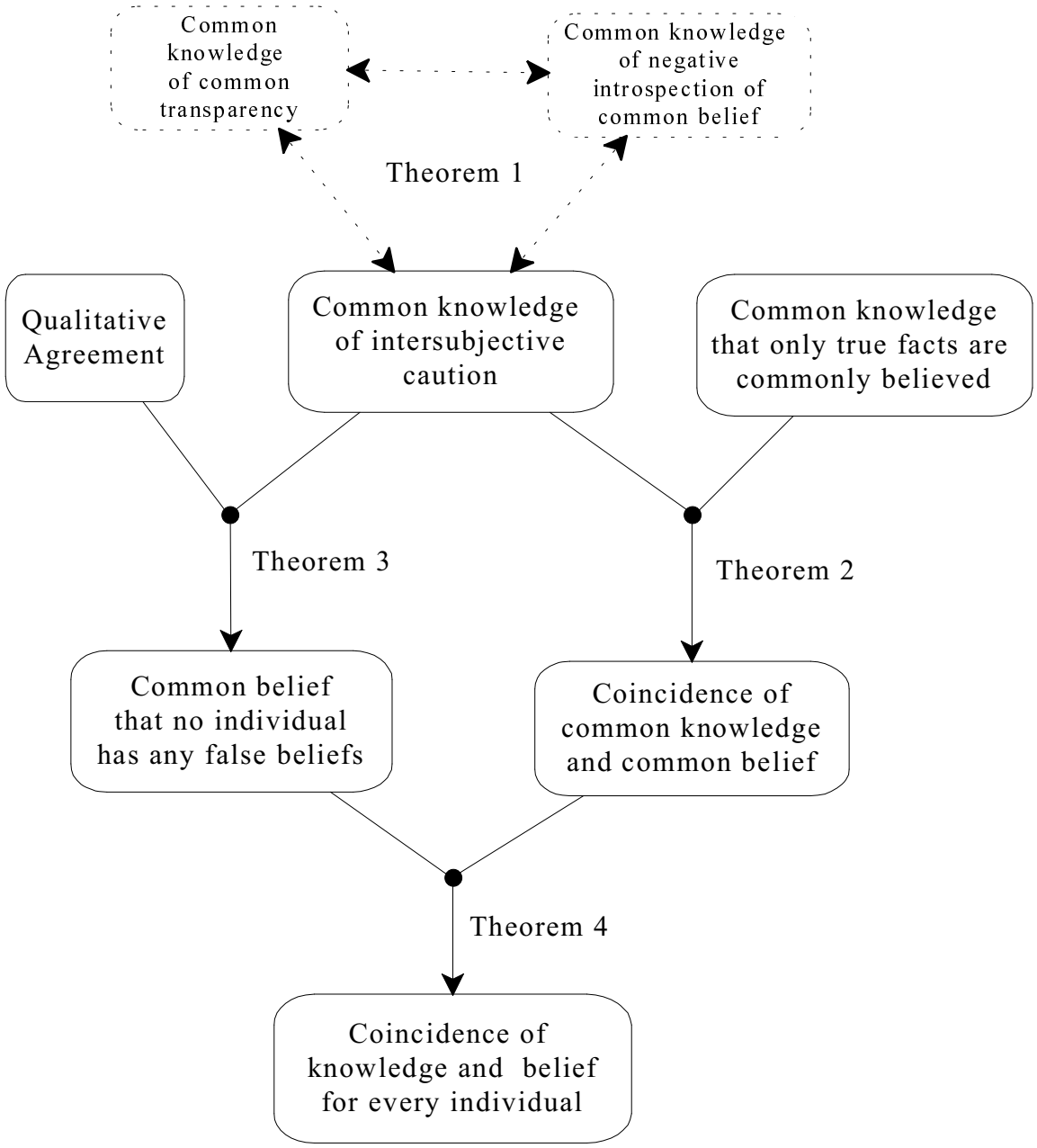
Common knowledge of common transparency
Common knowledge of negative introspection of common belief
Theorem 1
Qualitative Agreement
Common knowledge of intersubjective caution
Common knowledge that only true facts are commonly believed
Theorem 3
Theorem 2
Common belief that no individual has any false beliefs
Coincidence of common knowledge and common belief
Theorem 4
Coincidence of knowledge and belief for every individual

Figure 3

3. Interactive systems of knowledge and belief

Let Ω be a (possibly infinite) non-empty set of *states*. The subsets of Ω are called *events*. Let N be a set of *individuals*. For each individual $i \in N$ we postulate a *belief operator* $B_i : 2^{\Omega} \to 2^{\Omega}$ (where 2^{Ω} denotes the set of subsets of Ω) and a *knowledge operator* $K_i : 2^{\Omega} \to 2^{\Omega}$. For $E \subseteq \Omega$, B_iE (respectively, K_iE) is the event that individual i believes (resp. knows) E.[7] These operators are assumed to satisfy the following properties ($\neg$ denotes complement): $\forall i \in N$, $\forall E, F \in 2^{\Omega}$

Necessity: $B_i\Omega = \Omega$ and $K_i\Omega = \Omega$. (Ax.1)

Monotonicity: if $E \subseteq F$ then $B_iE \subseteq B_iF$ and $K_iE \subseteq K_iF$. (Ax.2)

Conjunction: $B_i(E \cap F) = B_iE \cap B_iF$ and $K_i(E \cap F) = K_iE \cap K_iF$. (Ax.3)

Consistency: $B_iE \subseteq \neg B_i \neg E$ and $K_iE \subseteq \neg K_i \neg E$. (Ax.4)

Positive Introspection: $B_iE \subseteq B_iB_iE$ and $K_iE \subseteq K_iK_iE$. (Ax.5)

Negative Introspection: $\neg B_iE \subseteq B_i \neg B_iE$ and $\neg K_iE \subseteq K_i \neg K_iE$. (Ax.6)

Priority of knowledge: $K_iE \subseteq B_iE$. (Ax.7)

Transparency: $B_iE \subseteq K_iB_iE$. (Ax.8)

[7] Belief and knowledge 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 one needs to introduce a formal language and a procedure for associating an event with every formula in the language: see, for example, Battigalli and Bonanno (1997, pp. 41-48).

We call a tuple $\langle \Omega,\ N,\ \{B_i\}_{i\in N}\ ,\ \{K_i\}_{i\in N}\ \rangle$ where the operators B_i and K_i satisfy (Ax.1)-(Ax.8) a $\mathbb{KB}$-system. Note that (unlike negative introspection) positive introspection of belief is redundant, since it can be deduced from the other properties[8].

Systems for knowledge and belief have been studied by philosophers (cf. Hintikka, 1962, Lenzen, 1978,), computer scientists (cf. Halpern, 1991, van der Hoek, 1993, van der Hoek and Meyer, 1995, Kraus and Lehmann, 1988), as well as economists and game theorists (cf. Battigalli and Bonanno, 1997, Dekel and Gul, 1996, Geanakoplos, 1994). In this literature it is customary to impose the Truth Axiom for knowledge ($K_iE \subseteq E$) so that – while incorrect beliefs are allowed – only true facts can be known. In our analysis veridicality of knowledge is required only for a small class of events (cf. (Ax.9) below).

We shall denote by $\mathcal{B} : \Omega \to 2^{\Omega}$ (respectively, $\mathcal{K} : \Omega \to 2^{\Omega}$) the possibility correspondence associated with the belief operator B_i (resp. the knowledge operator K_i). Thus, $\forall\, \alpha \in \Omega,\ \mathcal{B}(\alpha) = \{\omega \in \Omega : \alpha \in \neg B_i \neg \{\omega\}\}$ and $\mathcal{K}(\alpha) = \{\omega \in \Omega : \alpha \in \neg K_i \neg \{\omega\}\}$.

REMARK 1. It is well-known (cf. Chellas, 1984) that, for all $\alpha \in \Omega$ and $E \subseteq \Omega$, $\alpha \in B_iE$ (resp. $\alpha \in K_iE$) if and only if $\mathcal{B}(\alpha) \subseteq E$ (resp. $\mathcal{K}(\alpha) \subseteq E$). Furthermore, B_i satisfies consistency if and only if $\mathcal{B}$ is *serial* ($\forall\, \omega \in \Omega,\ \mathcal{B}(\omega) \neq \varnothing$), it satisfies positive introspection if and only if $\mathcal{B}$ is *transitive* ($\forall\, \alpha, \beta \in \Omega$, if $\beta \in \mathcal{B}(\alpha)$ then $\mathcal{B}(\beta) \subseteq \mathcal{B}(\alpha)$) and it satisfies negative introspection if and only if $\mathcal{B}$ is *euclidean* ($\forall\, \alpha, \beta \in \Omega$, if $\beta \in \mathcal{B}(\alpha)$ then

---

[8] By (Ax.8) $B_iE \subseteq K_iB_iE$; by (Ax.7) $K_iB_iE \subseteq B_iB_iE$. Thus $B_iE \subseteq B_iB_iE$. On the other hand, since we have not assumed the Truth Axiom for knowledge, negative introspection of beliefs *cannot* be deduced from the other properties, as the following example shows: $\Omega = \{\alpha, \beta, \gamma\}$, $K\{\beta,\gamma\} = K\Omega = \Omega$ and $KE = \varnothing$ for every other event E, $B\{\gamma\} = \{\beta,\gamma\}$, $B\{\beta,\gamma\} = B\Omega = \Omega$ and $BE = \varnothing$ for every other event E. In this example K satisfies (Ax.1)-(Ax.6), B satisfies (Ax.1)-(Ax.5) and they jointly satisfy (Ax.7)-(Ax.8). However, B fails to satisfy (Ax.6): lettting $E = \{\gamma\}$, we have that $\neg BE = \{\alpha\} \not\subseteq B\neg BE = \varnothing$.

When the truth axiom is postulated for K_i then both positive and negative introspection of B_i become redundant (cf. Kraus and Lehmann, 1988, and Meyer and vand der Hoek, 1995).

$\mathcal{B}(\alpha) \subseteq \mathcal{B}(\beta)$). The same is true of K_i and $\mathcal{K}$. It is also well-known (cf. van der Hoek, 1993) that (Ax.7) is equivalent to $\mathcal{B}(\omega) \subseteq \mathcal{K}(\omega)$, $\forall\, \omega \in \Omega$, and (Ax.8) is equivalent to the following: $\forall\, \alpha, \beta, \gamma \in \Omega$, if $\beta \in \mathcal{B}(\alpha)$ and $\gamma \in \mathcal{K}(\beta)$ then $\gamma \in \mathcal{B}(\alpha)$.

