Agreeing to disagree: a survey

Giacomo Bonanno

and

Klaus Nehring

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

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

May 1997. Revised, September 1997

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

1. Introduction

Aumann (1976) put forward a formal definition of common knowledge and used it to prove that two “like minded” individuals cannot “agree to disagree” in the following sense. If they start from a common prior and update the probability of an event E (using Bayes’ rule) on the basis of private information then it cannot be common knowledge between them that individual 1 assigns probability p to E and individual 2 assigns probability q to E with $p \neq q$. In other words, if their posteriors of event E are common knowledge then they must coincide. This celebrated result captures the intuition that the fact that somebody else has a different opinion from yours is an important piece of information which should induce you to revise your own opinion. This process of revision will continue until consensus is reached.

Aumann’s original result has given rise to a large literature on the topic, which we review in this paper. We shall base our exposition on the distinction between Bayesian (or quantitative) versions and non-Bayesian (or qualitative) versions of the notion of agreeing to disagree.

2. Illustration of the logic of agreeing to disagree

Imagine two scientists who agree on everything. They agree that the true law of Nature must be one of seven, call them α, β, γ, δ, ε, ζ, η. They also agree on the relative likelihood of these possibilities, which they take to be as illustrated in Figure 1:

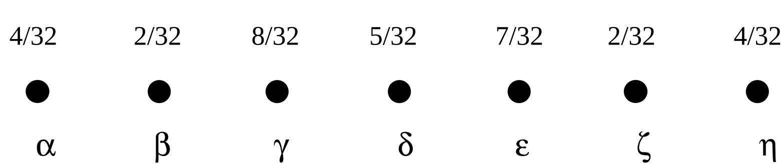

Figure 1

Experiments can be conducted to learn more. An experiment leads to a partition of the above set. For example, if the true law of Nature is α and you performed experiment 1 then you would learn that it cannot be δ or ε or ζ or η but you still would not know which is the true law of Nature among the remaining ones. Suppose that the scientists agree that Scientist 1 will perform experiment 1 and Scientist 2 will perform experiment 2. They also agree that each experiment would lead to a partition of the states as illustrated in Figure 2:

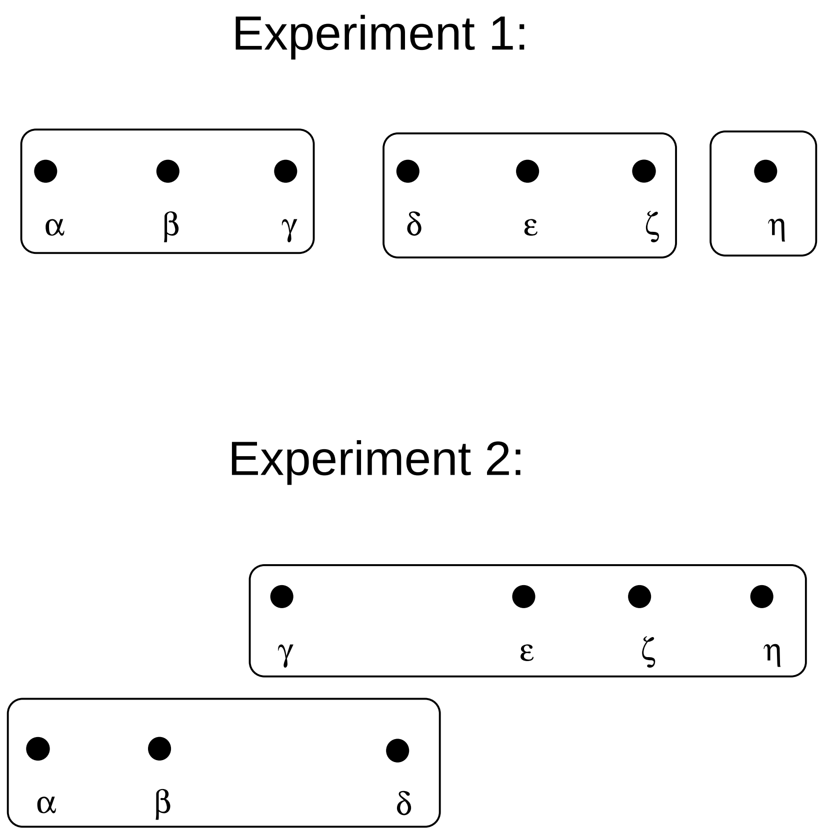

Figure 2

Suppose that they are interested in establishing the truth of a proposition that is represented by the event $E = \{\alpha, \gamma, \delta, \varepsilon\}$. Initially they agree that the probability that E is true is (cf. Figure 1):

