Intersubjective Consistency of Knowledge and Belief

Giacomo Bonanno and Klaus Nehring

Common belief, in contrast to common knowledge, may exhibit epistemically counterintuitive properties. In a *regular* intersubjective system of knowledge and belief the logic of common belief and the relationship between common belief and common knowledge are the same as for the individuals. We characterize regularity in terms of properties of individual beliefs and study its implications for intersubjective consistency conditions on beliefs.

1 Introduction

We consider intersubjective 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 by the truth axiom (whatever is known is true). The implications of the truth axiom in an intersubjective context are strong: not only are individuals never mistaken in what they know, but this fact is moreover 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. Such "common knowledge of no error of knowledge" is a prominent example of an intersubjective consistency condition on epistemic states, and plays a role in the foundations of game theory. 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.

While common knowledge always obeys the formal logic of agents' knowledge (S5), in situations where some individuals believe other individuals to be mistaken in their beliefs, common belief may fail to obey

*The authors are grateful to three anonymous referees for helpful and constructive comments.

Formalizing the Dynamics of Information
Martina Faller, Stefan Kaufmann & Marc Pauly (eds.)

the formal logic of individual beliefs (KD45). In particular, the event that E is not commonly believed need not be itself commonly believed. Even more strikingly, whereas individuals always know what they believe, this is not necessarily so at the "common" level: it may well be that the agents fail to commonly know what they commonly believe. Call an intersubjective system of knowledge and belief *regular* when the logic of common belief and the relationship between common belief and common knowledge are the same as for the individuals. In this paper we characterize regularity in terms of properties of individual beliefs and study its implications for intersubjective consistency conditions on beliefs.

Integrated epistemic systems that jointly consider knowledge and belief have been studied in philosophy (Hintikka, 1962; Lenzen, 1978), artificial intelligence and computer science (Halpern, 1991; van der Hoek, 1993; van der Hoek and Meyer, 1995; Kraus and Lehmann, 1988; Lamarre and Shoham, 1994), 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 possibility of belief collapsing 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 a player's beliefs after she observes an unexpected move of an opponent. 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 analysis of the assumption of common belief in no error (of beliefs) which plays a crucial role in the justification of backward induction for interesting classes of perfect information games (cf. Ben-Porath 1997; Stalnaker 1996; Stuart 1997), and in the interpretation of the Common Prior Assumption under incomplete information (Bonanno and Nehring, 1998c).

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

The following example shows that common belief may be "ill-behaved" in the sense that it may fail to satisfy the same logic as individual beliefs.

Example 1 *Individual 1 knows that she is an illegitimate child. Individual 2, on the other hand, mistakenly believes that 1 is a legitimate child. He even believes this to be common belief between them. These beliefs are represented by state α in Figure 1, where the rectangles represent the knowledge partitions and the arrows represent the belief accessibility relations (common belief and knowledge are defined formally in Section 3). Let E be the event that represents the proposition "individual 1 is a legitimate child", that is, $E = \{\beta\}$. Then, at state α, although E is not commonly believed (because individual 1 believes herself to be illegitimate), it is not common belief that E is not commonly believed (due to individual 2's belief that E is common belief). Thus the property of* Negative Introspection of Common Belief, *denoted by NI^{CB} ($\neg B_* E \to B_* \neg B_* E$, where B_* is the common belief operator), fails to hold at α. Furthermore, at state β, while E is commonly believed, 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). Thus the property of* Awareness of Common Belief, *denoted by AW^{CB} ($B_* E \to K_* B_* E$, where K_* is the common knowledge operator), fails to hold at β.*

How can one understand the properties of common belief in terms of properties of individual beliefs? This question is answered by the first main result (Theorem 5.1) in terms of a condition on individual beliefs called *Caution about Common Belief* (CAU^{CB}). An agent is cautious about common belief at a 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 in his belief about what is commonly believed. Note that in Example 1, individual 2 fails to be cautious in this sense at either state. Theorem 5.1 shows that common knowledge of Negative Introspection of Common Belief ($K_*\mathrm{NI}^{CB}$), common knowledge of Awareness of Common Belief ($K_*\mathrm{AW}^{CB}$) and common knowledge of Caution about Common Belief ($K_*\mathrm{CAU}^{CB}$) are pairwise equivalent.

Example 2 *In the modification of Example 1 illustrated in Figure 2, at state α individual 2 still mistakenly believes that individual 1 is a legitimate child, but no longer believes this to be commonly believed:*

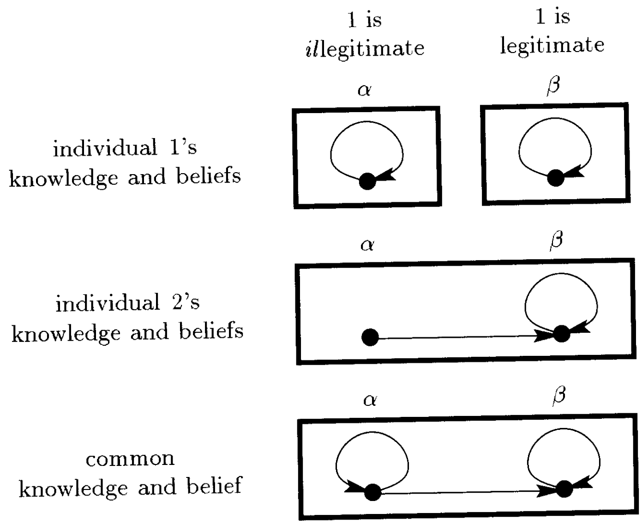

Figure 1

at state α he considers it possible (according to his beliefs) that the true state is β where individual 1 believes herself to be illegitimate. Here both agents are cautious about common belief at every state; as implied by Theorem 5.1, both Negative Introspection of Common Belief and Awareness of Common Belief are satisfied at every state. (Indeed, the only event that is commonly believed at any state is the universal event so that common belief and common knowledge coincide.)

Example 2 shows that common belief may be well-behaved even in the case where some individuals believe that others' beliefs are (or may be) wrong. Theorem 5.2 shows that, given common knowledge of Caution about Common Belief ($K_*\mathrm{CAU}^{CB}$), whenever individuals fail to have common belief in the correctness of each others' beliefs, this failure must in fact be commonly believed. We refer to this as *Disagreement* and to its complement as *Agreement*. Disagreement is a severe form of intersubjective inconsistency; in particular, it can be shown to characterize situations in which agents can make infinitely profitable bets with each other (see Bonanno and Nehring 1998a.)

In Example 2, common belief is not only well-behaved, it even coincides with common knowledge. When does this happen? The answer to this question is provided in Theorem 5.3 as follows. Let T^{CB} (for *Truth of Common Belief*) denote the property that, while some individuals may have incorrect beliefs, the group is never wrong *collectively* in the

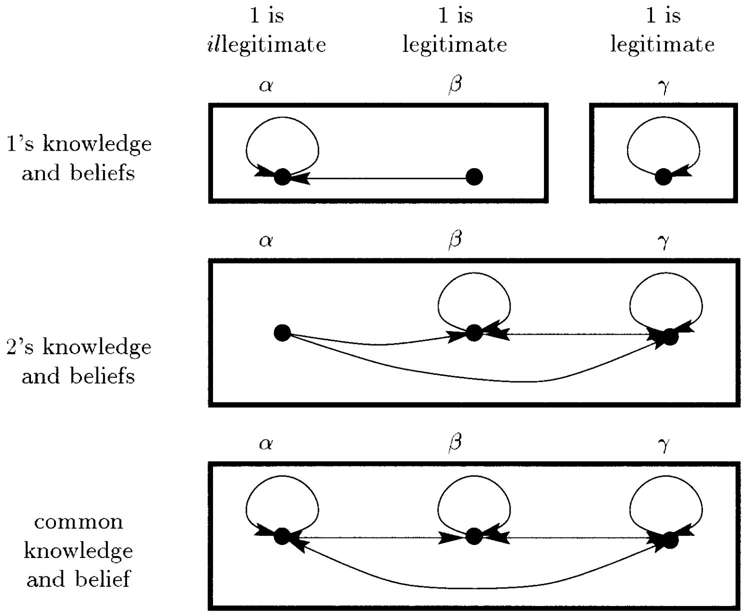

Figure 2

sense that whatever is commonly believed is in fact true. Let EQU^{CB} denote the *Equivalence of Common Belief and Common Knowledge.* Theorem 5.3 asserts the equivalence between common knowledge of EQU^{CB} and common knowledge of the conjunction of CAU^{CB} and T^{CB}.