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

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

Similarly, $K_*E = K_eE \cap K_eK_eE \cap K_eK_eK_eE \cap \ldots$

Let $\mathcal{B}_*: \Omega \to 2^{\Omega}$ and $\mathcal{K}_*: \Omega \to 2^{\Omega}$ be the corresponding possibility correspondences: $\forall\, \alpha \in \Omega$, $\mathcal{B}_*(\alpha) = \{\omega \in \Omega : \alpha \in \neg B_* \neg \{\omega\}\}$ and $\mathcal{K}_*(\alpha) = \{\omega \in \Omega : \alpha \in \neg K_* \neg \{\omega\}\}$. It is well known[9] that $\mathcal{B}_*$ can be characterized as the *transitive closure* of $\bigcup_{i \in N} \mathcal{B}_i$, that is,

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

Similarly, $\mathcal{K}_*$ is the transitive closure of $\bigcup_{i \in N} \mathcal{K}_i$.

---

[9] See, for example, Bonanno (1996), Fagin et al (1995), Halpern and Moses (1992), Lismont and Mongin (1994, 1995). These authors also show that the common belief (knowledge) operator can be alternatively defined by means of a finite list of axioms, rather than as an infinite conjunction.

Although we do not require individual knowledge to satisfy the Truth Axiom, for most of the results we do require a weakening of it, namely the assumption that individuals cannot be mistaken in what they know to be common knowledge: $\forall i \in N, \forall E \in 2^{\Omega}$,

$$K_i K_* E \subseteq K_* E \qquad \text{(Ax.9)}$$

It is shown in the Appendix (Lemma A.3) that (Ax.9) amounts to assuming Negative Introspection of common knowledge ($\neg K_* E \subseteq K_* \neg K_* E$). A $\mathbb{KB}$-system that satisfies (Ax.9) will be called a $\mathbb{KB}^+$-system. Note that in such a system individuals might have (not only incorrect beliefs but also) incorrect knowledge of any event which is not a common knowledge event. Thus the logic of knowledge is still KD45 (using the terminology of Chellas, 1984) as is the case with beliefs.

REMARK 2. In a $\mathbb{KB}^+$-system (*a fortiori* in a $\mathbb{KB}$-system) the properties of *individual* beliefs / knowledge are not inherited by *common* belief / knowledge. In particular, negative introspection of common belief ($\neg B_* E \subseteq B_* \neg B_* E$) and Common Transparency (the counterpart to (Ax.8): $B_* E \subseteq K_* B_* E$) are *not* satisfied in general, as the following example shows: $N = \{1, 2\}$, $\Omega = \{\alpha, \beta\}$, $\mathcal{K}_1(\alpha) = \mathcal{B}_1(\alpha) = \{\alpha\}$, $\mathcal{K}_1(\beta) = \mathcal{B}_1(\beta) = \{\beta\}$, $\mathcal{K}_2(\alpha) = \mathcal{K}_2(\beta) = \{\alpha, \beta\}$, $\mathcal{B}_2(\alpha) = \mathcal{B}_2(\beta) = \{\beta\}$. Thus $\mathcal{K}_*(\alpha) = \mathcal{K}_*(\beta) = \{\alpha, \beta\}$, $\mathcal{B}_*(\alpha) = \{\alpha, \beta\}$ and $\mathcal{B}_*(\beta) = \{\beta\}$. This is illustrated in Figure 4 (which extends the representation of Figure 1 by adding the common belief/knowledge correspondences) according to the following convention which will be used throughout the paper. States are denoted by dots and a (individual or common) belief possibility correspondence $\mathcal{B} : \Omega \to 2^{\Omega}$ is represented by thin arrows and thin ellipses as follows: $\omega' \in \mathcal{B}(\omega)$ if and only if either ω' and ω are enclosed in the same ellipsis or there is an arrow from ω, or from the ellipsis containing ω, to the ellipsis containing ω'. Similarly, a (individual or common) knowledge possibility correspondence $\mathcal{K} : \Omega \to 2^{\Omega}$ is represented by thick arrows and thick rectangles using the same convention: $\omega' \in \mathcal{K}(\omega)$ if and only if either

ω' and ω a re enclosed in the same rectangle or there is an arrow from ω, or from the rectangle containing ω, to the rectangle containing ω'.[10]

Insert Figure 4

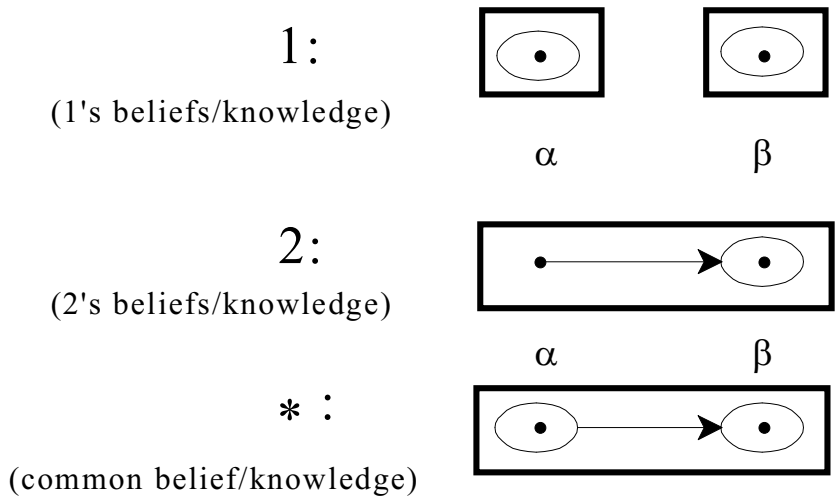

Figure 4

Let E = {β}. Then $B_*E = \{\beta\}$, $\neg B_*E = \{\alpha\}$ and $B_*\neg B_*E = \varnothing$. Thus Negative Introspection of common belief fails: $\neg B_*E \not\subseteq B_*\neg B_*E$. Furthermore, $K_*B_*E = \varnothing$. Thus Common Transparency fails at the intersubjective level: $B_*E \not\subseteq K_*B_*E$. [Note that (Ax.9) is satisfied, that is, the system of Figure 4 is indeed a $\mathbb{KB}^+$-system since in this example individual knowledge satisfies the Truth Axiom, in other words $\mathcal{K}_i$ is partitional.]

Given two events E and F, we denote by $(E \to F) \subseteq \Omega$ the following event

$$E \to F \underset{def}{=} \neg E \cup F .$$

[10] In most of the examples, for simplicity, each $\mathcal{K}_i$ will be a partition, that is, the Truth Axiom for knowledge will be satisfied (hence, a fortiori, (Ax.9)).

Thus $\alpha \in (E \rightarrow F)$ if and only if $\alpha \in E$ implies $\alpha \in F$ (hence $E \subseteq F$ is equivalent to $(E \rightarrow F) = \Omega$). Furthermore, let

$$E \leftrightarrow F \underset{def}{=} (E \rightarrow F) \cap (F \rightarrow E).$$

Thus $\alpha \in (E \leftrightarrow F)$ is equivalent to "$\alpha \in E$ if and only if $\alpha \in F$" (thus $E = F$ if and only if $(E \leftrightarrow F) = \Omega$).

4. Results

The following events capture important intersubjective properties of beliefs/knowledge (throughout the paper, events that represent properties of beliefs / knowledge are denoted by bold-face capital letters). Let

Negative Introspection of common belief $\quad \mathbf{NIB}^* = \bigcap_{E \in 2^\Omega} (\neg B_* E \rightarrow B_* \neg B_* E)$

Common Transparency $\quad \mathbf{TRN}^* = \bigcap_{E \in 2^\Omega} (B_* E \rightarrow K_* B_* E)$

Intersubjective caution $\quad \mathbf{ICAU} = \bigcap_{i \in N} \bigcap_{E \in 2^\Omega} (B_i B_* E \rightarrow K_i B_* E)$

Truth of common belief $\quad \mathbf{TB}^* = \bigcap_{E \in 2^\Omega} (B_* E \rightarrow E)$

Equivalence of common belief and common knowledge $\quad \mathbf{EQU}^* = \bigcap_{E \in 2^\Omega} (B_* E \leftrightarrow K_* E)$

Thus $\omega \in \mathbf{NIB}^*$ if and only if, for every event E, if $\omega \in \neg B_* E$ then $\omega \in B_* \neg B_* E$; $\omega \in \mathbf{TRN}^*$ if and only if, for every event E, if $\omega \in B_* E$ then $\omega \in K_* B_* E$; $\omega \in \mathbf{ICAU}$ if and only if, for every individual i and every event E, if $\omega \in B_i B_* E$ then $\omega \in K_i B_* E$; $\omega \in \mathbf{TB}^*$ if and only if, for every

event E, if $\omega \in B_*E$ then $\omega \in E$; finally, $\omega \in$ **EQU*** if and only if, for every event E, if $\omega \in B_*E$ then $\omega \in K_*E$ and vice versa.

NIB* is the analogue, for common belief, of (Ax.6) for individual beliefs, while **TRN*** is the analogue, for common belief and knowledge, of property (Ax.8) of individual beliefs/knowledge. **ICAU**, on the other hand, captures the notion of intersubjective caution of individual beliefs: individuals are cautious in what they believe to be common belief, in the sense that, while – in general – they allow for the possibility that they have incorrect beliefs, such mistakes are ruled out for common belief events. **TB*** captures the property that only true facts are commonly believed and **EQU*** the property that common belief and common knowledge coincide.

Lemma 1 gives the semantic properties that characterize these five events. For example, in the $\mathbb{KB}^+$-system of Figure 4, **NIB*** = $\{\beta\}$, **TRN*** = $\{\alpha\}$, **ICAU** = $\varnothing$, **TB*** = Ω and **EQU*** = $\{\alpha\}$. That **ICAU** = $\varnothing$ can be seen directly by noting that at every state individual 2 believes that E = $\{\beta\}$ is common belief ($B_2B_*E = \Omega$), but she does not know this, since $\mathcal{K}_2(\omega) = \Omega$, for every ω, while $B_*E = \{\beta\}$.

All the proofs are contained in the appendix.

LEMMA 1. In a $\mathbb{KB}$-system (thus without assuming (Ax.9)) the following holds for every $\alpha \in \Omega$:

(i) $\alpha \in$ **NIB*** if and only if $\mathcal{B}_*$ is euclidean at α, that is, $\forall\, \beta, \gamma \in \mathcal{B}_*(\alpha)$, $\gamma \in \mathcal{B}_*(\beta)$.