$$P(E) = P(\alpha) + P(\gamma) + P(\delta) + P(\varepsilon) = \frac{24}{32} = 75\%$$

Before they perform the experiments they also realize that, depending on what the true law of Nature is, after the experiment they will have an updated probability of event E conditional on what the experiment has revealed. For example, they agree that if one performs Experiment 1 and the true state is β (so that E is actually false) then the experiment will yield the information $I = \{\alpha, \beta, \gamma\}$ and Bayesian updating (which they agree to be the correct way to update probabilities) will lead to the following new probability of event E:

$$P(E \mid I) = \frac{P(E \cap I)}{P(I)} = \frac{P(\{\alpha,\gamma\})}{P(\{\alpha,\beta,\gamma\})} = \frac{\frac{4}{32}+\frac{8}{32}}{\frac{4}{32}+\frac{2}{32}+\frac{8}{32}} = \frac{12}{14} = 86\%.$$ [1]

Similarly for every other possibility. Thus we can attach to every cell of each experiment a new updated probability of E, as illustrated in Figure 3.

[1] Note the interesting fact that sometimes experiments, although informative (they reduce uncertainty), might actually induce one to become more confident of the truth of something that is false: in this case one increases one's subjective probability that E is true from 75% to 86%, although E is actually false! (Recall that we have assumed that the true state is β.)