If one adds Agreement to the assumptions of Theorem 5.3, belief collapses into knowledge for every individual (Theorem 5.4). 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 5.4 as follows: if it is common knowledge that individuals are cautious about common belief, and if the group is always correct (whatever is commonly believed is in fact true), 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.)

In view of the degeneracy uncovered by Theorem 5.4, the two conditions of Theorem 5.2 (Agreement and common knowledge of Caution about Common Belief) define the strongest plausible integrated intersubjective logic of knowledge and belief.

Figure 3 contains a summary of the results proved in this paper.[1]

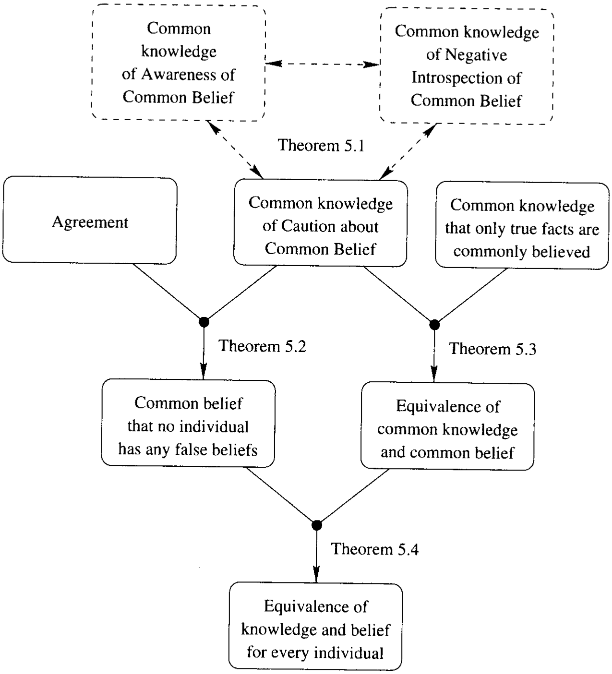

Figure 3

The following section defines the formal systems of knowledge and belief in the event-based framework which is common in game theory and economics. Section 4 clarifies the relationship between the event-based and the axiomatic (syntactic) approach. Section 5 contains the results, while Section 6 concludes by providing an assessment of the three fundamental conditions on individual beliefs as conditions of intersubjective rationality.

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

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

$i \in N$ 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. These operators are assumed to satisfy the following properties ($\neg$ denotes complement): $\forall i \in N$, $\forall E, F \in 2^\Omega$,

(Ax.1)	Necessity:	$B_i\Omega = \Omega$ and $K_i\Omega = \Omega$.
(Ax.2)	Monotonicity:	if $E \subseteq F$ then $K_iE \subseteq K_iF$ and $B_iE \subseteq B_iF$
(Ax.3)	Conjunction:	$K_i(E \cap F) = K_iE \cap K_iF$ and $B_i(E \cap F) = B_iE \cap B_iF$
(Ax.4)	Truth Axiom for knowledge:	$K_iE \subseteq E$
(Ax.5)	Consistency of beliefs:	$B_iE \subseteq \neg B_i \neg E$
(Ax.6)	Positive Introspection:	$K_iE \subseteq K_iK_iE$ and $B_iE \subseteq B_iB_iE$
(Ax.7)	Negative Introspection:	$\neg K_iE \subseteq K_i\neg K_iE$ and $\neg B_iE \subseteq B_i\neg B_iE$
(Ax.8)	Priority of knowledge:	$K_iE \subseteq B_iE$
(Ax.9)	Awareness of own beliefs:	$B_iE \subseteq K_iB_iE$

We call a tuple $\langle \Omega, N, \{B_i\}_{i\in N}, \{K_i\}_{i\in N}\rangle$ that satisfies (Ax.1)–(Ax.9) a *KB-system.* Thus the logic of knowledge is S5, the logic of belief is KD45 and (Ax.8) and (Ax.9) establish the relationship between knowledge and belief.[2]

We shall denote by $\mathcal{B}_i : \Omega \to 2^\Omega$ (respectively, $\mathcal{K}_i : \Omega \to 2^\Omega$) the *possibility correspondence* associated with the belief operator B_i (resp. the knowledge operator K_i). Thus, $\forall \alpha \in \Omega$, $\mathcal{B}_i(\alpha) = \{\omega \in \Omega : \alpha \in \neg B_i \neg\{\omega\}\}$ and $\mathcal{K}_i(\alpha) = \{\omega \in \Omega : \alpha \in \neg K_i \neg \{\omega\}\}$.[3]

[2] Note that positive and negative introspection of belief are redundant, since they can be deduced from the other properties (cf. Kraus and Lehmann 1988.)

[3] In the philosophy and AI literature it is more common to express the (Kripkean) semantics in terms of accessibility relations. However, the notions of accessibility relation and possibility correspondence are entirely equivalent. Given an accessibility relation R on Ω, one defines the corresponding possibility correspondence P as follows: $P(\omega) = \{\omega' \in \Omega : \omega R \omega'\}$. Conversely, given a possibility correspondence P one obtains the associated accessibility relation as follows: $\omega R\omega'$ if and only if $\omega' \in P(\omega)$.
It is well-known that, $\forall \omega \in \Omega$, $\forall E \subseteq \Omega$, $\omega \in B_iE$ (resp. $\omega \in K_iE$) if and only if $\mathcal{B}_i(\omega) \subseteq E$ (resp. $\mathcal{K}_i(\omega) \subseteq E$). Furthermore, B_i satisfies consistency if and only if $\mathcal{B}_i$ is serial ($\forall \omega \in \Omega$, $\mathcal{B}_i(\omega) \neq \emptyset$), K_i satisfies the truth axiom if and only if $\mathcal{K}_i$ is reflexive ($\forall \omega \in \Omega$, $\omega \in \mathcal{K}_i(\omega)$), B_i satisfies positive introspection if and only if $\mathcal{B}_i$ is transitive ($\forall \alpha, \beta \in \Omega$, if $\beta \in \mathcal{B}_i(\alpha)$ then $\mathcal{B}_i(\beta) \subseteq \mathcal{B}_i(\alpha)$) and it satisfies negative introspection if and only if $\mathcal{B}_i$ is euclidean ($\forall \alpha, \beta \in \Omega$, if $\beta \in \mathcal{B}_i(\alpha)$ then $\mathcal{B}_i(\alpha) \subseteq \mathcal{B}_i(\beta)$). The

The *common belief* operator $B_* : 2^\Omega \to 2^\Omega$ and the *common knowledge* operator $K_* : 2^\Omega \to 2^\Omega$ are defined as follows. First, $\forall E \subseteq \Omega$ let $B_e E = \bigcap_{i\in N} B_i E$ and $K_e E = \bigcap_{i\in N} K_i E$, that is, $B_e E$ (resp. $K_e E$) is the event that everybody believes (resp. knows) E. Let $B_e^1 E = B_e E$ (resp. $K_e^1 E = K_e E$) and for $r > 1$ let $B_e^r E = B_e B_e^{r-1} E$ (resp. $K_e^r E = K_e K_e^{r-1} E$). Then the event that E is commonly believed (resp. commonly known) is defined by:

$$B_* E = \bigcap_{r=1}^{\infty} B_e^r E \qquad \left(resp. \;\; K_* E = \bigcap_{r=1}^{\infty} K_e^r E \right)$$

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[4] that $\mathcal{B}_*$ coincides with 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 $\langle i_1, ..., i_m \rangle$ in N and a sequence $\langle \eta_0, \eta_1, ..., \eta_m \rangle$ in Ω such that: (i) $\eta_0 = \alpha$, (ii) $\eta_m = \beta$ and (iii) $\forall k = 0, ..., m-1$, $\eta_{k+1} \in \mathcal{B}_{i_{k+1}}(\eta_k)$.

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

Example 3 *In a KB-system not all the properties of individual beliefs / knowledge are inherited by common belief / knowledge. In particular, Negative Introspection of Common Belief* ($\neg B_* E \subseteq B_* \neg B_* E$) *and Awareness of Common Belief (the counterpart to (Ax.9):* $B_* E \subseteq K_* B_* E$) *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 1 according to the following convention which will be used throughout the paper. States are denoted by points and a (individual or common)* knowledge *possibility correspondence* $\mathcal{K} : \Omega \to 2^\Omega$ *is represented by rectangles which partition the set of states, while a (individual or common)* belief *possibility correspondence* $\mathcal{B} : \Omega \to 2^\Omega$ *is*

same is true of K_i and $\mathcal{K}_i$. It is also well-known (cf. van der Hoek 1993) that (Ax.8) is equivalent to $\mathcal{B}_i(\omega) \subseteq \mathcal{K}_i(\omega)$, $\forall \omega \in \Omega$, and (Ax.9) is equivalent to the following: $\forall \alpha, \beta, \gamma \in \Omega$, if $\beta \in \mathcal{K}_i(\alpha)$ and $\gamma \in \mathcal{B}_i(\beta)$ then $\gamma \in \mathcal{B}_i(\alpha)$.