(ii) $\alpha \in$ **TRN*** if and only if, $\forall\, \beta, \gamma \in \Omega$, if $\beta \in \mathcal{K}_*(\alpha)$ and $\gamma \in \mathcal{B}_*(\beta)$ then $\gamma \in \mathcal{B}_*(\alpha)$.

(iii) $\alpha \in$ **ICAU** if and only if the following property holds:

(P_{ICAU}) $\forall\, \beta, \gamma \in \Omega$, if $\beta \in \mathcal{K}_i(\alpha)$ and $\gamma \in \mathcal{B}_*(\beta)$ then there exists a $\delta \in \mathcal{B}_i(\alpha)$ such that $\gamma \in \mathcal{B}_*(\delta)$.

(iv) $\alpha \in$ **TB*** if and only if $\alpha \in \mathcal{B}_*(\alpha)$.

(v) $\alpha \in$ **EQU*** if and only if $\mathcal{B}_*(\alpha) = \mathcal{K}_*(\alpha)$.

The following Theorem states that at the public (= common knowledge) level the three properties of Negative Introspection of common belief, Common Transparency and Intersubjective Caution coincide (although locally they typically do not).

THEOREM 1. In a $\mathbb{KB}^+$-system the following holds:

$$K_*\mathbf{NIB}^* = K_*\mathbf{TRN}^* = K_*\mathbf{ICAU}.$$

The next theorem shows that if one adds to common knowledge of intersubjective caution the hypothesis that it is common knowledge that only true facts are commonly believed, one obtains the collapse (both locally and globally) of common belief into common knowledge. This theorem, unlike the others, holds in general systems where no veridicality assumptions of any sort are imposed on the knowledge of individuals.

THEOREM 2. In a $\mathbb{KB}$-system (thus without assuming (Ax.9)) the following holds:

$$K_*\mathbf{ICAU} \cap K_*\mathbf{TB}^* = \mathbf{EQU}^* \cap K_*\mathbf{EQU}^* = K_*\mathbf{EQU}^*$$

In both Theorems 1 and 2 (as well as in the ones to follow), common knowledge of the events under consideration is crucial. For instance, in Figure 4, at state β, while there is common knowledge of the truth of common belief, there is only intersubjective caution but not common knowledge of it; in line with Theorems 1 and 2, common transparency fails at that state (and thus of necessity common knowledge and common belief fail to coincide); similar counterexamples can be constructed in each case.[11]

[11] Note also that, although $K_*\mathbf{EQU}^* \subseteq \mathbf{EQU}^*$, in general, $\mathbf{EQU}^* \not\subseteq K_*\mathbf{EQU}^*$ as the example of Figure 4 shows, where $\mathbf{EQU}^* = \{\alpha\}$ and $K_*\mathbf{EQU}^* = \varnothing$.

REMARK 3. None of the properties of beliefs/knowledge introduced so far embody agreement-type restrictions on individual beliefs, as the following example (illustrated in Figure 5) shows. : N = {1, 2}, $\Omega = \{\alpha, \beta\}$, $\mathcal{K}_1(\omega) = \mathcal{K}_2(\omega) = \{\alpha, \beta\}$, for all $\omega \in \Omega$, $\mathcal{B}_1(\alpha) = \mathcal{B}_1(\beta) = \{\beta\}$, $\mathcal{B}_2(\alpha) = \mathcal{B}_2(\beta) = \{\alpha\}$. Thus $\mathcal{K}_*(\omega) = \mathcal{B}_*(\omega) = \Omega$, for all $\omega \in \Omega$.

Insert Figure 5

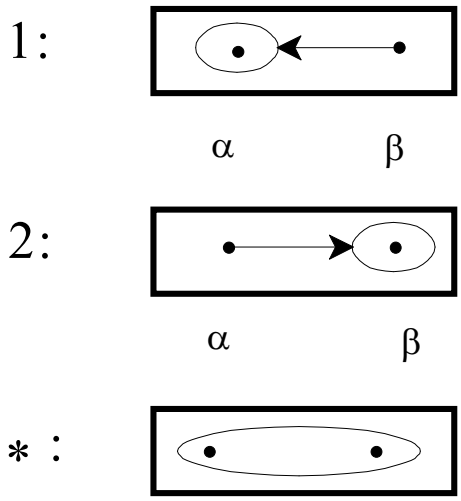

Figure 5

Here $\mathbf{NIB}^* = \mathbf{TRN}^* = \mathbf{ICAU} = \mathbf{TB}^* = \mathbf{EQU}^* = \Omega$ and yet the two individuals "agree to strongly disagree" in the sense that, at every state, it is common knowledge and common belief that individual 1 believes E = $\{\beta\}$ while individual 2 believes ¬E.

We now introduce two more properties of beliefs, one of which captures the property of Agreement. Let

Truth of individuals' beliefs $\quad \mathbf{TB}^{\mathbf{i}} = \bigcap_{i \in N} \bigcap_{E \in 2^{\Omega}} (B_i E \rightarrow E)$

Qualitative Agreement $\quad \mathbf{QA} = \neg \mathrm{B}_* \neg \mathrm{B}_* \mathbf{TB}^{\mathbf{i}}$

Thus $\alpha \in \mathbf{TB}^i$ if no individual has any false beliefs at α, that is, for every $i \in N$ and every $E \subseteq \Omega$, if $\alpha \in B_i E$ then $\alpha \in E$. It is well-known (see Chellas, 1984) that $\alpha \in \mathbf{TB}^i$ if and only if, $\forall i \in N$, $\alpha \in \mathscr{B}_i(\alpha)$. The event $B_* \mathbf{TB}^i$ captures a property known in the game theoretic literature as *common belief in no error* (cf. Ben Porath, 1997, Stalnaker, 1994, 1996, Stuart, 1997). **QA** is a weaker property than this in that it only requires the *common possibility* of common belief in no error, that is, $\alpha \in \mathbf{QA}$ if and only if for *some* $\beta \in \mathscr{B}_*(\alpha)$, $\beta \in B_* \mathbf{TB}^i$. Qualitative Agreement has been shown (Bonanno and Nehring, 1997b) to capture the notion of Agreement in two equivalent ways: (1) $\alpha \in \mathbf{QA}$ if and only if at α unbounded gains from betting are impossible (assuming moderately risk-averse preferences), (2) $\alpha \in \mathbf{QA}$ if and only if at α the individuals cannot "agree to disagree" about "union consistent" qualitative belief indices (a generalization of the Agreement property introduced by Aumann, 1976).

The following theorem shows that adding Qualitative Agreement to common knowledge of intersubjective caution yields common belief in no error. Indeed the conjunction of the former two properties is equivalent to the conjunction of common belief in no error and common knowledge of common transparency[12].

THEOREM 3. In a $\mathbb{KB}^+$-system the following holds:[13]

$$\mathbf{QA} \cap K_* \mathbf{ICAU} = B_* \mathbf{TB}^i \cap K_* \mathbf{TRN}^*.$$

---

[12] Lemma A.8 in the appendix shows that, in turn, common knowledge of common transparency implies common transparency, that is, $K_* \mathbf{TRN}^* = \mathbf{TRN}^* \cap K_* \mathbf{TRN}^*$.

[13] In a $\mathbb{KB}$-system (i.e. if (Ax.9) is not satisfied) in general $\mathbf{QA} \cap K_* \mathbf{ICAU} \not\subseteq B_* \mathbf{TB}^i$, as can be seen by modifying the example of Figure 4 in such a way that knowledge and belief coincide also for individual 2: $N = \{1, 2\}$, $\Omega = \{\alpha, \beta\}$, $\mathscr{K}_1(\alpha) = \mathscr{B}_1(\alpha) = \{\alpha\}$, $\mathscr{K}_1(\beta) = \mathscr{B}_1(\beta) = \{\beta\}$, $\mathscr{K}_2(\alpha) = \mathscr{K}_2(\beta) = \mathscr{B}_2(\alpha) = \mathscr{B}_2(\beta) = \{\beta\}$. Thus $\mathscr{K}_*(\alpha) = \mathscr{B}_*(\alpha) = \{\alpha, \beta\}$ and $\mathscr{K}_*(\beta) = \mathscr{B}_*(\beta) = \{\beta\}$. In this example, $\mathbf{QA} = \mathbf{ICAU} = K_* \mathbf{ICAU} = \Omega$, while $B_* \mathbf{TB}^i = \{\beta\}$ (since $\mathbf{TB}^i = \{\beta\}$).

The following theorem shows that putting together the three conditions of Qualitative Agreement, common knowledge of Intersubjective Caution and common knowledge that only true facts are commonly believed leads to the collapse (both locally and globally) of belief into knowledge for every individual. The theorem also states that such collapse of individual belief into knowledge is brought about also by the hypothesis of common knowledge that every individual has correct beliefs.

Let

Caution (coincidence of belief and knowledge for every individual)

$$\mathbf{CAU} = \bigcap_{i \in N} \bigcap_{E \in 2^{\Omega}} \left(B_i E \leftrightarrow K_i E \right)$$

Thus $\alpha \in \mathbf{CAU}$ if and only if, for every individual i and event E, at α individual i believes E ($\alpha \in B_i E$) if and only if she knows E ($\alpha \in K_i E$).

THEOREM 4. In a $\mathbb{KB}^{+}$-system the following holds:[14]

$$\mathbf{QA} \cap K_{*}\mathbf{ICAU} \cap K_{*}\mathbf{TB}^{*} = \mathbf{CAU} \cap K_{*}\mathbf{CAU} \cap \mathbf{QA} = K_{*}\mathbf{TB}^{i}.$$

[14] In a $\mathbb{KB}$-system Theorem 4 does not hold. For instance, in the example of Footnote 13, $\mathbf{CAU} \cap K_{*}\mathbf{CAU} \cap \mathbf{QA} \not\subseteq K_{*}\mathbf{TB}^{i}$ since $\mathbf{CAU} \cap K_{*}\mathbf{CAU} = \mathbf{QA} = \{\alpha, \beta\}$ while $\mathbf{TB}^{i} = K_{*}\mathbf{TB}^{i} = \{\beta\}$.

5. Conclusion: On the intersubjective rationality of the three fundamental consistency conditions

The analysis of this paper has spanned the intersubjective gap between belief and knowledge by three intersubjective consistency conditions: Qualitative Agreement, Intersubjective Caution, and Common knowledge that only true facts are commonly believed. How plausible are these conditions? Can they perhaps even be viewed as "*intersubjective rationality"* conditions?