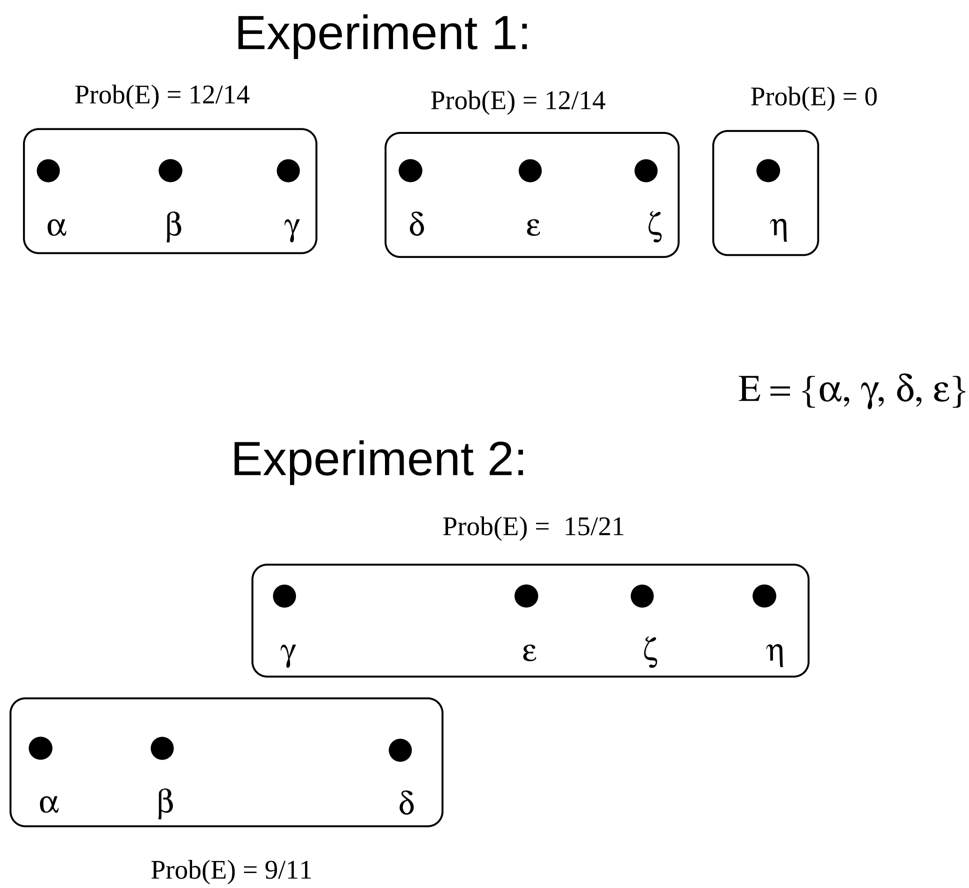

Figure 3

Suppose now that each scientist goes to her laboratory and performs the respective experiment (Scientist 1 Experiment 1 and Scientist 2 Experiment 2). **Assume** also **that the true state of Nature is** ζ. Afterwards they exchange e-mail messages informing each other of their new subjective estimates of event E. Scientist 1 says that she now attaches probability 12/14 to E, while Scientist 2 says that she attaches probability 15/21 to E. So their estimates disagree (not surprisingly, since they have performed different experiments and obtained different information). Should they be happy with these estimates? Obviously not. Consider Scientist 2. She learns that Scientist 1 has a new updated probability of 12/14. From this she can deduce that the true state is **not** η (had the true state been η she would have been communicated by Scientist 1 an updated probability of E of 0). She can thus revise her knowledge by eliminating η from her information set. Then she will need to re-compute the probability of E as shown in the following figure. Similarly, Scientist 1 learns that the true state cannot be δ, hence revises her information partition and estimate of E as illustrated in Figure 4.

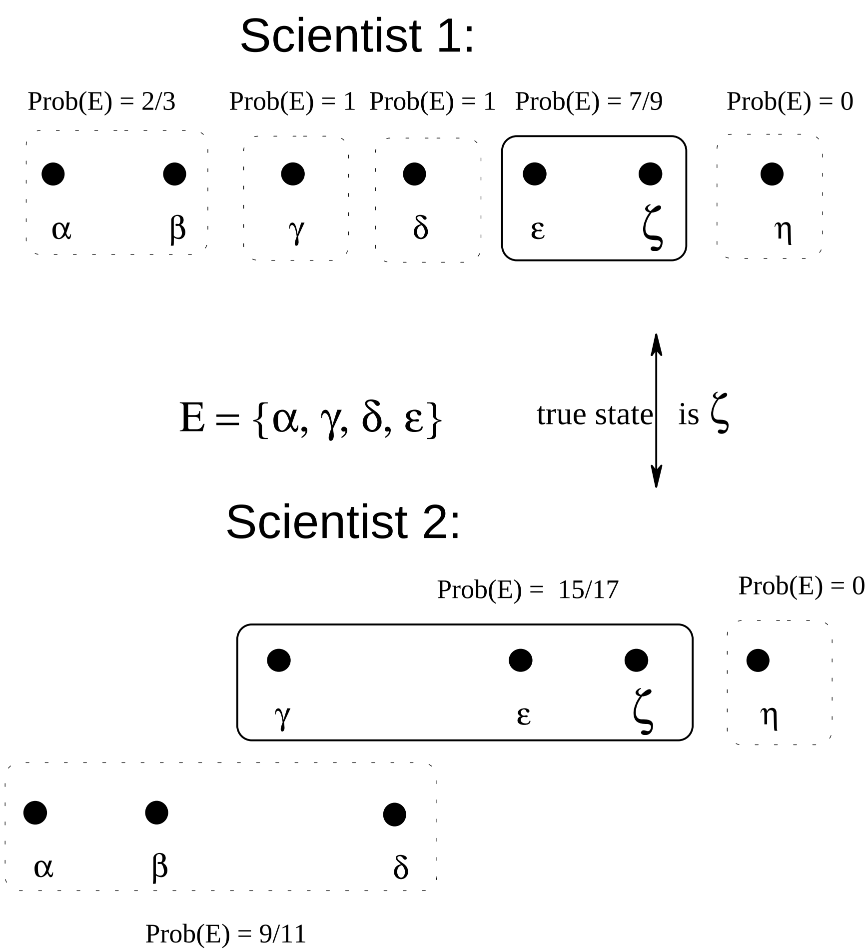

Figure 4

Now they inform each other of their new subjective estimates: 7/9 for Scientist 1 and 15/17 for Scientist 2. Again, there is disagreement. Should they accept such disagreement? The answer is, again, No. Scientist 1 does not learn anything from the new estimate of Scientist 2, but Scientist 2 does learn something, namely that the state cannot be γ. Hence she will revise her information partition and estimate of the probability of E, as illustrated in Figure 5:

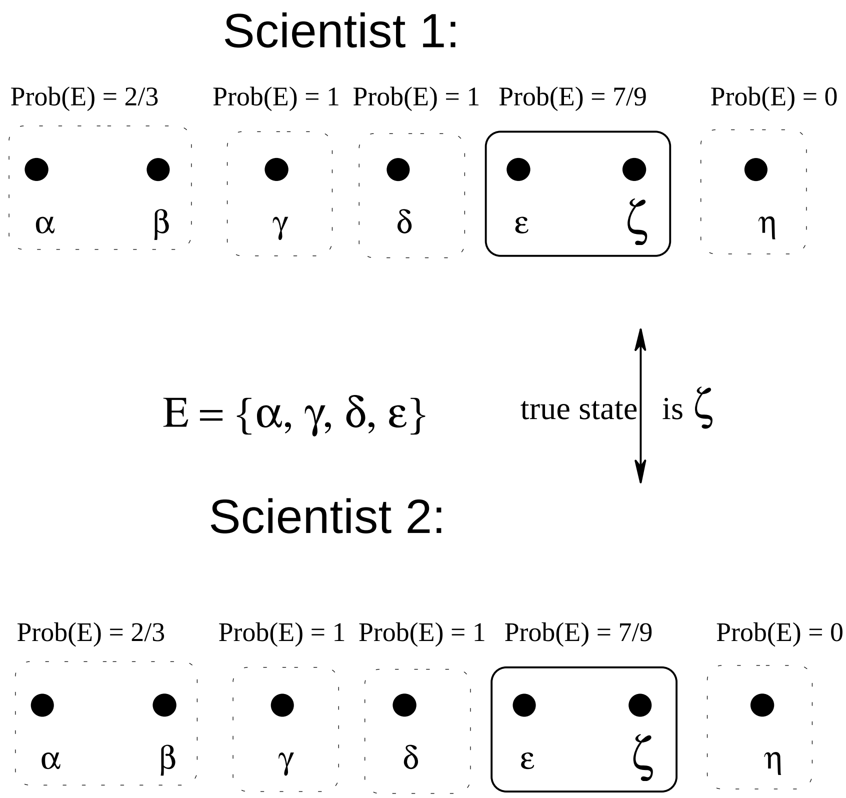

Figure 5

Notice that at this stage they finally agree on the probability of E and indeed it becomes common knowledge that both estimate this probability to be 7/9 = 78% (before the experiments the ex ante probability of E was 24/32 = 75%; note that with the experiments and the exchange of information they have gone further from the truth!).

Notice that before the last step (leading to it being common knowledge that P(E) = 7/9 for both scientists) it was not common knowledge between the two what probability each scientist attached to E. When one scientist announced his subjective estimate, the other scientist found that announcement informative and revised her own estimate accordingly. At the end of the process of exchanges, the announcement by one scientist of his estimate did not make the other scientist change her estimate. In a sense further announcements became pointless, occasioned no surprise, revealed nothing new.

Although this example suggests that the scientists will end up with exactly the same information, this is not true in general.

3. Formal statement of Aumann's result

In this section we provide a formal and precise statement of Aumann's (1976) result – known as the *Agreement Theorem* – which was proved within the context of knowledge and common knowledge. Extensions of the result to the more general case of belief and common belief will be examined in Sections 5-7.