[4]See, for example, Bonanno (1996); Fagin et al. (1995); Halpern and Moses (1992); Lismont and Mongin (1994). 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.

represented by arrows as follows: $\omega' \in \mathcal{B}(\omega)$ if and only if there is an arrow from ω to ω'. Let $E = \{\beta\}$. Then $B_* E = \{\beta\}$, $\neg B_* E = \{\alpha\}$ and $B_* \neg B_* E = \emptyset$. Thus Negative Introspection of common belief fails: $\neg B_* E \not\subseteq B_* \neg B_* E$. Furthermore, $K_* B_* E = \emptyset$. Hence also Awareness of Common Belief fails: $B_* E \not\subseteq K_* B_* E$.

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

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

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

$$E \leftrightarrow F \stackrel{def}{=} (E \to F) \cap (F \to E).$$

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

4 Event-based versus syntactic approach

In this paper we have employed the event-based approach which is common in game theory and economics. On the other hand, the philosophy and computer science literature usually relies on the axiomatic-syntactic approach. In this section we clarify the relationship between the two.

The axiomatic approach starts with a propositional language augmented with (individual and common) belief and knowledge operators. With abuse of notation, we shall denote these operators by B_i, B_*, K_i and K_*. Furthermore, we shall denote formulae by capital letters, such as E and F and use the symbols '$\to$' and '$\leftrightarrow$' for the 'if ... then...' and 'if and only if' operators on formulae, respectively. Again with abuse of notation we denote by KB the system of multimodal normal logic where the K_i operators satisfy the S5 logic (Truth, Positive and Negative Introspection), the B_i operators satisfy the KD45 logic (Consistency, Positive and Negative Introspection) and individual knowledge and belief are connected by the two axioms corresponding to (Ax.8) and (Ax.9), namely $K_i E \to B_i E$ and $B_i E \to K_i B_i E$. Furthermore, the usual axioms for common belief and common knowledge are assumed.[5]

[5] See, for example, Bonanno (1996); Fagin et al. (1995); Halpern and Moses (1992); van der Hoek (1993); van der Hoek and Meyer (1995); Kraus and Lehmann (1988); Lismont and Mongin (1994).

Our characterization results are typically in the form of equality between two common knowledge events, say $K_*E = K_*F$. Now, equality of the events K_*E and K_*F is equivalent to validity of the formula $K_*E \leftrightarrow K_*F$, that is, its truth set is the universal set Ω. (However, in general, the formula $E \leftrightarrow F$ is *not* valid.) Thus, *by completeness of the system* KB [6], a result of the form $K_*E = K_*F$ in the event-based approach corresponds to the following syntactic result: $KB \vdash K_*E \leftrightarrow K_*F$ (that is, the formula $K_*E \leftrightarrow K_*F$ is a theorem of system KB). In turn, this is equivalent to saying that F is a theorem of $KB + E$ (the system obtained by adding E to KB) and E is a theorem of $KB + F$.

5 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	$\mathbf{NI}^{CB} = \bigcap\limits_{E \in 2^\Omega} (\neg B_*E \to B_*\neg B_*E)$
Awareness of Common Belief	$\mathbf{AW}^{CB} = \bigcap\limits_{E \in 2^\Omega} (B_*E \to K_*B_*E)$
Caution about Common Belief	$\mathbf{CAU}^{CB} = \bigcap\limits_{i \in N} \bigcap\limits_{E \in 2^\Omega} (B_iB_*E \to K_iB_*E)$

Thus $\omega \in \mathbf{NI}^{CB}$ if and only if, for every event E, if $\omega \in \neg B_*E$ then $\omega \in B_*\neg B_*E$; $\omega \in \mathbf{AW}^{CB}$ if and only if, for every event E, if $\omega \in B_*E$ then $\omega \in K_*B_*E$; $\omega \in \mathbf{CAU}^{CB}$ if and only if, for every individual i and every event E, if $\omega \in B_iB_*E$ then $\omega \in K_iB_*E$.

$\mathbf{NI}^{CB}$ is the analogue, for common belief, of (Ax.7) for individual beliefs, while $\mathbf{AW}^{CB}$ is the analogue, for common belief and knowledge, of property (Ax.9) of individual beliefs / knowledge. $\mathbf{CAU}^{CB}$, 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.

The following proposition gives the properties of the possibility correspondences that characterize these three events. For example, in Figure 1, $\mathbf{NI}^{CB} = \{\beta\}$, $\mathbf{AW}^{CB} = \{\alpha\}$, $\mathbf{CAU}^{CB} = \emptyset$. That $\mathbf{CAU}^{CB} = \emptyset$

[6] See Kraus and Lehmann (1988, Theorem 2.6, p. 160) and van der Hoek and Meyer (1995, Theorem 2.13.5, p. 93).

can be seen directly by noting that at every state individual 2 believes that $E = \{\beta\}$ is common belief ($B_2 B_* E = \Omega$), but she does not know this, since $\mathcal{K}_2(\omega) = \Omega$, for every ω, while $B_* E = \{\beta\}$.

All the proofs are given in the appendix. [7]

Proposition 1 *The following holds for every $\alpha \in \Omega$:*

(i) $\alpha \in$ $\mathbf{NI}^{CB}$ 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) $\alpha \in$ $\mathbf{AW}^{CB}$ if and only if, $\forall \alpha, \beta, \gamma \in \Omega$, if $\beta \in \mathcal{K}_*(\alpha)$ and $\gamma \in \mathcal{B}_*(\beta)$ then $\gamma \in \mathcal{B}_*(\alpha)$.

(iii) $\alpha \in$ $\mathbf{CAU}^{CB}$ if and only if $\forall i \in N$, $\forall \beta, \gamma \in \Omega$, 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 $\gamma \in \mathcal{B}_*(\delta)$.

Although, typically, $\mathbf{NI}^{CB} \neq \mathbf{AW}^{CB}$, $\mathbf{AW}^{CB} \neq \mathbf{CAU}^{CB}$ and $\mathbf{NI}^{CB} \neq \mathbf{CAU}^{CB}$, the three properties of Negative Introspection of Common Belief, Awareness of Common Belief and Caution about Common Belief coincide *when commonly known.*

Theorem 5.1 $K_*\mathbf{NI}^{CB} = K_*\mathbf{AW}^{CB} = K_*\mathbf{CAU}^{CB}$.

Definition 1 *A state α is* regular *if at α any of the events $\mathbf{NI}^{CB}$, $\mathbf{AW}^{CB}$ or $\mathbf{CAU}^{CB}$ are common knowledge (e.g. if $\alpha \in K_*\mathbf{CAU}^{CB}$); similarly, a KB-system is regular if any of those events coincides with the universal set (e.g. if $\mathbf{CAU}^{CB} = \Omega$).*

Example 4 *None of the above properties of beliefs / knowledge embody agreement-type restrictions on individual beliefs, as the following example, illustrated in Figure 4, shows: $N = \{1,2\}$, $\Omega = \{\alpha, \beta\}$, $\mathcal{K}_1(\omega) = \mathcal{K}_2(\omega) = \{\alpha, \beta\}$ $\forall \omega \in \Omega$, $\mathcal{B}_1(\alpha) = \mathcal{B}_1(\beta) = \{\alpha\}$, $\mathcal{B}_2(\alpha) = \mathcal{B}_2(\beta) = \{\beta\}$. Thus $\mathcal{K}_*(\omega) = \mathcal{B}_*(\omega) = \Omega$ $\forall \omega \in \Omega$. Here $\mathbf{NI}^{CB} = \mathbf{AW}^{CB} = \mathbf{CAU}^{CB} = \Omega$ 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 $E = \{\alpha\}$ while individual 2 believes $\neg E$.*

We now introduce two more properties of beliefs:

Truth of individual beliefs	$\mathbf{T}^{IB} = \bigcap_{i \in N} \bigcap_{E \in 2^\Omega} (B_i E \to E)$
Disagreement	$\mathbf{DIS} = B_* \neg B_* \mathbf{T}^{IB}$

[7] All the results are proved for weaker systems than KB-systems, in particular the truth axiom for knowledge is not required for any of the results. To simplify the exposition we have adopted in the text the standard KB-systems discussed in the literature.