As a reference point, it is instructive to consider the condition of "common belief in no error" (corresponding to the event $B_*\mathbf{TB}^i$). *Prima facie*, a case for it as a requirement of "intersubjective rationality" can be made by viewing it as an intersubjective generalization of secondary reflexivity[15]: every agent is willing to underwrite epistemically every other agent's beliefs to the extent that he knows about them.

However, a reinterpretation of Example 1 shows that this condition cannot be *always* applicable, which casts some doubt on the intersubjective rationality interpretation. Consider the following augmentation of the story underlying Example 1. At date zero, both the philosopher and his counterpart took it for granted that his name was spelled "Rabinowi*zc*"; however, after having per chance a private look at his birth certificate, he discovers to his great surprise that the correct spelling is "… *cz*". Formally, this can be described in a two-state universe augmenting Figure 1; the (non-augmented) figure now describes individuals' beliefs at date 1, after the (one-sided) inspection of the birth certificate. There are now two additional "epistemic agents" describing the individuals' beliefs at date 0; these in fact coincide with individual 2's beliefs at date 1. At date 0, both individuals' beliefs coincide and thus satisfy any meaningful intersubjective rationality condition. Individuals' beliefs at date 1, in particular the philosopher's

[15] Secondary reflexivity of individual beliefs is the property that each individual believes not to be mistaken in his own beliefs (the individual believes that if he believes E then E is true). Secondary reflexivity is implied by Negative Introspection (cf. Remark A.2 in the appendix).

certainty of the falsity of his counterparts' beliefs, are a necessary result of the information received in the interim; thus *neither* individual's beliefs at date 1 can be criticized for lack of intersubjective rationality.

Qualitative Agreement ($\neg B_* \neg B_* \mathbf{TB}^i$) can be viewed as an appropriate weakening of common belief in no error ($B_* \mathbf{TB}^i$) : if the epistemic assessments of an event E (that E is believed or that E is not believed, and more generally of a "qualitative belief index") of both agents are common belief, they must coincide. If any intersubjective consistency can stake a claim on rationality, it would seem to be Qualitative Agreement: its equivalence to the absence of unbounded gains form betting (cf. Bonanno and Nehring, 1997b) lends it strong normative appeal. Moreover, it is not subject to the contingencies of history, as it restricts agent's beliefs only when they are jointly commonly known. In Example 1, for instance, only trivial beliefs are *jointly* commonly known[16]. It would even make perfect sense to require Qualitative Agreement after counterfactual histories in a game!

Intersubjective Caution, by contrast, is exposed to the same problems in a dynamic setting that plague common belief in no error; note that it fails even *within* individual 1 who at date 0 took the wrong spelling for granted (and believed that he would continue to take it for granted), recognizing the possibility (in terms of knowledge) that he might live to change his mind. On the other hand, while not categorical, Intersubjective Caution seems highly reasonable as a constraint on how individuals "initially" construct their intersubjective belief hierarchies, prior to the receipt of specific private information (but incompletely informed of each other's beliefs), for example prior to the actual play of the game. This would be sufficient to justify the striking Stalnaker-Stuart justification of non-cooperative play in the repeated prisoner's dilemma game (Stalnaker, 1996, Stuart, 1997) which preempts the "maintained rationality" critique of the backward induction solution in extensive form games.

[16] Note that while individual 2's beliefs about the event $\{\beta\}$ are commonly known, 1's beliefs about $\{\beta\}$ are not.

Common knowledge of the truth of common belief clearly has the flavor of an empirical rather than a rationality assumption. It seems implausible as the latter; note, for example, that applied to a group of one, it coincides with caution[17]. In view of Theorem 2, and taking into account the plausibility of both **QA** and **ICAU**, it seems implausible even as an empirical assumption, in spite of the appeal to the reasonable intuition that a group's beliefs may enjoy higher epistemic dignity than any individual's beliefs.

[17] By contrast, both **QA** and **ICAU** are automatically satisfied in this case.

Appendix

PROOF OF LEMMA 1. (i) and (iv) are well-known (see Chellas, 1984); (ii) follows from Theorem 4.3 (c) in van der Hoek (1993, p. 183) and (v) is straightforward. Thus we shall only prove (iii).

(Not P_{ICAU} at $\alpha \Rightarrow \alpha \notin$ **ICAU**). Suppose P_{ICAU} does not hold at α. Then there exist $\beta, \gamma \in \Omega$ such that $\beta \in \mathcal{K}_i(\alpha)$, $\gamma \in \mathcal{B}_*(\beta)$ and, $\forall\ \delta \in \mathcal{B}_i(\alpha)$, $\gamma \notin \mathcal{B}_*(\delta)$. Let

$$E = \{\omega \in \Omega : \omega \in \mathcal{B}_*(\omega') \text{ for some } \omega' \in \mathcal{B}_i(\alpha)\}.$$

Then

$$\gamma \notin E \tag{1}$$

and, by construction,

$$\alpha \in B_i B_* E \tag{2}$$

By (1), since $\gamma \in \mathcal{B}_*(\beta)$, $\beta \notin B_*E$. Hence, since $\beta \in \mathcal{K}_i(\alpha)$, $\alpha \notin K_i B_* E$. Thus, by (2), $\alpha \notin (B_i B_* E \rightarrow K_i B_* E)$. Hence $\alpha \notin$ **ICAU**.

($\alpha \notin$ **ICAU** $\Rightarrow$ Not P_{ICAU} at α). Suppose that $\alpha \notin$ **ICAU**. Then there exist $E \subseteq \Omega$ and $i \in N$ such that $\alpha \in B_i B_* E \cap \neg K_i B_* E$. Since $\alpha \in \neg K_i B_* E$, there exist $\beta, \gamma \in \Omega$ such that $\beta \in \mathcal{K}_i(\alpha)$ and $\gamma \in \mathcal{B}_*(\beta) \cap \neg E$. Since $\alpha \in B_i B_* E$, $\forall\, \delta \in \mathcal{B}_i(\alpha)$, $\delta \in B_* E$, that is, $\mathcal{B}_*(\delta) \subseteq E$. Hence $\gamma \notin \mathcal{B}_*(\delta)$. Thus P_{ICAU} does not hold at α. ■

The proof of Theorem 1 will be carried out in three steps. The first step is given by Lemma A.1, which holds for all $\mathbb{KB}$-systems (thus without assuming (Ax.9)). The second step is given by Proposition A.1, which is a restatement of Theorem 1 for $\mathbb{KB}$-systems that satisfy an additional property. The third and final step is given by Lemma A.3 which shows that this additional property is equivalent to (Ax.9).

Let (VB^{i*} stands for "Veridicality of individual belief about common belief")

$$\mathbf{VB}^{i*} = \bigcap_{i \in N} \bigcap_{E \in 2^{\Omega}} \left(B_i B_* E \rightarrow B_* E\right).$$

Thus $\omega \in \mathbf{VB}^{i*}$ if and only if for every individual i and event E, if $\omega \in B_i B_* E$ then $\omega \in B_* E$, that is, at ω no individual has mistaken beliefs about what is commonly believed.

REMARK A.1. For every $\alpha \in \Omega$, $\alpha \in \mathbf{VB}^{i*}$ if and only if $\forall\, i \in N$, $\forall\, \gamma \in \mathscr{B}_*(\alpha)$, $\exists\, \delta \in \mathscr{B}_i(\alpha)$ such that $\gamma \in \mathscr{B}_*(\delta)$. For a proof see Lemma 2 in Bonanno and Nehring (1997b).[18]

LEMMA A.1. In a $\mathbb{KB}$-system (thus without postulating (Ax.9)) the following holds:

$$K_*\mathbf{NIB}^* \subseteq K_*\mathbf{VB}^{i*} \subseteq K_*\mathbf{ICAU} \subseteq K_*\mathbf{TRN}^*.$$

Proof. ($K_*\mathbf{NIB}^* \subseteq K_*\mathbf{VB}^{i*}$). First we show that $\mathbf{NIB}^* \subseteq \mathbf{VB}^{i*}$. Let $\alpha \in \mathbf{NIB}^*$. Fix an arbitrary $i \in N$ and $E \subseteq \Omega$. We want to show that $\alpha \in (B_iB_*E \rightarrow B_*E)$, or, equivalently, that $\alpha \in (\neg B_*E \rightarrow \neg B_iB_*E)$. Since $\alpha \in \mathbf{NIB}^*$, $\alpha \in (\neg B_*E \rightarrow B_*\neg B_*E)$. Suppose that $\alpha \in \neg B_*E$. Then $\alpha \in B_*\neg B_*E$. By definition of B_*, $B_*\neg B_*E \subseteq B_i\neg B_*E$. By Consistency of i's beliefs (cf. Ax.4), $B_i\neg B_*E \subseteq \neg B_iB_*E$. Thus $B_*\neg B_*E \subseteq \neg B_iB_*E$. Hence $\alpha \in \neg B_iB_*E$. Thus $\mathbf{NIB}^* \subseteq \mathbf{VB}^{i*}$. By Monotonicity of K_*, it follows that $K_*\mathbf{NIB}^* \subseteq K_*\mathbf{VB}^{i*}$.

($K_*\mathbf{VB}^{i*} \subseteq K_*\mathbf{ICAU}$). Let $\alpha \in K_*\mathbf{VB}^{i*}$ and fix an arbitrary $\beta \in \mathscr{K}_*(\alpha)$. We want to show that $\beta \in \mathbf{ICAU}$. Fix arbitrary $i \in N$ and $E \subseteq \Omega$ such that $\beta \in B_iB_*E$. Fix an arbitrary $\gamma \in \mathscr{K}_i(\beta)$. We need to show that $\gamma \in B_*E$. Since $B_iB_*E \subseteq K_iB_iB_*E$ (cf. Ax.8), $\beta \in K_iB_iB_*E$ hence $\gamma \in B_iB_*E$. By definition of $\mathscr{K}_*$, since $\beta \in \mathscr{K}_*(\alpha)$ and $\gamma \in \mathscr{K}_i(\beta)$, $\gamma \in \mathscr{K}_*(\alpha)$. Thus, since $\alpha \in K_*\mathbf{VB}^{i*}$, $\gamma \in \mathbf{VB}^{i*}$. Hence, since $\gamma \in B_iB_*E$, $\gamma \in B_*E$.