Let Ω be a set of states. There are two individuals and suppose that they start with the same prior probability distribution $\mu : \Omega \to [0,1]$ on Ω $\left(\text{define, for every } E \subseteq \Omega,\ \mu(E) = \sum_{\omega \in E} \mu(\omega)\right)$. Individual i receives private information (the nature of the private information is common knowledge between the two) according to the information partition $\mathcal{I}_i$. For every state $\omega \in \Omega$, $I_i(\omega)$ denotes the cell of i's partition that contains ω. Assume that, for every $i \in \{1, 2\}$ and $\omega \in \Omega$, $\mu(I_i(\omega)) \neq 0$.

The knowledge operator of individual i, $K_i : 2^{\Omega} \to 2^{\Omega}$, is defined by $K_iE = \{\omega \in \Omega : I_i(\omega) \subseteq E\}$. The common knowledge operator $K_* : 2^{\Omega} \to 2^{\Omega}$ is defined by:

$$K_*E = K_1E \cap K_2E \cap K_1K_2E \cap K_2K_1E \cap K_1K_2K_1E \cap K_2K_1K_2E \cap \ldots$$

That is, an event E is common knowledge between the two if both know E, both know that both know E, etc. *ad infinitum*. Let $\mathcal{I}_*$ be the meet (finest common coarsening) of the partitions $\mathcal{I}_1$ and $\mathcal{I}_2$ and denote by $I_*(\omega)$ the cell of $\mathcal{I}_*$ that contains ω. Aumann (1976) proved that, for every event E, $K_*E = \{\omega \in \Omega : I_*(\omega) \subseteq E\}$.

Fix an event E and let $\|\mu_i(E) = a\| = \{\omega \in \Omega : \mu(E \mid I_i(\omega)) = a\}$.[2] That is, $\|\mu_i(E) = a\|$ is the event that individual i's posterior probability of E is a. In the example of the previous section, for Scientist 2 we have that – after the experiment and before any exchange of information with the other scientist – $\|\mu_2(E) = \frac{9}{11}\| = \{\alpha, \beta, \delta\}$, where $E = \{\alpha, \gamma, \delta, \varepsilon\}$.

Aumann's Agreement Theorem. Let Ω be a set of states and suppose that two individuals 1 and 2 have the same prior probability distribution $\mu : \Omega \to [0,1]$ on Ω sartisfying the property that for every $i \in \{1, 2\}$ and $\omega \in \Omega$, $\mu(I_i(\omega)) \neq 0$. Let E be an event and let $a,b \in [0,1]$. Suppose that at some state ω it is common knowledge that individual 1's posterior probability of E (given his information at ω) is a and 2's posterior probability of E (given her information at ω) is b. Then $a = b$. In other words, individuals who start with the same prior *cannot agree to disagree*. Formally,

$$\text{if} \quad K_*(\|\mu_1(E) = a\| \cap \|\mu_2(E) = b\|) \neq \varnothing \quad \text{then} \quad a = b.$$

The theorem is a consequence of the following lemma.

LEMMA 1. Fix an arbitrary event F and let $\{P_1, ..., P_m\}$ be a partition of F (thus $F = P_1 \cup ... \cup P_m$ and any two P_j and P_k with $j \neq k$ are non-empty and disjoint). Suppose that $\mu(E \mid P_j) = a$ for all j = 1, ..., m. Then $\mu(E \mid F) = a$.

Proof of Lemma 1. $\mu(E \mid P_j) = \dfrac{\mu(E \cap P_j)}{\mu(P_j)}$. Hence, since $\mu(E \mid P_j) = a$, we have that $\mu(E \cap P_j) = a\mu(P_j)$. Adding over j, the LHS becomes $\mu(E \cap F)$ and the RHS becomes $a\mu(F)$ (by definition of probability measure). Hence $\mu(E \mid F) \equiv \dfrac{\mu(E \cap F)}{\mu(F)} = a$.

[2] For every E, $I \subseteq \Omega$ such that $\mu(I) \neq 0$, $\mu(E|I)$ is the conditional probability of E given I, defined by $\mu(E|I) = \dfrac{\mu(E \cap I)}{\mu(I)}$.