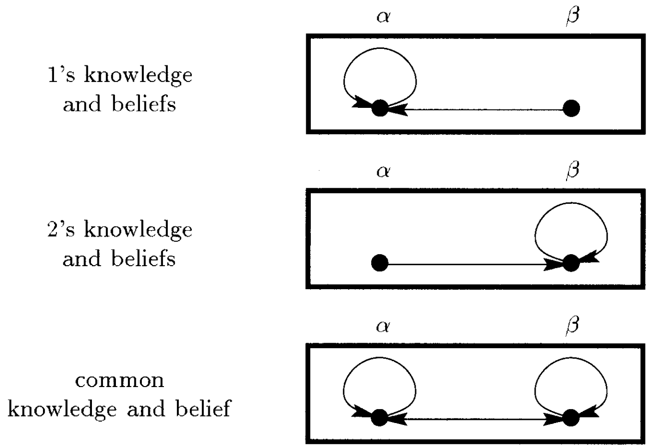

Figure 4

Thus $\alpha \in \mathbf{T}^{IB}$ 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 that $\alpha \in \mathbf{T}^{IB}$ if and only if, $\forall i \in N$, $\alpha \in \mathcal{B}_i(\alpha)$. The event $B_*\mathbf{T}^{IB}$ captures a property known in the game theoretic literature as *common belief in no error* (cf. Ben-Porath 1997, Stalnaker 1996, Stuart 1997.) Disagreement is defined as common belief in the *lack* of common belief in no error. We refer to its negation, $\neg\mathbf{DIS}$, as Agreement. Thus at state α there is Agreement if and only if for *some* $\beta \in \mathcal{B}_*(\alpha)$, $\beta \in B_*\mathbf{T}^{IB}$.[8]

Theorem 5.2 $\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} = B_*\mathbf{T}^{IB} \cap K_*\mathbf{AW}^{CB}$.

Thus regularity and Agreement ensure strong intersubjective consistency properties; arguably the strongest plausible (see remark after Theorem 5.4).

For the next result we need to introduce two more properties:

Truth of Common Belief	$\mathbf{T}^{CB} = \bigcap_{E \in 2^\Omega} (B_*E \to E)$
Equivalence of Common Belief and Common Knowledge	$\mathbf{EQU}^{CB} = \bigcap_{E \in 2^\Omega} (B_*E \leftrightarrow K_*E)$

[8] Bonanno and Nehring (1998a) motivate the notion of Agreement as $\neg\mathbf{DIS}$ in two distinct but equivalent ways: (1) the absence of "agreeing to disagree" about "union consistent" qualitative belief indices (a generalization of the Agreement property introduced by Aumann (1976), and (2) the absence of *unbounded* gains from betting (assuming moderately risk-averse preferences).

$\mathbf{T}^{CB}$ captures the property that only true facts are commonly believed ($\omega \in \mathbf{T}^{CB}$ if and only if, for every event E, if $\omega \in B_*E$ then $\omega \in E$) while $\mathbf{EQU}^{CB}$ is the property that common belief and common knowledge coincide ($\omega \in \mathbf{EQU}^{CB}$ if and only if, for every event E, if $\omega \in B_*E$ then $\omega \in K_*E$ and *vice versa)*. [9]

If one adds to regularity the hypothesis that it is common knowledge that only true facts are commonly believed, one obtains the collapse of common belief into common knowledge.

Theorem 5.3 $K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB} = K_*\mathbf{EQU}^{CB}$.

Remark 1 *In all the theorems common knowledge of the events under consideration is crucial. For instance, in Figure 1, at state β, while there is common knowledge of the Truth of Common Belief ($\beta \in K_*\mathbf{T}^{CB}$), there is only Caution about Common Belief but not common knowledge of it ($\beta \in \mathbf{CAU}^{CB}$ but $\beta \notin K_*\mathbf{CAU}^{CB}$); in line with the above theorems, Awareness of Common Belief fails at that state ($\beta \notin \mathbf{AW}^{CB}$) and thus common knowledge and common belief fail to coincide ($\beta \notin \mathbf{EQU}^{CB}$). Similar counterexamples can be constructed in each case.*

Our last theorem shows that putting together the three conditions of Agreement, common knowledge of Caution about Common Belief and common knowledge that only true facts are commonly believed leads to the collapse of belief into knowledge *for every individual*. The theorem moreover states that such collapse of individual belief into knowledge is also equivalent to the hypothesis of common knowledge that every individual has correct beliefs. Let

Equivalence of belief and knowledge for every individual $\quad \mathbf{EQU}^{IB} = \bigcap_{i \in N} \bigcap_{E \in 2^\Omega} (B_iE \leftrightarrow K_iE)$

Thus $\alpha \in \mathbf{EQU}^{IB}$ if and only if, for every individual i and event E, at α individual i believes E ($\alpha \in B_iE$) if and only if she knows E ($\alpha \in K_iE$).

Theorem 5.4 $\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB} = K_*\mathbf{EQU}^{IB} \cap \neg\mathbf{DIS} = K_*\mathbf{T}^{IB}$.

In view of the degeneracy uncovered by Theorem 5.4, the two conditions of Theorem 5.2 (Agreement and common knowledge of Caution about Common Belief) define the strongest plausible integrated intersubjective logic of knowledge and belief.

[9] It is straightforward that $\alpha \in \mathbf{T}^{CB}$ if and only if $\alpha \in \mathcal{B}_*(\alpha)$ and $\alpha \in \mathbf{EQU}^{CB}$ if and only if $\mathcal{B}_*(\alpha) = \mathcal{K}_*(\alpha)$. In the example of Figure 1, $\mathbf{T}^{CB} = \Omega$ and $\mathbf{EQU}^{CB} = \{\alpha\}$. In the example of Figure 4, $\mathbf{T}^{CB} = \mathbf{EQU}^{CB} = \Omega$.

6 Conclusion

The analysis of this paper has spanned the intersubjective gap between belief and knowledge by three intersubjective consistency conditions: Agreement ($\neg\mathbf{DIS}$), common knowledge of Caution abut Common Belief ($K_*\mathbf{CAU}^{CB}$), and common knowledge that only true facts are commonly believed ($K_*\mathbf{T}^{CB}$). 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" ($B_*\mathbf{T}^{IB}$). *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[10]: every agent is willing to underwrite epistemically every other agent's beliefs to the extent that he is certain of 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 individuals took it for granted that individual 1 was a legitimate child; however, after having a private look at her birth certificate, individual 1 discovers to her great surprise that she is an illegitimate child. Formally, this can be described in a two-state universe augmenting Figure 1. The original Figure 1 now describes the individuals' beliefs *at date 1*, after the (one-sided) inspection of the birth certificate. We now augment Figure 1 by adding two new "epistemic agents" describing the individuals' beliefs at date 0; the beliefs of each of these two new agents are a replica of individual 2's beliefs in the original Figure 1 (thus of 2's beliefs at date 1). At date 0, both individuals' beliefs coincide and therefore satisfy any meaningful intersubjective rationality condition. The individuals' beliefs at date 1, in particular individual 1's certainty of the falsity of her 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.

Agreement ($\neg B_*\neg B_*\mathbf{T}^{IB}$) can be viewed as an appropriate weakening of common belief in no error ($B_*\mathbf{T}^{IB}$) not subject to an objection of this kind: 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

[10] 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: $B_i(B_iE \rightarrow E) = \Omega$). Secondary reflexivity is implied by Negative Introspection.

(cf. Bonanno and Nehring 1998a.) If any intersubjective consistency can stake a claim on rationality, it would seem to be Agreement: its equivalence to the absence of unbounded gains form betting (cf. Bonanno and Nehring 1998a) lends it strong normative appeal. Moreover, it is not subject to the contingencies of history, as it restricts agents' beliefs only when they are jointly commonly believed. In Example 1, for instance, only trivial beliefs are jointly commonly believed.[11]It would even make perfect sense to require Agreement in a game after a player observes an unexpected move by an opponent!