($K_*\mathbf{ICAU} \subseteq K_*\mathbf{TRN}^*$). Let $\alpha \in K_*\mathbf{ICAU}$. Fix an arbitrary $\beta \in \mathscr{K}_*(\alpha)$. We want to show that $\beta \in \mathbf{TRN}^*$. Fix arbitrary $E \subseteq \Omega$ such that $\beta \in B_*E$. We need to show that $\beta \in K_*B_*E$. Fix arbitrary sequences $\langle i_1, \ldots, i_m \rangle$ in N and $\langle \beta_0, \beta_1, \ldots, \beta_m \rangle$ in Ω such that $\beta_0 = \beta$ and, for every $k = 1, \ldots, m$, $\beta_k \in \mathscr{K}_{i_k}(\beta_{k-1})$. We need to show that $\beta_m \in B_*E$. First of all, note that, since $\beta \in \mathscr{K}_*(\alpha)$, by definition of $\mathscr{K}_*$, $\beta_k \in \mathscr{K}_*(\alpha)$ for all $k = 0, \ldots, m$. Hence, since $\alpha \in K_*\mathbf{ICAU}$,

$$\beta_k \in \mathbf{ICAU} \quad \text{for all } k = 0, \ldots, m. \tag{3}$$

Since $\beta_0 = \beta \in B_*E$ and, by definition of B_*, $B_*E \subseteq B_{i_1}B_*E$, $\beta_0 \in B_{i_1}B_*E$. Hence, by (3), $\beta_0 \in K_{i_1}B_*E$. Thus, since $\beta_1 \in \mathscr{K}_{i_1}(\beta_0)$, $\beta_1 \in B_*E$. Since $B_*E \subseteq B_{i_2}B_*E$, $\beta_1 \in B_{i_2}B_*E$. Hence, by

[18] There the event $\mathbf{VB}^{i*}$ is denoted by $\mathbf{T_{CB}}$ and the possibility correspondence $\mathscr{B}_*$ (resp. $\mathscr{B}_i$) is denoted by I_* (resp. I_i).

(3), $\beta_1 \in K_{i_2} B_* E$. Thus, since $\beta_2 \in \mathcal{K}_{i_2}(\beta_1)$, $\beta_2 \in B_* E$. Repeating this argument m times we get that $\beta_m \in B_* E$. ■

REMARK A.2. A possibility correspondence $\mathcal{P} : \Omega \to 2^{\Omega}$ is *secondary reflexive* if $\forall\, \alpha,\beta \in \Omega$, $\beta \in \mathcal{P}(\alpha)$ implies $\beta \in \mathcal{P}(\beta)$. Secondary reflexivity is implied by euclideanness. Hence, for every $i \in N$, $\mathcal{B}_i$ and $\mathcal{K}_i$ are secondary reflexive. It follows from the definition of $\mathcal{B}_*$ and $\mathcal{K}_*$ that both $\mathcal{B}_*$ and $\mathcal{K}_*$ are secondary reflexive.

Let ($\mathbf{NIK}^*$ stands for "Negative Introspection of common knowledge")

$$\mathbf{NIK}^* = \bigcap_{E \in 2^{\Omega}} \left(\neg K_* E \to K_* \neg K_* E \right).$$

REMARK A.3. Analogously to (i) of Lemma 1, it can be shown that $\beta \in \mathbf{NIK}^*$ if and only if $\mathcal{K}_*$ is euclidean at β, that is, $\forall\, \gamma, \delta \in \mathcal{K}_*(\beta)$, $\delta \in \mathcal{K}_*(\gamma)$.

PROPOSITION A.1. In a $\mathbb{KB}$-system satisfying $K_* \mathbf{NIK}^* = \Omega$ the following holds:

$$K_* \mathbf{NIB}^* = K_* \mathbf{VB}^{i*} = K_* \mathbf{ICAU} = K_* \mathbf{TRN}^*.$$

Proposition A.1 follows directly from Lemma A.1 and the following lemma which can be viewed as a generalization of Lemma 2.2 in Kraus and Lehmann (1988) to the case where individual knowledge satisfies the KD45 (rather than the S5) logic.

LEMMA A.2. In a $\mathbb{KB}$-system the following holds:

$$K_* \mathbf{NIK}^* \cap K_* \mathbf{TRN}^* \subseteq K_* \mathbf{NIB}^*.$$

Proof. Let $\alpha \in K_* \mathbf{NIK}^* \cap K_* \mathbf{TRN}^*$ and fix an arbitrary $\beta \in \mathcal{K}_*(\alpha)$. We need to show that $\beta \in \mathbf{NIB}^*$, that is (cf. (i) of Lemma A.1), for all $\delta, \gamma \in \mathcal{B}_*(\beta)$, $\delta \in \mathcal{B}_*(\gamma)$. Fix arbitrary $\delta, \gamma \in \mathcal{B}_*(\beta)$. By secondary reflexivity of $\mathcal{B}_*$ (cf. Remark A.2),

$$\delta \in \mathcal{B}_*(\delta) \tag{4}$$

Since, for all $\omega \in \Omega$, $\mathcal{B}_*(\omega) \subseteq \mathcal{K}_*(\omega)$, $\gamma, \delta \in \mathcal{K}_*(\beta)$. Since $\beta \in \mathcal{K}_*(\alpha)$ and $\alpha \in K_* \mathbf{NIK}^*$, $\beta \in \mathbf{NIK}^*$. Hence (cf. Remark A.3),

$$\delta \in \mathcal{K}_*(\gamma) \tag{5}$$

Since $\beta \in \mathcal{K}_*(\alpha)$ and $\gamma \in \mathcal{K}_*(\beta)$, by transitivity of $\mathcal{K}_*$, $\gamma \in \mathcal{K}_*(\alpha)$. Thus, since $\alpha \in K_* \mathbf{TRN}^*$,

$$\gamma \in \mathbf{TRN}^*. \tag{6}$$

It follows from (4)-(6) and (ii) of Lemma 1 that $\delta \in \mathcal{B}_*(\gamma)$. ■

Let ($\mathbf{VK}^{i*}$ stands for "Veridicality of individual knowledge about common knowledge")

$$\mathbf{VK}^{i*} = \bigcap_{i \in N} \bigcap_{E \in 2^{\Omega}} \left(K_i K_* E \rightarrow K_* E \right).$$

Thus $\omega \in \mathbf{VK}^{i*}$ if and only if for every individual i and event E, if $\omega \in K_i K_* E$ then $\omega \in K_* E$, that is, at ω no individual has mistaken knowledge about what is commonly known.

LEMMA A.3. In a $\mathbb{KB}^{+}$-system $\mathbf{NIK}^{*} = \Omega$.

Proof. Note that in a $\mathbb{KB}^{+}$-system $\mathbf{VK}^{i*} = \Omega$, since (Ax.9) is equivalent to $\mathbf{VK}^{i*} = \Omega$. We want to show that, in turn, $\mathbf{VK}^{i*} = \Omega$ is equivalent to $\mathbf{NIK}^{*} = \Omega$. We show this to be true in general, for *any* "common" operator. Let $\{B_i : 2^{\Omega} \rightarrow 2^{\Omega}\}_{i \in N}$ be any collection of individual operators satisfying (Ax.1) – (Ax. 6) (that is, Necessity, Monotonicity, Conjunction, Consistency, Positive and Negative Introspection) and let B_* be the corresponding common operator. We want to show that $\mathbf{VB}^{i*} = \Omega$ if and only if $\mathbf{NIB}^{*} = \Omega$.[19] Let $\mathscr{B}_i : \Omega \rightarrow 2^{\Omega}$ be the possibility correspondence associated with B_i. For every $i \in N$ construct the possibility correspondence $\mathscr{K}_i : \Omega \rightarrow 2^{\Omega}$ as follows: $\forall\, \omega, \omega' \in \Omega$, $\omega' \in \mathscr{K}_i(\omega)$ if and only if $\mathscr{B}_i(\omega') = \mathscr{B}_i(\omega)$. Then $\mathscr{K}_i$ gives rise to a partition of Ω, that is, $\forall\, \omega, \omega' \in \Omega$, $\omega \in \mathscr{K}_i(\omega)$ and if $\omega' \in \mathscr{K}_i(\omega)$ then $\mathscr{K}_i(\omega') = \mathscr{K}_i(\omega)$ (in the economics and game-theory literature this partition is called the *type partition* of individual i). Let K_i be the associated knowledge operator of individual i. The system so constructed is a $\mathbb{KB}$-system (it is straightforward to verify that (Ax.1)-(Ax.8) are satisfied). Let $\mathscr{K}_* : \Omega \rightarrow 2^{\Omega}$ be the transitive closure of $\bigcup_{i \in N} \mathscr{K}_i$ and K_* the associated common knowledge operator ($\forall E \in 2^{\Omega}$, $K_* E = \{\omega \in \Omega : \mathscr{K}_*(\omega) \subseteq E\}$). Then $\mathscr{K}_*$ also gives rise to a partition of Ω and therefore is euclidean, that is (cf. Lemma 1), $\mathbf{NIK}^{*} = \Omega$. Thus we can invoke Proposition A.1 and conclude that

$$K_* \mathbf{NIB}^{*} = K_* \mathbf{VB}^{i*} \tag{7}$$

Furthermore, since $\mathscr{K}_*$ is partitional, K_* satisfies the Truth Axiom, that is, $\forall E \in 2^{\Omega}$, $K_* E \subseteq E$. Hence

$$\forall E \in 2^{\Omega}, \qquad K_* E = \Omega \text{ if and only if } E = \Omega \tag{8}$$

[19] This result can also be proved as a Corollary to Theorem 1 in Bonanno and Nehring (1997a).

Suppose now that $\mathbf{VB}^{i*} = \Omega$. Then, by Necessity, $K_*\mathbf{VB}^{i*} = \Omega$. Thus, by (7), $K_*\mathbf{NIB}^{*} = \Omega$ and, by (8), $\mathbf{NIB}^{*} = \Omega$. By the same argument, if $\mathbf{NIB}^{*} = \Omega$ then $\mathbf{VB}^{i*} = \Omega$. ■

PROOF OF THEOREM 1. By Lemma A.3, in a $\mathbb{KB}^{+}$-system $\mathbf{NIK}^{*} = \Omega$. Thus, by Monotonicity of K_*, $K_*\mathbf{NIK}^{*} = \Omega$. Hence Theorem 1 follows from Proposition A.1. ■

The proof of Theorem 2 is split into several steps, given by Lemmas A.4-A.6.

LEMMA A.4. In a $\mathbb{KB}$-system $K_*\mathbf{EQU}^{*} \subseteq \mathbf{EQU}^{*}$.

Proof. Let $\alpha \in K_*\mathbf{EQU}^{*}$. We want to show (cf. (v) of Lemma 1) that $\mathcal{K}_*(\alpha) = \mathcal{B}_*(\alpha)$, that is (since $\mathcal{B}_*(\omega) \subseteq \mathcal{K}_*(\omega)\ \forall\ \omega \in \Omega$), that $\mathcal{K}_*(\alpha) \subseteq \mathcal{B}_*(\alpha)$. Fix an arbitrary $\gamma \in \mathcal{K}_*(\alpha)$. Then there exist $i \in N$ and $\beta \in \mathcal{K}_i(\alpha)$ such that $\gamma \in \mathcal{K}_*(\beta)$ [in the case where $\gamma \in \mathcal{K}_i(\alpha)$ we can take $\beta = \gamma$ and use secondary reflexivity of $\mathcal{K}_*$]. Since $\mathcal{K}_i(\alpha) \subseteq \mathcal{K}_*(\alpha)$, $\beta \in \mathcal{K}_*(\alpha)$ and, therefore (since $\alpha \in K_*\mathbf{EQU}^{*}$), $\beta \in \mathbf{EQU}^{*}$. Thus, by (v) of Lemma 1, $\mathcal{K}_*(\beta) = \mathcal{B}_*(\beta)$. Hence (since $\gamma \in \mathcal{K}_i(\beta) \subseteq \mathcal{K}_*(\beta)$)

$$\gamma \in \mathcal{B}_*(\beta). \tag{9}$$

By seriality of $\mathcal{B}_i$ (cf. Remark 1), $\mathcal{B}_i(\alpha) \neq \varnothing$. Fix an arbitrary $\delta \in \mathcal{B}_i(\alpha)$. Since $\mathcal{B}_i(\alpha) \subseteq \mathcal{K}_i(\alpha)$, $\delta \in \mathcal{K}_i(\alpha)$. Thus $\beta, \delta \in \mathcal{K}_i(\alpha)$; hence, by euclideanness of $\mathcal{K}_i$, $\beta \in \mathcal{K}_i(\delta)$. Thus, since $\mathcal{K}_i(\delta) \subseteq \mathcal{K}_*(\delta)$

$$\beta \in \mathcal{K}_*(\delta). \tag{10}$$

Since $\delta \in \mathcal{K}_i(\alpha) \subseteq \mathcal{K}_*(\alpha)$ and $\alpha \in K_*\mathbf{EQU}^{*}$, $\delta \in \mathbf{EQU}^{*}$. Hence, by (v) of Lemma 1, $\mathcal{K}_*(\delta) = \mathcal{B}_*(\delta)$. Thus, by (10),

$$\beta \in \mathcal{B}_*(\delta). \tag{11}$$

Since $\delta \in \mathcal{B}_i(\alpha)$ and $\mathcal{B}_i(\alpha) \subseteq \mathcal{B}_*(\alpha)$, $\delta \in \mathcal{B}_*(\alpha)$. It follows from this and (11), by transitivity of $\mathcal{B}_*$, that $\beta \in \mathcal{B}_*(\alpha)$. Using this and (9) (and transitivity of $\mathcal{B}_*$) we conclude that $\gamma \in \mathcal{B}_*(\alpha)$. Since $\gamma \in \mathcal{K}_*(\alpha)$ was chosen arbitrarily, we have shown that $\mathcal{K}_*(\alpha) \subseteq \mathcal{B}_*(\alpha)$. ■

The following two lemmas are one-operator results.[20]

---

[20] In the case where the cardinality of N is 1, Lemma A.5 states that adding to Transparency the assumption that the individual knows his beliefs to be correct leads to the collapse of knowledge and beliefs. In the philosophy literature an alternative way in which this collapse can take place has been discussed, namely adding to Transparency the assumption that if the agent believes something he believes that he knows it (cf. Hintikka, 1962, Lentzen, 1978; see also van der Hoek, 1993 and van der Hoek and Meyer, 1995): this is the content of Lemma A.6 (i).

LEMMA A.5. In a $\mathbb{KB}$-system (thus without assuming (A.9))

$$\mathbf{TRN}^* \cap \mathrm{K}_*\mathbf{TB}^* \subseteq \mathbf{EQU}^*$$

Proof. Let $\alpha \in \mathbf{TRN}^* \cap \mathrm{K}_*\mathbf{TB}^*$. We want to show that $\mathscr{K}_*(\alpha) \subseteq \mathscr{B}_*(\alpha)$. Fix an arbitrary $\beta \in \mathscr{K}_*(\alpha)$. Then, since $\alpha \in \mathrm{K}_*\mathbf{TB}^*$, $\beta \in \mathbf{TB}^*$. Thus (cf. (iv) of Lemma 1) $\beta \in \mathscr{B}_*(\beta)$. Since $\alpha \in \mathbf{TRN}^*$, $\beta \in \mathscr{K}_*(\alpha)$ and $\beta \in \mathscr{B}_*(\beta)$, by (ii) of Lemma 1 (with $\gamma = \beta$) it follows that $\beta \in \mathscr{B}_*(\alpha)$. ■

REMARK A.4. By transitivity and secondary reflexivity of $\mathscr{B}_*$ and $\mathscr{K}_*$, for every event E, $\mathrm{B}_*\mathrm{E} = \mathrm{B}_*\mathrm{B}_*\mathrm{E}$ and $\mathrm{K}_*\mathrm{E} = \mathrm{K}_*\mathrm{K}_*\mathrm{E}$.

COROLLARY A.1. In a $\mathbb{KB}$-system (thus without assuming (A.9))

$$\mathrm{K}_*\mathbf{TRN}^* \cap \mathrm{K}_*\mathbf{TB}^* \subseteq \mathrm{K}_*\mathbf{EQU}^*$$

Proof. By Lemma A.5 and monotonicity of K_*, $\mathrm{K}_*\mathbf{TRN}^* \cap \mathrm{K}_*\mathrm{K}_*\mathbf{TB}^* \subseteq \mathrm{K}_*\mathbf{EQU}^*$. By Remark A.4, $\mathrm{K}_*\mathrm{K}_*\mathbf{TB}^* = \mathrm{K}_*\mathbf{TB}^*$. ■

LEMMA A.6. In a $\mathbb{KB}$-system (thus without assuming (A.9)) $\forall \mathrm{E} \subseteq \Omega$, $\forall \mathrm{i} \in \mathrm{N}$

$$\text{(i)}\ \mathrm{B}_\mathrm{i}\mathrm{K}_\mathrm{i}\mathrm{E} \subseteq \mathrm{K}_\mathrm{i}\mathrm{E} \quad \text{and} \quad \text{(ii)}\ \mathrm{B}_\mathrm{i}\mathrm{K}_*\mathrm{E} = \mathrm{K}_\mathrm{i}\mathrm{K}_*\mathrm{E}.$$

Proof. (i) From $\neg\mathrm{K}_\mathrm{i}\mathrm{E} \subseteq \mathrm{K}_\mathrm{i}\neg\mathrm{K}_\mathrm{i}\mathrm{E}$ (Ax.6) and $\mathrm{K}_\mathrm{i}\neg\mathrm{K}_\mathrm{i}\mathrm{E} \subseteq \mathrm{B}_\mathrm{i}\neg\mathrm{K}_\mathrm{i}\mathrm{E}$ ((Ax.7) applied to the event $\neg\mathrm{K}_\mathrm{i}\mathrm{E}$) we get $\neg\mathrm{K}_\mathrm{i}\mathrm{E} \subseteq \mathrm{B}_\mathrm{i}\neg\mathrm{K}_\mathrm{i}\mathrm{E}$, which is equivalent to $\neg\mathrm{B}_\mathrm{i}\neg\mathrm{K}_\mathrm{i}\mathrm{E} \subseteq \mathrm{K}_\mathrm{i}\mathrm{E}$. This, in conjunction with $\mathrm{B}_\mathrm{i}\mathrm{K}_\mathrm{i}\mathrm{E} \subseteq \neg\mathrm{B}_\mathrm{i}\neg\mathrm{K}_\mathrm{i}\mathrm{E}$ ((Ax.4) applied to the event $\mathrm{K}_\mathrm{i}\mathrm{E}$), yields $\mathrm{B}_\mathrm{i}\mathrm{K}_\mathrm{i}\mathrm{E} \subseteq \mathrm{K}_\mathrm{i}\mathrm{E}$.
(ii) Since (by definition of K_*) $\mathrm{K}_*\mathrm{E} \subseteq \mathrm{K}_\mathrm{i}\mathrm{K}_*\mathrm{E}$, by monotonicity of B_i, $\mathrm{B}_\mathrm{i}\mathrm{K}_*\mathrm{E} \subseteq \mathrm{B}_\mathrm{i}\mathrm{K}_\mathrm{i}\mathrm{K}_*\mathrm{E}$ and, by (i), $\mathrm{B}_\mathrm{i}\mathrm{K}_\mathrm{i}\mathrm{K}_*\mathrm{E} \subseteq \mathrm{K}_\mathrm{i}\mathrm{K}_*\mathrm{E}$. Thus $\mathrm{B}_\mathrm{i}\mathrm{K}_*\mathrm{E} \subseteq \mathrm{K}_\mathrm{i}\mathrm{K}_*\mathrm{E}$. On the other hand, by (Ax. 7) $\mathrm{K}_\mathrm{i}\mathrm{K}_*\mathrm{E} \subseteq \mathrm{B}_\mathrm{i}\mathrm{K}_*\mathrm{E}$. ■