Common knowledge of Caution about Common Belief $K_*\mathbf{CAU}^{CB}$, 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 it for granted that she was a legitimate child (and believed that she would continue to take it for granted), recognizing the possibility (in terms of knowledge) that she might live to change her mind. On the other hand, while not categorical, Caution about Common Belief seems highly reasonable as a constraint on how individuals "initially" construct their intersubjective beliefs, prior to the receipt of specific private information (but incompletely informed of each other's beliefs), for example prior to the actual play of a game. This would be sufficient to justify the striking Stalnaker-Stuart justification of non-cooperative play in the repeated prisoners' dilemma game (Stalnaker, 1996; Stuart, 1997).

Finally, common knowledge of the truth of common belief ($K_*\mathbf{T}^{CB}$) has the flavor of an empirical rather than a rationality assumption. As the latter, it seems implausible; note, for example, that applied to a group of one, it coincides with Equivalence of knowledge and belief. [12] In view of Theorem 5.2, and taking into account the plausibility of both $\neg\mathbf{DIS}$ and $K_*\mathbf{CAU}^{CB}$, it seems implausible even as an empirical assumption, in spite of the appeal to the *prima facie* reasonable intuition that a group's beliefs may enjoy higher epistemic dignity than any individual's beliefs.

Appendix

For the sake of generality, all the proofs will be given for the weaker systems obtained by replacing (Ax.4) (truth axiom for knowledge) with the following weaker axiom: $\forall i \in N, \forall E \subseteq \Omega$,

[11]Note that, while individual 2's beliefs about the event $\{\beta\}$ are common belief, 1's beliefs about $\{\beta\}$ are not.

[12]By contrast, both $\neg\mathbf{DIS}$ and $\mathbf{CAU}^{CB}$ are automatically satisfied in this case.

$$(Ax.4') \qquad\qquad K_i K_* E \subseteq K_* E$$

Indeed, as pointed out below, some results hold even without assuming (Ax.4′). Throughout this appendix, the systems considered are those obtained from KB-systems by weakening (Ax.4) to (Ax.4′). By *weak KB-systems*, on the other hand, we shall mean systems obtained from KB-systems by *dropping* (Ax.4) (without replacing it with another axiom, in particular, without assuming (Ax.4′)).

Proof of Proposition 1.

(i) is well-known (see Chellas 1984) and (ii) follows from Theorem 4.3 (c) in van der Hoek (1993, p. 183). Thus we shall only prove (iii). Let P be the property stated there. First we show that if P is not satisfied at α then $\alpha \notin \mathbf{CAU}^{CB}$. Suppose P does not hold at α. Then there exist $i \in N$ and $\beta, \gamma \in \Omega$ 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 $E = \{\omega \in \Omega : \omega \in \mathcal{B}_*(\omega') \ for\ some\ \omega' \in \mathcal{B}_i(\alpha)\}$. Then $\gamma \notin E$, and, by construction, $\alpha \in B_i B_* E$. Since $\gamma \in \mathcal{B}_*(\beta)$ and $\gamma \notin E$, $\mathcal{B}_*(\beta) \nsubseteq E$, that is, $\beta \notin B_* E$. Hence, since $\beta \in \mathcal{K}_i(\alpha)$, $\alpha \notin K_i B_* E$. Thus, since $\alpha \in B_i B_* E$, $\alpha \notin (B_i B_* E \rightarrow K_i B_* E)$. It follows that $\alpha \notin \mathbf{CAU}^{CB}$. Next we show that if $\alpha \notin \mathbf{CAU}^{CB}$ then P is not satisfied at α. Suppose that $\alpha \notin \mathbf{CAU}^{CB}$. 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 does not hold at α. □

Proof of Theorem 5.1.

The proof of Theorem 5.1 will be carried out in three steps. The first step is given by Lemma 1, which is true in weak systems (that is, without assuming (Ax.4′)). The second step is given by Lemma 2, which is a restatement of Theorem 5.1 for weak systems that satisfy an additional property. The third and final step is given by Lemma 4 which shows that this additional property is equivalent to (Ax.4′).

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

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

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 2 *For every $\alpha \in \Omega$, $\alpha \in \mathbf{VB}^{i*}$ if and only if, $\forall i \in N$, $\forall \gamma \in \mathcal{B}_*(\alpha)$, $\exists \delta \in \mathcal{B}_i(\alpha)$ such that $\gamma \in \mathcal{B}_*(\delta)$. For a proof see Lemma 2 in Bonanno and Nehring (1998a).*[13]

Lemma 1 *In a weak system (thus without assuming (Ax.4′)) the following holds: $K_*\mathbf{NI}^{CB} \subseteq K_*\mathbf{VB}^{i*} \subseteq K_*\mathbf{CAU}^{CB} \subseteq K_*\mathbf{AW}^{CB}$.*

Proof. ($K_*\mathbf{NI}^{CB} \subseteq K_*\mathbf{VB}^{i*}$). First we show that $\mathbf{NI}^{CB} \subseteq \mathbf{VB}^{i*}$. Let $\alpha \in \mathbf{NI}^{CB}$. Fix an arbitrary $i \in N$ and $E \subseteq \Omega$. We want to show that $\alpha \in (B_i B_* E \to B_* E)$, or, equivalently, that $\alpha \in (\neg B_* E \to \neg B_i B_* E)$. Since $\alpha \in \mathbf{NI}^{CB}$, $\alpha \in (\neg B_* E \to B_* \neg B_* E)$. 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.5), $B_i \neg B_* E \subseteq \neg B_i B_* E$. Thus $B_* \neg B_* E \subseteq \neg B_i B_* E$. Hence $\alpha \in \neg B_i B_* E$. Thus $\mathbf{NI}^{CB} \subseteq \mathbf{VB}^{i*}$. By Monotonicity of K_*, it follows that $K_*\mathbf{NI}^{CB} \subseteq K_*\mathbf{VB}^{i*}$.
($K_*\mathbf{VB}^{i*} \subseteq K_*\mathbf{CAU}^{CB}$). Let $\alpha \in K_*\mathbf{VB}^{i*}$ and fix an arbitrary $\beta \in \mathcal{K}_*(\alpha)$. We want to show that $\beta \in \mathbf{CAU}^{CB}$. Fix arbitrary $i \in N$ and $E \subseteq \Omega$ such that $\beta \in B_i B_* E$. Fix an arbitrary $\gamma \in \mathcal{K}_i(\beta)$. We need to show that $\gamma \in B_* E$. Since $B_i B_* E \subseteq K_i B_i B_* E$ (cf. Ax.9), $\beta \in K_i B_i B_* E$ hence $\gamma \in B_i B_* E$. By definition of $\mathcal{K}_*$, since $\beta \in \mathcal{K}_*(\alpha)$ and $\gamma \in \mathcal{K}_i(\beta)$, $\gamma \in \mathcal{K}_*(\alpha)$. Thus, since $\alpha \in K_*\mathbf{VB}^{i*}$, $\gamma \in \mathbf{VB}^{i*}$. Hence, since $\gamma \in B_i B_* E$, $\gamma \in B_* E$.
($K_*\mathbf{CAU}^{CB} \subseteq K_*\mathbf{AW}^{CB}$). Let $\alpha \in K_*\mathbf{CAU}^{CB}$. Fix an arbitrary $\beta \in \mathcal{K}_*(\alpha)$. We want to show that $\beta \in \mathbf{AW}^{CB}$. Fix arbitrary $E \subseteq \Omega$ such that $\beta \in B_* E$. We need to show that $\beta \in K_* B_* E$. Fix arbitrary sequences $\langle i_1, ... i_m \rangle$ in N and $\langle \eta_0, \eta_1, ..., \eta_m \rangle$ in Ω such that $\eta_0 = \beta$, and, for every $k = 1, \ldots, m$, $\eta_k \in \mathcal{K}_{i_k}(\eta_{k-1})$. We need to show that $\eta_m \in B_* E$. First of all, note that, since $\beta \in \mathcal{K}_*(\alpha)$, by definition of $\mathcal{K}_*$, $\eta_k \in \mathcal{K}_*(\alpha)$ for all $k = 0, \ldots, m$. Hence, since $\alpha \in K_*\mathbf{CAU}^{CB}$,

$$\forall k = 0, ..., m, \quad \eta_k \in \mathbf{CAU}^{CB} \tag{5.1}$$

Since $\eta_0 = \beta \in B_* E$ and, by definition of B_*, $B_* E \subseteq B_{i_1} B_* E$, $\eta_0 \in B_{i_1} B_* E$. Hence, by (5.1), $\eta_0 \in K_{i_1} B_* E$. Thus, since $\eta_1 \in \mathcal{K}_{i_1}(\eta_0)$, $\eta_1 \in B_* E$. Since $B_* E \subseteq B_{i_2} B_* E$, $\eta_1 \in B_{i_2} B_* E$. Hence, by (5.1), $\eta_1 \in K_{i_2} B_* E$. Thus, since $\eta_2 \in \mathcal{K}_{i_2}(\eta_1)$, $\eta_2 \in B_* E$. Repeating this argument m times we get that $\eta_m \in B_* E$. □

[13] There the event $\mathbf{VB}^{i*}$ is denoted by $\mathbf{T}_{CB}$ and the possibility correspondence $\mathcal{B}_*$ (resp. $\mathcal{B}_i$) is denoted by $\mathcal{I}_*$ (resp. $\mathcal{I}_i$).