COROLLARY A.2. In a $\mathbb{KB}$-system (thus without assuming (A.9))

$$\mathrm{K}_*\mathbf{EQU}^* \subseteq \mathbf{ICAU}.$$

Proof. Let $\alpha \in \mathrm{K}_*\mathbf{EQU}^*$. Fix arbitrary $\mathrm{i} \in \mathrm{N}$ and $\mathrm{E} \subseteq \Omega$ such that $\alpha \in \mathrm{B}_\mathrm{i}\mathrm{B}_*\mathrm{E}$. We want to show that $\alpha \in \mathrm{K}_\mathrm{i}\mathrm{B}_*\mathrm{E}$. First we show that $\alpha \in \mathrm{B}_\mathrm{i}\mathrm{K}_*\mathrm{E}$. Fix an arbitrary $\beta \in \mathscr{B}_\mathrm{i}(\alpha)$. Then $\beta \in \mathrm{B}_*\mathrm{E}$. Since $\mathscr{B}_\mathrm{i}(\alpha) \subseteq \mathscr{K}_\mathrm{i}(\alpha) \subseteq \mathscr{K}_*(\alpha)$, $\beta \in \mathscr{K}_*(\alpha)$ and therefore $\beta \in \mathbf{EQU}^*$. Hence, since $\beta \in \mathrm{B}_*\mathrm{E}$, $\beta \in \mathrm{K}_*\mathrm{E}$. Thus $\alpha \in \mathrm{B}_\mathrm{i}\mathrm{K}_*\mathrm{E}$. By (ii) of Lemma A.6, $\alpha \in \mathrm{K}_\mathrm{i}\mathrm{K}_*\mathrm{E}$. Now choose an

arbitrary $\gamma \in \mathscr{K}_i(\alpha)$. Then $\gamma \in K_*E$. Furthermore, since $\mathscr{K}_i(\alpha) \subseteq \mathscr{K}_*(\alpha)$, $\gamma \in \mathscr{K}_*(\alpha)$ and therefore $\gamma \in \mathbf{EQU}^*$. Thus, since $\gamma \in K_*E$, $\gamma \in B_*E$. Hence $\alpha \in K_iB_*E$. ■

PROOF OF THEOREM 2. ($K_*\mathbf{ICAU} \cap K_*\mathbf{TB}^* \subseteq K_*\mathbf{EQU}^*$) By Lemma A.1, $K_*\mathbf{ICAU} \cap K_*\mathbf{TB}^* \subseteq K_*\mathbf{TRN}^* \cap K_*\mathbf{TB}^*$ and by Corollary A.1, $K_*\mathbf{TRN}^* \cap K_*\mathbf{TB}^* \subseteq K_*\mathbf{EQU}^*$.

($K_*\mathbf{EQU}^* \subseteq K_*\mathbf{ICAU}$) By Corollary A.2 and Monotonicity of K_*, $K_*K_*\mathbf{EQU}^* \subseteq K_*\mathbf{ICAU}$. By Remark A.4, $K_*K_*\mathbf{EQU}^* = K_*\mathbf{EQU}^*$.

($K_*\mathbf{EQU}^* \subseteq K_*\mathbf{TB}^*$). Let $\alpha \in K_*\mathbf{EQU}^*$ and fix an arbitrary $\beta \in \mathscr{K}_*(\alpha)$. We want to show that $\beta \in \mathbf{TB}^*$. Since $\alpha \in K_*\mathbf{EQU}^*$, $\beta \in \mathbf{EQU}^*$. Thus, by (v) of Lemma 1, $\mathscr{B}_*(\beta) = \mathscr{K}_*(\beta)$. Since $\beta \in \mathscr{K}_*(\alpha)$, by secondary reflexivity of $\mathscr{K}_*$ (cf. Remark A.2), $\beta \in \mathscr{K}_*(\beta)$. Thus $\beta \in \mathscr{B}_*(\beta)$. Hence, by (iv) of Lemma 1, $\beta \in \mathbf{TB}^*$.

Finally, by Lemma A.4, $K_*\mathbf{EQU}^* \subseteq \mathbf{EQU}^*$ hence $\mathbf{EQU}^* \cap K_*\mathbf{EQU}^* = K_*\mathbf{EQU}^*$. ■

The proof of Theorem 3 makes use of the following lemma.

LEMMA A.7. In a $\mathbb{KB}$-system (thus without assuming (Ax.9)) the following holds:

$$\mathbf{QA} \cap K_*\mathbf{ICAU} \cap \mathbf{NIK}^* \subseteq B_*\mathbf{TB}^i$$

Proof. Let $\alpha \in \mathbf{QA} \cap K_*\mathbf{ICAU} \cap \mathbf{NIK}^*$. Since $\alpha \in QA$, there exists a $\beta \in \mathscr{B}_*(\alpha)$ such that $\beta \in B_*\mathbf{TB}^i$. Suppose that $\alpha \notin B_*\mathbf{TB}^i$. Then there exists a $\gamma \in \mathscr{B}_*(\alpha)$ such that

$$\gamma \notin B_*\mathbf{TB}^i. \tag{12}$$

Since $\mathscr{B}_*(\alpha) \subseteq \mathscr{K}_*(\alpha)$, $\beta, \gamma \in \mathscr{K}_*(\alpha)$. Since $\alpha \in \mathbf{NIK}^*$, $\mathscr{K}_*$ is euclidean at α, hence

$$\gamma \in \mathscr{K}_*(\beta). \tag{13}$$

Since $\alpha \in K_*\mathbf{ICAU}$ and $\beta \in \mathscr{K}_*(\alpha)$ and, by Lemma A.1, $K_*\mathbf{ICAU} \subseteq K_*\mathbf{TRN}^*$, $\beta \in \mathbf{TRN}^*$. Thus, since $\beta \in B_*\mathbf{TB}^i$, $\beta \in K_*B_*\mathbf{TB}^i$. Hence, by (13), $\gamma \in B_*\mathbf{TB}^i$, contradicting (12). ■

PROOF OF THEOREM 3. (**QA** $\cap$ K_***ICAU** $\subseteq$ B_***TB**^{i} $\cap$ K_***TRN***) By Lemmas A.3 and A.7, **QA** $\cap$ K_***ICAU** $\subseteq$ B_***TB**^{i}. By Lemma A.1, K_***ICAU** $\subseteq$ K_***TRN***.
(B_***TB**^{i} $\cap$ K_***TRN*** $\subseteq$ **QA** $\cap$ K_***ICAU**). By Remark A.4, B_***TB**^{i} = B_*B_***TB**^{i} and by seriality of $\mathscr{B}_*$, B_*B_***TB**^{i} $\subseteq$ $\neg B_*\neg B_*$**TB**^{i} = **QA**. Thus B_***TB**^{i} $\subseteq$ **QA**. By Theorem 1, K_***TRN*** $\subseteq$ K_***ICAU**. ■

REMARK A.5. Although **QA** $\cap$ K_***ICAU** $\subseteq$ B_***TB**^{i} $\cap$ **TRN***, the converse is not true as the following example, illustrated in Figure A.1, shows. N = {1,2}, $\Omega = \{\alpha, \beta, \gamma\}$, $\mathscr{K}_1(\alpha) = \mathscr{K}_1(\gamma) = \{\alpha, \gamma\}$, $\mathscr{K}_1(\beta) = \mathscr{B}_1(\beta) = \{\beta\}$, $\mathscr{B}_1(\alpha) = \mathscr{B}_1(\gamma) = \{\gamma\}$, $\mathscr{K}_2(\alpha) = \mathscr{K}_2(\beta) = \{\alpha,\beta\}$, $\mathscr{K}_2(\gamma) = \mathscr{B}_2(\gamma) = \{\gamma\}$, $\mathscr{B}_2(\alpha) = \mathscr{B}_2(\beta) = \{\beta\}$. Thus, $\forall\, \omega \in \Omega$, $\mathscr{B}_*(\omega) = \{\beta,\gamma\}$ and $\mathscr{K}_*(\omega) = \Omega$.

Insert Figure A.1

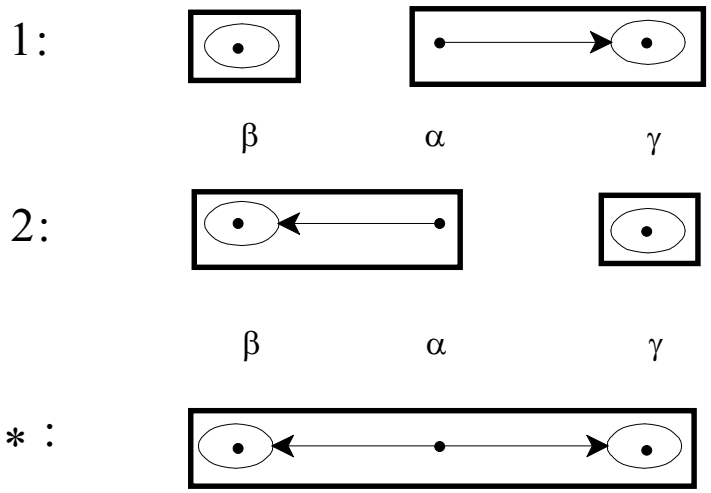

Figure A.1

Here we have that **TB**^{i} = $\{\beta, \gamma\}$, B_***TB**^{i} = Ω and **TRN*** = $\{\alpha\}$. Thus B_***TB**^{i} $\cap$ **TRN*** = $\{\alpha\}$. On the other hand, **ICAU** = K_***ICAU** = $\varnothing$.

The following lemma highlights an interesting property of common transparency.

LEMMA A.8. In a $\mathbb{KB}$-system (thus without assuming (Ax.9))

$$K_*\mathbf{TRN}^{*} \subseteq \mathbf{TRN}^{*}.$$

Proof. Let $\alpha \in K_* \mathbf{TRN}^*$ and suppose by contradiction that $\alpha \notin \mathbf{TRN}^*$. Then there exists an event E such that

$$\alpha \in B_* E \tag{14}$$

and

$$\alpha \in \neg K_* B_* E \tag{15}.$$

By (15) there exists a $\beta \in \mathcal{K}_*(\alpha)$ and a $\gamma \in \mathcal{B}_*(\beta)$ such that

$$\gamma \in \neg E \tag{16}.$$

Since $\beta \in \mathcal{K}_*(\alpha)$, by definition of $\mathcal{K}_*$ there exist $i \in N$ and $\delta \in \Omega$ such that

$$\delta \in \mathcal{K}_i(\alpha) \tag{17}$$

and

$$\beta \in \mathcal{K}_*(\delta) \tag{18}$$

[in the case where $\beta \in \mathcal{K}_i(\alpha)$ we can take $\delta = \beta$ and invoke secondary reflexivity of $\mathcal{K}_*$]. By seriality of $\mathcal{B}_i$ (cf. Remark 1), $\mathcal{B}_i(\alpha) \neq \emptyset$. Fix an arbitrary $\eta \in \mathcal{B}_i(\alpha)$. Since $\mathcal{B}_i(\alpha) \subseteq \mathcal{K}_i(\alpha)$,

$$\eta \in \mathcal{K}_i(\alpha) \tag{19}.$$

By (17), (19) and euclideanness of $\mathcal{K}_i$,

$$\delta \in \mathcal{K}_i(\eta) \tag{20}.$$

Since $\delta \in \mathcal{K}_i(\alpha) \subseteq \mathcal{K}_*(\alpha)$ and $\alpha \in K_* \mathbf{TRN}^*$, $\delta \in \mathbf{TRN}^*$. Thus by (ii) of Lemma 1, since $\beta \in \mathcal{K}_*(\delta)$ (cf. (18)) and $\gamma \in \mathcal{B}_*(\beta)$,

$$\gamma \in \mathcal{B}_*(\delta) \tag{21}.$$

Since $\eta \in \mathcal{B}_i(\alpha) \subseteq \mathcal{K}_i(\alpha) \subseteq \mathcal{K}_*(\alpha)$ and $\alpha \in K_* \mathbf{TRN}^*$, $\eta \in \mathbf{TRN}^*$. Hence by (20) [using the fact that $\mathcal{K}_i(\eta) \subseteq \mathcal{K}_*(\eta)$] and (21) and (ii) of Lemma 1, $\gamma \in \mathcal{B}_*(\eta)$. Thus, by definition of $\mathcal{B}_*$ [since $\eta \in \mathcal{B}_i(\alpha)$], $\gamma \in \mathcal{B}_*(\alpha)$. Hence, by (16), $\alpha \in \neg B_* E$, contradicting (14). ■

PROOF OF THEOREM 4. First we prove that

$$K_* \mathbf{TB}^i \subseteq \mathbf{QA} \cap K_* \mathbf{ICAU} \cap K_* \mathbf{TB}^* . \tag{22}$$

First note that $K_* \mathbf{TB}^i \subseteq B_* \mathbf{TB}^i$ and, as shown in the proof of Theorem 3, $B_* \mathbf{TB}^i \subseteq \mathbf{QA}$. Thus $K_* \mathbf{TB}^i \subseteq \mathbf{QA}$. Furthermore, since $\mathbf{TB}^i \subseteq \mathbf{TB}^*$, by Monotonicity of K_*, $K_* \mathbf{TB}^i \subseteq K_* \mathbf{TB}^*$. Finally, since $\mathbf{TB}^i \subseteq \mathbf{VB}^{i*}$, by Monotonicity of K_*, $K_* \mathbf{TB}^i \subseteq K_* \mathbf{VB}^{i*}$. By Proposition A.1 and Lemma A.3, $K_* \mathbf{VB}^{i*} = K_* \mathbf{ICAU}$.

Next we prove that

$$\mathbf{QA} \cap K_* \mathbf{ICAU} \cap K_* \mathbf{TB}^* \subseteq K_* \mathbf{TB}^i \tag{23}$$

Let $\alpha \in \mathbf{QA} \cap \mathrm{K}_*\mathbf{ICAU} \cap \mathrm{K}_*\mathbf{TB}^*$. By Theorems 2 and 3, $\alpha \in \mathrm{B}_*\mathbf{TB}^i \cap \mathbf{EQU}^*$. Hence $\alpha \in \mathrm{K}_*\mathbf{TB}^i$.

Thus, by (22) and (23),

$$\mathbf{QA} \cap \mathrm{K}_*\mathbf{ICAU} \cap \mathrm{K}_*\mathbf{TB}^* = \mathrm{K}_*\mathbf{TB}^i. \tag{24}$$

Next we prove that

$$\mathrm{K}_*\mathbf{TB}^i \subseteq \mathbf{CAU} \tag{25}$$

Let $\alpha \in \mathrm{K}_*\mathbf{TB}^i$. Fix arbitrary $i \in N$ and $E \subseteq \Omega$ and suppose that $\alpha \in \mathrm{B}_i E$. We need to show that $\alpha \in \mathrm{K}_i E$. Fix an arbitrary $\beta \in \mathscr{K}_i(\alpha)$. We have to prove that $\beta \in E$. Since $\alpha \in \mathrm{B}_i E \subseteq \mathrm{K}_i\mathrm{B}_i E$ and $\beta \in \mathscr{K}_i(\alpha)$,

$$\beta \in \mathrm{B}_i E. \tag{26}$$

Since $\alpha \in \mathrm{K}_*\mathbf{TB}^i$ and $\beta \in \mathscr{K}_i(\alpha) \subseteq \mathscr{K}_*(\alpha)$, $\beta \in \mathbf{TB}^i$. Hence, by (2), $\beta \in E$.

By (25) and Monotonicity of K_*, $\mathrm{K}_*\mathrm{K}_*\mathbf{TB}^i \subseteq \mathrm{K}_*\mathbf{CAU}$. By Remark A.4, $\mathrm{K}_*\mathrm{K}_*\mathbf{TB}^i = \mathrm{K}_*\mathbf{TB}^i$. Thus $\mathrm{K}_*\mathbf{TB}^i \subseteq \mathrm{K}_*\mathbf{CAU}$. It follows from this and (25) that

$$\mathrm{K}_*\mathbf{TB}^i \subseteq \mathbf{CAU} \cap \mathrm{K}_*\mathbf{CAU} \tag{27}$$

From (24) we get (by intersecting both sides with **QA**) that $\mathbf{QA} \cap \mathrm{K}_*\mathbf{ICAU} \cap \mathrm{K}_*\mathbf{TB}^* = \mathbf{QA} \cap \mathrm{K}_*\mathbf{TB}^i$ and from (27) $\mathbf{QA} \cap \mathrm{K}_*\mathbf{TB}^i \subseteq \mathbf{CAU} \cap \mathrm{K}_*\mathbf{CAU} \cap \mathbf{QA}$. Thus

$$\mathbf{QA} \cap \mathrm{K}_*\mathbf{ICAU} \cap \mathrm{K}_*\mathbf{TB}^* \subseteq \mathbf{CAU} \cap \mathrm{K}_*\mathbf{CAU} \cap \mathbf{QA} \tag{28}$$

We conclude the proof by showing that

$$\mathbf{CAU} \cap \mathrm{K}_*\mathbf{CAU} \cap \mathbf{QA} \subseteq \mathrm{K}_*\mathbf{TB}^i. \tag{29}$$

Let $\alpha \in \mathbf{CAU} \cap \mathrm{K}_*\mathbf{CAU} \cap \mathbf{QA}$. By Lemma A.3, $\mathbf{NIK}^\star = \Omega$. Since $\alpha \in \mathbf{CAU} \cap \mathrm{K}_*\mathbf{CAU}$, $\mathbf{NIK}^\star = \mathbf{NIB}^\star$. Thus $\alpha \in \mathbf{NIB}^\star \cap \mathbf{QA}$. By definition of **QA**, $\mathbf{NIB}^\star \cap \mathbf{QA} \subseteq \mathrm{B}_*\mathbf{TB}^i$. Thus $\alpha \in \mathrm{B}_*\mathbf{TB}^i$. Since $\alpha \in \mathbf{CAU} \cap \mathrm{K}_*\mathbf{CAU}$, $\alpha \in \mathrm{B}_*\mathbf{TB}^i$ if and only if $\alpha \in \mathrm{K}_*\mathbf{TB}^i$. ■

References

Aumann, Robert. 1976. 'Agreeing to disagree'. *Annals of Statistics,* 4: 1236-1239.

Battigalli, Paolo and Giacomo Bonanno. 1997. 'The logic of belief persistence'. *Economics and Philosophy*, 13: 39-59.

Ben-Porath, Elchanan. 1997. 'Rationality, Nash equilibrium and backward induction in perfect information games'. *Review of Economic Studies,* 64: 23-46.

Bonanno, Giacomo. 1996. 'On the logic of common belief'. *Mathematical Logic Quarterly*, 42: 305-311.

Bonanno, Giacomo and Klaus Nehring. 1996. 'Making sense of the common prior assumption under incomplete information', Working Paper, University of California, Davis.

Bonanno, Giacomo and Klaus Nehring. 1997a. 'On the logic and role of negative introspection of common belief', Working Paper, University of California, Davis [forthcoming in *Mathematical Social Sciences*].

Bonanno, Giacomo and Klaus Nehring, 1997b, 'Assessing the Truth Axiom under incomplete information', Working Paper, University of California, Davis [forthcoming in *Mathematical Social Sciences*].

Chellas, Brian. 1984. *Modal logic: an introduction*. Cambridge University Press.

Colombetti, Marco. 1993. 'Formal semantics for mutual beliefs'. *Artificial intelligence*, 62: 341-353.

Dekel, Eddie and Faruk Gul. 1997. 'Rationality and knowledge in game theory', in *Advances in Economic Theory, Seventh World Congress*. Kreps D. M. and K. F. Wallis (eds.). Cambridge University Press.

Fagin, Ronald, Joseph Halpern, Yoram Moses and Moshe Vardi. 1995, *Reasoning about knowledge*. MIT Press.

Geanakoplos, John. 1994. 'Common knowledge'. In *Handbook of game theory*, vol. 2, pp. 1437-1496. Robert Aumann and Sergiu Hart (eds). Elsevier.

Halpern, Joseph. 1991. 'The relationship between knowledge, belief and certainty'. *Annals of Mathematics and Artificial Intelligence*, 4: 301-322.

Halpern, Joseph and Yoram Moses. 1992, 'A guide to completeness and complexity for modal logics of knowledge and belief', *Artificial intelligence*, 54: 319-379.

Hintikka, Jaakko. 1962. *Knowledge and belief*. Cornell University Press.

van der Hoek, Wiebe. 1993. 'Systems for knowledge and belief'. *Journal of Logic and Computation*, 3: 173-195.

van der Hoek, Wiebe and J.-J. Ch. Meyer. 1995. *Epistemic logic for Artificial Intelligence and computer science*. Cambridge University Press.

Kraus, Sarit and Danile Lehmann. 1988. 'Knowledge, belief and time'. *Theoretical Computer Science,* 58: 155-174.

Lenzen, Wolfgang. 1978. 'Recent work in epistemic logic'. *Acta Philosophica Fennica*, 30: 1-220.

Lewis, David. 1969. *Convention: a philosophical study.* Harvard University Press.

Lismont, Luc and Philippe Mongin. 1994. 'On the logic of common belief and common knowledge', *Theory and Decision*, 37: 75-106.

Lismont, Luc and Philippe Mongin. 1995. 'Belief closure: a semantics for common knowledge for modal propositional logic'. *Mathematical Social Sciences*, 30: 127-153.

Morris, Steven. 1994. 'Trade with heterogeneous prior beliefs and asymmetric information', *Econometrica,* 62: 1327-1347.

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

Stalnaker, Robert. 1996. 'Knowledge, belief and counterfactual reasoning in games', *Economics and Philosophy,* 12: 133-163.

Stuart, H. 1997. 'Common belief of rationality in the finitely repeated Prisoners' Dilemma', *Games and Economic Behavior,* 19: 133-143.