Remark 3 *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 (**NIK*** stands for "Negative Introspection of common knowledge")

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

Remark 4 *Analogously to (i) of Proposition 1, it can be shown that, $\forall \beta \in \Omega$, $\beta \in \mathbf{NIK}^*$ if and only if $\mathcal{K}_*$ is euclidean at β, that is, $\forall \gamma, \delta \in \mathcal{K}_*(\beta)$, $\delta \in \mathcal{K}_*(\gamma)$.*

Lemma 2 *In a weak system that satisfies $K_* \mathbf{NIK}^* = \Omega$ the following holds:*

$$K_* \mathbf{NI}^{CB} = K_* \mathbf{VB}^{i*} = K_* \mathbf{CAU}^{CB} = K_* \mathbf{AW}^{CB}$$

Lemma 2 follows directly from Lemma 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 3 *In a weak system (thus without assuming (Ax.4′)) the following holds: $K_* \mathbf{NIK}^* \cap K_* \mathbf{AW}^{CB} \subseteq K_* \mathbf{NI}^{CB}$.*

Proof. Let $\alpha \in K_* \mathbf{NIK}^* \cap K_* \mathbf{AW}^{CB}$ and fix an arbitrary $\beta \in \mathcal{K}_*(\alpha)$. We need to show that $\beta \in \mathbf{NI}^{CB}$, that is (cf. (i) of Proposition 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 3),

$$\delta \in \mathcal{B}_*(\delta) \tag{5.2}$$

Since, for all $\omega \in \Omega$, $\mathcal{B}_*(\omega) \subseteq \mathcal{K}_*(\omega)$, $\delta, \gamma \in \mathcal{K}_*(\beta)$. Since $\beta \in \mathcal{K}_*(\alpha)$ and $\alpha \in K_* \mathbf{NIK}^*$, $\beta \in \mathbf{NIK}^*$. Hence (cf. Remark 4),

$$\delta \in \mathcal{K}_*(\gamma) \tag{5.3}$$

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

$$\gamma \in \mathbf{AW}^{CB} \tag{5.4}$$

It follows from (5.2)-(5.4) and (ii) of Proposition 1 that $\delta \in \mathcal{B}_*(\gamma)$. □

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

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

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 every individual is correct in her knowledge of what is commonly known.

Lemma 4 $\mathbf{NIK}^* = \Omega$.

Proof. First note that (Ax.4′) 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 $\{B_i : 2^{\Omega} \to 2^{\Omega}\}_{i \in N}$ be any collection of operators satisfying Necessity, Monotonicity, Conjunction, Consistency, Positive and Negative Introspection (cf. (Ax.1) - (Ax. 3) and (Ax.5)-(Ax.7)), and let B_* be the corresponding common operator. We want to show that $\mathbf{VB}^{i*} = \Omega$ if and only if $\mathbf{NI}^{CB} = \Omega$. Let $\mathcal{B}_i : \Omega \to 2^{\Omega}$ be the possibility correspondence associated with B_i. For every $i \in N$ construct a new possibility correspondence $\mathcal{K}_i : \Omega \to 2^{\Omega}$ as follows: $\forall \omega, \omega' \in \Omega$, $\omega' \in \mathcal{K}_i(\omega)$ if and only if $\mathcal{B}_i(\omega') = \mathcal{B}_i(\omega)$. Then $\mathcal{K}_i$ gives rise to a partition of Ω, that is, $\forall \omega, \omega' \in \Omega$, $\omega \in \mathcal{K}_i(\omega)$ and if $\omega' \in \mathcal{K}_i(\omega)$ then $\mathcal{K}_i(\omega') = \mathcal{K}_i(\omega)$ (in the economics and game-theory literature this partition is called the *type partition* of individual i). Let $K_i : 2^{\Omega} \to 2^{\Omega}$ be the associated operator ($\forall \omega \in \Omega, \forall E \subseteq \Omega, \omega \in K_i E$ if and only if $\mathcal{K}_i(\omega) \subseteq E$). It is straightforward to verify that the system so constructed is a KB-system. Let K_* (resp. $\mathcal{K}_*$) be the associated common operator (resp. possibility correspondence). Then $\mathcal{K}_*$ also gives rise to a partition of Ω and therefore is euclidean, that is (cf. Proposition 1), $\mathbf{NIK}^* = \Omega$. Thus we can invoke Lemma 2 and conclude that

$$K_* \mathbf{NI}^{CB} = K_* \mathbf{VB}^{i*} \tag{5.5}$$

Since $\mathcal{K}_*$ is reflexive, K_* satisfies the Truth Axiom, that is, $\forall E \in 2^{\Omega}$, $K_* E \subseteq E$. Hence

$$\forall E \in 2^{\Omega}, \;\; K_* E = \Omega \;\; \textit{if and only if} \;\; E = \Omega \tag{5.6}$$

Suppose now that $\mathbf{VB}^{i*} = \Omega$. Then, by Necessity, $K_* \mathbf{VB}^{i*} = \Omega$. Thus, by (5.5), $K_* \mathbf{NI}^{CB} = \Omega$ and, by (5.6), $\mathbf{NI}^{CB} = \Omega$. By the same argument, if $\mathbf{NI}^{CB} = \Omega$ then $\mathbf{VB}^{i*} = \Omega$. □

Completion of proof of Theorem 5.1:
by Lemma 4, $\mathbf{NIK}^* = \Omega$; thus, by Monotonicity of K_*, $K_*\mathbf{NIK}^* = \Omega$. Hence Theorem 5.1 follows from Lemma 2. □

Remark 5 *By transitivity and secondary reflexivity of $\mathcal{B}_*$ and $\mathcal{K}_*$, for every event E, $B_*E = B_*B_*E$ and $K_*E = K_*K_*E$.*

Proof of Theorem 5.2.

The proof of Theorem 5.2 makes use of the following lemma.

Lemma 5 *In a weak KB-system (thus without assuming (A.4′)), the following holds: $\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap \mathbf{NIK}^* \subseteq B_*\mathbf{T}^{IB}$.*

Proof. Let $\alpha \in \neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap \mathbf{NIK}^*$. Since $\alpha \in \neg\mathbf{DIS}$, there exists a $\beta \in \mathcal{B}_*(\alpha)$ such that $\beta \in B_*\mathbf{T}^{IB}$. Suppose that $\alpha \notin B_*\mathbf{T}^{IB}$. Then there exists a $\gamma \in \mathcal{B}_*(\alpha)$ such that

$$\gamma \notin B_*\mathbf{T}^{IB}. \tag{5.7}$$

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

$$\gamma \in \mathcal{K}_*(\beta). \tag{5.8}$$

Since $\alpha \in K_*\mathbf{CAU}^{CB}$ and $\beta \in \mathcal{K}_*(\alpha)$ and, by Lemma 1, $K_*\mathbf{CAU}^{CB} \subseteq K_*\mathbf{AW}^{CB}$, $\beta \in \mathbf{AW}^{CB}$. Thus, since $\beta \in B_*\mathbf{T}^{IB}$, $\beta \in K_*B_*\mathbf{T}^{IB}$. Hence, by (5.8), $\gamma \in B_*\mathbf{T}^{IB}$, contradicting (5.7). □

Completion of proof of Theorem 5.2:
($\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \subseteq B_*\mathbf{T}^{IB} \cap K_*\mathbf{AW}^{CB}$) By Lemmas 4 and 5, $\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \subseteq B_*\mathbf{T}^{IB}$. By Lemma 1, $K_*\mathbf{CAU}^{CB} \subseteq K_*\mathbf{AW}^{CB}$.
($B_*\mathbf{T}^{IB} \cap K_*\mathbf{AW}^{CB} \subseteq \neg\mathbf{DIS} \cap \mathbf{K}_*\mathbf{CAU}^{CB}$.) By Remark 5, $B_*\mathbf{T}^{IB} = B_*B_*\mathbf{T}^{IB}$ and by seriality of $\mathcal{B}_*$, $B_*B_*\mathbf{T}^{IB} \subseteq \neg B_*\neg B_*\mathbf{T}^{IB} = \neg\mathbf{DIS}$. Thus $B_*\mathbf{T}^{IB} \subseteq \neg\mathbf{DIS}$. By Theorem 5.1, $K_*\mathbf{AW}^{CB} \subseteq \mathbf{K}_*\mathbf{CAU}^{CB}$. □

Remark 6 *Although $\neg\mathbf{DIS} \cap \mathbf{K}_*\mathbf{CAU}^{CB} \subseteq B_*\mathbf{T}^{IB} \cap \mathbf{AW}^{CB}$, the converse is not true as the following example shows. $N = \{1,2\}$, $\Omega = \{\alpha, \beta, \gamma\}$, $\mathcal{K}_1(\alpha) = \mathcal{K}_1(\gamma) = \{\alpha, \gamma\}$, $\mathcal{K}_1(\beta) = \mathcal{B}_1(\beta) = \{\beta\}$, $\mathcal{B}_1(\alpha) = \mathcal{B}_1(\gamma) = \{\gamma\}$, $\mathcal{K}_2(\alpha) = \mathcal{K}_2(\beta) = \{\alpha, \beta\}$, $\mathcal{K}_2(\gamma) = \mathcal{B}_2(\gamma) = \{\gamma\}$, $\mathcal{B}_2(\alpha) = \mathcal{B}_2(\beta) = \{\beta\}$. Thus, $\forall\omega \in \Omega$, $\mathcal{B}_*(\omega) = \{\beta, \gamma\}$ and $\mathcal{K}_*(\omega) = \Omega$. Here we have that $\mathbf{T}^{IB} = \{\beta, \gamma\}$, $B_*\mathbf{T}^{IB} = \Omega$ and $\mathbf{AW}^{CB} = \{\alpha\}$. Thus $B_*\mathbf{T}^{IB} \cap \mathbf{AW}^{CB} = \{\alpha\}$. On the other hand, $\mathbf{CAU}^{CB} = K_*\mathbf{CAU}^{CB} = \emptyset$.*

Proof of Theorem 5.3.

The proof of Theorem 5.3 is based on the following two lemmas, which are essentially one-agent results.

Lemma 6 *In a weak KB-system (thus without assuming (Ax.4′)),* $\mathbf{AW}^{CB} \cap K_*\mathbf{T}^{CB} \subseteq \mathbf{EQU}^{CB}$.

Proof. Let $\alpha \in \mathbf{AW}^{CB} \cap K_*\mathbf{T}^{CB}$. We want to show that $\mathcal{K}_*(\alpha) \subseteq \mathcal{B}_*(\alpha)$. Fix an arbitrary $\beta \in \mathcal{K}_*(\alpha)$. Then, since $\alpha \in K_*\mathbf{T}^{CB}$, $\beta \in \mathbf{T}^{CB}$. Thus (cf. Proposition 1) $\beta \in \mathcal{B}_*(\beta)$. Since $\alpha \in \mathbf{AW}^{CB}$ and $\beta \in \mathcal{K}_*(\alpha)$ and $\beta \in \mathcal{B}_*(\beta)$, by Proposition 1 it follows that $\beta \in \mathcal{B}_*(\alpha)$. □

Corollary 1 *In a weak KB-system (thus without assuming (A.4′)),* $K_*\mathbf{AW}^{CB} \cap K_*\mathbf{T}^{CB} \subseteq K_*\mathbf{EQU}^{CB}$.

Proof. By Lemma 6 and monotonicity of K_*, $K_*\mathbf{AW}^{CB} \cap K_*K_*\mathbf{T}^{CB} \subseteq K_*\mathbf{EQU}^{CB}$. By Remark 5, $K_*K_*\mathbf{T}^{CB} = K_*\mathbf{T}^{CB}$. □

Lemma 7 *In a weak KB-system (thus without assuming (A.4′)),* $\forall E \subseteq \Omega$, $\forall i \in N$, *(i)* $B_iK_iE \subseteq K_iE$ *and (ii)* $B_iK_*E = K_iK_*E$.

Proof. (i) From $\neg K_iE \subseteq K_i\neg K_iE$ (Ax.7) and $K_i\neg K_iE \subseteq B_i\neg K_iE$ ((Ax.8) applied to the event $\neg K_iE$) we get $\neg K_iE \subseteq B_i\neg K_iE$, which is equivalent $\neg B_i\neg K_iE \subseteq K_iE$. This, in conjunction with $B_iK_iE \subseteq \neg B_i\neg K_iE$ ((Ax.5) applied to the event K_iE), yields $B_iK_iE \subseteq K_iE$. (ii) Since (by definition of K_*) $K_*E \subseteq K_iK_*E$, by monotonicity of B_i, $B_iK_*E \subseteq B_iK_iK_*E$ and, by (i), $B_iK_iK_*E \subseteq K_iK_*E$. Thus $B_iK_*E \subseteq K_iK_*E$. On the other hand, by (Ax.8) $K_iK_*E \subseteq B_iK_*E$. □

Corollary 2 *In a weak KB-system (thus without assuming (A.4′)),* $K_*\mathbf{EQU}^{CB} \subseteq \mathbf{CAU}^{CB}$.

Proof. Let $\alpha \in K_*\mathbf{EQU}^{CB}$. Fix arbitrary $i \in N$ and $E \subseteq \Omega$ such that $\alpha \in B_iB_*E$. We want to show that $\alpha \in K_iB_*E$. First we show that a $\alpha \in B_iK_*E$. Fix an arbitrary $\gamma \in \mathcal{B}_i(\alpha)$. Then $\beta \in B_*E$. Since $\mathcal{B}_i(\alpha) \subseteq \mathcal{K}_i(\alpha) \subseteq \mathcal{K}_*(\alpha)$, $\beta \in \mathcal{K}_*(\alpha)$ and therefore (since $\alpha \in K_*\mathbf{EQU}^{CB}$) $\beta \in \mathbf{EQU}^{CB}$. Hence, since $\beta \in B_*E$, $\beta \in K_*E$. Thus a $\alpha \in B_iK_*E$. By (ii) of Lemma 7, $\alpha \in K_iK_*E$. Now choose an arbitrary $\gamma \in \mathcal{K}_i(\alpha)$. Then $\gamma \in K_*E$. Furthermore, since $\mathcal{K}_i(\alpha) \subseteq \mathcal{K}_*(\alpha)$, $\gamma \in \mathcal{K}_*(\alpha)$ and therefore $\gamma \in \mathbf{EQU}^{CB}$. Thus, since $\gamma \in K_*E$, $\gamma \in B_*E$. Hence $\alpha \in K_iB_*E$. □

Completion of proof of Theorem 5.3:
($K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB} \subseteq K_*\mathbf{EQU}^{CB}$) By Lemma 1, $K_*\mathbf{CAU}^{CB} \cap$

$K_*\mathbf{T}^{CB} \subseteq K_*\mathbf{AW}^{CB} \cap K_*\mathbf{T}^{CB}$ and by Corollary 1, $K_*\mathbf{AW}^{CB} \cap K_*\mathbf{T}^{CB} \subseteq K_*\mathbf{EQU}^{CB}$.
($K_*\mathbf{EQU}^{CB} \subseteq K_*\mathbf{CAU}^{CB}$) By Corollary 2 and Monotonicity of K_*, $K_*K_*\mathbf{EQU}^{CB} \subseteq K_*\mathbf{CAU}^{CB}$. By Remark 5, $K_*K_*\mathbf{EQU}^{CB} = K_*\mathbf{EQU}^{CB}$.
($K_*\mathbf{EQU}^{CB} \subseteq K_*\mathbf{T}^{CB}$). Let $\alpha \in K_*\mathbf{EQU}^{CB}$ and fix an arbitrary $\beta \in \mathcal{K}_*(\alpha)$. We want to show that $\beta \in \mathbf{T}^{CB}$. Since $\alpha \in K_*\mathbf{EQU}^{CB}$, $\beta \in \mathbf{EQU}^{CB}$. Thus, by Proposition 1, $\mathcal{B}_*(\beta) = \mathcal{K}_*(\beta)$. Since $\beta \in \mathcal{K}_*(\alpha)$, by secondary reflexivity of $\mathcal{K}_*$ (cf. Remark 3), $\beta \in \mathcal{K}_*(\beta)$. Thus $\beta \in \mathcal{B}_*(\beta)$. Hence, by Proposition 1, $\beta \in \mathbf{T}^{CB}$. □

Proof of Theorem 5.4.

First we prove that

$$K_*\mathbf{T}^{IB} \subseteq \neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB} \tag{5.9}$$

First note that $K_*\mathbf{T}^{IB} \subseteq B_*\mathbf{T}^{IB}$ and, as shown in the proof of Theorem 5.2, $B_*\mathbf{T}^{IB} \subseteq \neg\mathbf{DIS}$. Thus $K_*\mathbf{T}^{IB} \subseteq \neg\mathbf{DIS}$. Furthermore, since $\mathbf{T}^{IB} \subseteq \mathbf{T}^{CB}$, by Monotonicity of K_*, $K_*\mathbf{T}^{IB} \subseteq K_*\mathbf{T}^{CB}$. Finally, since $\mathbf{T}^{IB} \subseteq \mathbf{VB}^{i*}$, by Monotonicity of K_*, $K_*\mathbf{T}^{IB} \subseteq K_*\mathbf{VB}^{i*}$. By Lemmas 2 and 4, $K_*\mathbf{VB}^{i*} = K_*\mathbf{CAU}^{CB}$. Next we prove that

$$\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB} \subseteq K_*\mathbf{T}^{IB} \tag{5.10}$$

Let $\alpha \in \neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB}$. By Theorems 5.2 and 5.3, $\alpha \in B_*\mathbf{T}^{IB} \cap \mathbf{EQU}^{CB}$. Hence $\alpha \in K_*\mathbf{T}^{IB}$. Thus, by (5.9) and (5.10),

$$\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB} = K_*\mathbf{T}^{IB} \tag{5.11}$$

Next we prove that

$$K_*\mathbf{T}^{IB} \subseteq \mathbf{EQU}^{IB} \tag{5.12}$$

Let $\alpha \in K_*\mathbf{T}^{IB}$. Fix arbitrary $i \in N$ and $E \subseteq \Omega$ and suppose that $\alpha \in B_iE$. We need to show that $\alpha \in K_iE$. Fix an arbitrary $\beta \in \mathcal{K}_i(\alpha)$. We have to prove that $\beta \in E$. Since $\alpha \in B_iE \subseteq K_iB_iE$ and $\beta \in \mathcal{K}_i(\alpha)$,

$$\beta \in B_iE. \tag{5.13}$$

Since $\alpha \in K_*\mathbf{T}^{IB}$ and $\beta \in \mathcal{K}_i(\alpha) \subseteq \mathcal{K}_*(\alpha)$, $\beta \in \mathbf{T}^{IB}$. Hence, by (5.13), $\beta \in E$. By (5.12) and Monotonicity of K_*, $K_*K_*\mathbf{T}^{IB} \subseteq K_*\mathbf{EQU}^{IB}$. By Remark 5, $K_*K_*\mathbf{T}^{IB} = K_*\mathbf{T}^{IB}$. Thus $K_*\mathbf{T}^{IB} \subseteq K_*\mathbf{EQU}^{IB}$. It follows from this and (5.12) that

$$K_*\mathbf{T}^{IB} \subseteq \mathbf{EQU}^{IB} \cap K_*\mathbf{EQU}^{IB} \tag{5.14}$$

From (5.11) we get (by intersecting both sides with $\neg\mathbf{DIS}$) that $\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB} = \neg\mathbf{DIS} \cap K_*\mathbf{T}^{IB}$ and from (5.14) $\neg\mathbf{DIS} \cap K_*\mathbf{T}^{IB} \subseteq \neg\mathbf{DIS} \cap \mathbf{EQU}^{IB} \cap K_*\mathbf{EQU}^{IB}$. Thus $\neg\mathbf{DIS} \cap K_*\mathbf{CAU}^{CB} \cap K_*\mathbf{T}^{CB} \subseteq \neg\mathbf{DIS} \cap \mathbf{EQU}^{IB} \cap K_*\mathbf{EQU}^{IB}$.

We conclude the proof by showing that $\neg\mathbf{DIS} \cap \mathbf{EQU}^{IB} \cap K_*\mathbf{EQU}^{IB} \subseteq K_*\mathbf{T}^{IB}$. Let $\alpha \in \neg\mathbf{DIS} \cap \mathbf{EQU}^{IB} \cap K_*\mathbf{EQU}^{IB}$. By Lemma 4, $\mathbf{NIK}^* = \Omega$. Since $\alpha \in \mathbf{EQU}^{IB} \cap K_*\mathbf{EQU}^{IB}$, $\mathbf{NIK}^* = \mathbf{NI}^{CB}$. Thus $\alpha \in \mathbf{NI}^{CB} \cap \neg\mathbf{DIS}$. By definition of $\neg\mathbf{DIS}$, $\mathbf{NI}^{CB} \cap \neg\mathbf{DIS} \subseteq B_*\mathbf{T}^{IB}$. Thus $\alpha \in B_*\mathbf{T}^{IB}$. Since $\alpha \in \mathbf{EQU}^{IB} \cap K_*\mathbf{EQU}^{IB}$, $\alpha \in B_*\mathbf{T}^{IB}$ if and only if $\alpha \in K_*\mathbf{T}^{IB}$. □

The following proposition highlights an interesting property of $\mathbf{AW}^{CB}$ and $\mathbf{EQU}^{CB}$ (recall that throughout this appendix we have *not* assumed the Truth Axiom for knowledge). For a proof see Bonanno and Nehring (1998b).

Proposition 2 *In a weak KB-system (thus without assuming (Ax.4′)),* $K_*\mathbf{AW}^{CB} \subseteq \mathbf{AW}^{CB}$ *and* $K_*\mathbf{EQU}^{CB} \subseteq \mathbf{EQU}^{CB}$.

References

Aumann, R. 1976. Agreeing to disagree. *Annals of Statistics*, 4:1236–1239.

Battigalli, P. and G. Bonanno. 1997. The logic of belief persistence. *Economics and Philosophy*, 13:39–59.

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

Bonanno, G. 1996. On the logic of common belief. *Mathematical Logic Quarterly*, 42:305–311.

Bonanno, G. and K. Nehring. 1998a. Assessing the Truth Axiom under incomplete information. *Mathematical Social Sciences*, 36:3–29.

Bonanno, G. and K. Nehring. 1998b. Intersubjective consistency of knowledge and belief. Technical report, University of California, Davis. Department of Economics Working Paper #98–03.

Bonanno, G. and K. Nehring. 1998c. Making sense of the common prior assumption under incomplete information. *International Journal of Game Theory*. forthcoming.

Bonanno, G. and K. Nehring. 1998d. On the logic and role of negative introspection of common belief. *Mathematical Social Sciences*, 35:17–36.

Chellas, B. 1984. *Modal Logic: an Introduction.* Cambridge University Press.

Colombetti, M. 1993. Formal semantics for mutual beliefs. *Artificial intelligence*, 62:341–353.

Dekel, E. and F. Gul. 1997. Rationality and knowledge in game theory. In Kreps, D. M. and K. F. Wallis, editors, *Advances in Economic Theory, Seventh World Congress.* Cambridge University Press.

Fagin, R., J. Halpern, Y. Moses, and M. Vardi. 1995. *Reasoning About Knowledge.* MIT Press.

Geanakoplos, J. 1994. Common knowledge. In Aumann, R. and S. Hart, editors, *Handbook of Game Theory*, volume 2, pages 1437–1496. Elsevier.

Halpern, J. 1991. The relationshi between knowledge, belief and certainty. *Annals of Mathematics and Artificial Intelligence*, 4:301–322.

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.

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

van der Hoek, W. and J.-J. C. Meyer. 1995. *Epistemic Logic for Artificial Intelligence and Computer Science.* Cambridge University Press.

Kraus, S. and D. Lehmann. 1988. Knowledge, belief and time. *Theoretical Computer Science*, 58:155–174.

Lamarre, P. and Y. Shoham. 1994. Knowledge, certainty, belief and conditionalisation. In Doyle, J., E. Sandewall, and P. Torasso, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR'94).* Morgan Kaufmann.

Lenzen, W. 1978. Recent work in epistemic logic. *Acta Philosophica Fennica*, 30:1–220.

Lewis, D. 1969. *Convention: A Philosophical Study.* Harvard University Press.

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

Stalnaker, R. 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.