Proof of the Agreement Theorem. Let $I_*(\omega)$ be the cell of the common knowledge partition containing ω. Consider individual 1. $I_*(\omega)$ is a union of cells of $\mathcal{I}_1$, the information partition of individual 1. On each such cell 1's conditional probability of E is a. By Lemma 1, $\mu(E \mid I_*(\omega)) = a$. A similar reasoning for individual 2 leads to $\mu(E \mid I_*(\omega)) = b$. Hence $a = b$. ■

4. Bayesian extensions of Aumann's result

Aumann's Agreement theorem has been extended in several directions. In this section we consider probabilistic or "Bayesian" extensions.

I. From events to expectations of random variables. Aumann's result can be extended from the probability of an event to the expectation of a random variable. In particular, the following extensions were proved by Milgrom and Stokey (1982), Sebenius and Geanakoplos (1983), Rubinstein and Wolinski (1990).

1. Let f be a random variable on Ω and a and b two distinct numbers. Then there is no ω at which it is common knowledge that, conditional on her information, individual 1 believes that the expectation of f is a and, conditional on his information, 2 believes that the expectation is b.[3]

2. Let f be a random variable on Ω and a a number. Then there is no ω at which it is common knowledge that, conditional on her information, individual 1 believes that the expectation of

[3] Note that Aumann's result is a special case of this: take f to be the characteristic function of event E.

f is greater than or equal to a and, conditional on his information, 2 believes that the expectation is less than a.[4]

The latter result can be interpreted as saying that it cannot be common knowledge between two risk-neutral individuals they both expect to profit from a bet: take $f(\omega)$ to be the payment from individual 1 to individual 2, if positive, and from 2 to 1, if negative, in case the true state turns out to be ω. If it is common knowledge that they both expect to gain from the bet, then it is common knowledge that for 1 the expectation of f is positive and for 2 the expectation of f is negative (the expectation of $-f$ is positive).

II. "No trade" theorems. Along these lines, Milgrom (1981) and Milgrom Stokey (1982) proved a result which is often interpreted as establishing the impossibility of speculative trade. Assume that two traders agree on an ex ante efficient allocation of goods. Then, after the traders get new information, there is no transaction with the property that it is common knowledge that both traders are willing to carry it out.

Morris (1994) explores further this "no trade" result, by looking at the case of *heterogeneous* prior beliefs. He shows how different notions of efficiency under asymmetric information (ex ante, ex interim and ex post) are related to agents' prior beliefs. These efficiency results are used to obtain necessary and sufficient conditions on agents' beliefs for no trade theorems in different environments.

III. From probabilities of events to aggregates. In many economic settings, instead of assuming that individual opinions (conditional probabilities or expectations) are

[4] Note, however, that it is possible that at a state ω it is common knowledge that for individual 1 the expectation of f is a and for individual 2 it is different from a. For example, let $\Omega = \{\alpha, \beta\}$, $I_1 = \{\Omega\}$ and $I_2 = \{\{\alpha\}, \{\beta\}\}$. Let $\mu(\alpha) = \mu(\beta) = \frac{1}{2}$ and $f(\alpha) = 1$, $f(\beta) = 3$. Then at α is it common knowledge that for 2 the expectation of f is 2 and for 1 it is different from 2 (it is either 1 or 3).

common knowledge, it is more natural to suppose that only some *aggregate* of individual opinions (e.g., a price) becomes common knowledge. In this context McKelvey and Page (1986) proved that if a stochastically monotone[5] aggregate of individual conditional probabilities is common knowledge, then it is still true that all the conditional probabilities must be equal. Nielsen *et al* (1990) show that also this result − like Aumann's result − can be extended from conditional probabilities of an event to conditional expectations of a random variable.

IV. Communication and common knowledge. One way in which common knowledge can be achieved is through communication. Suppose that the state space is finite. If the agents communicate to each other the probability of an event (or the expectation of a random variable, etc.) and revise their information partitions and subjective estimates accordingly, then a time will be reached after which communication induces no further revision. Then at that time it will become common knowledge at every state that each agent can predict the opinion she will hear from the other agent in the future. Then, by Aumann's theorem, at that time the opinions must be the same. This "convergence to common knowledge" theorem was proved by Geanakoplos and Polemarchakis (1982) and Sebenius and Geanakoplos (1983).

Along the same lines, Parikh and Krasucki (1990) consider communication protocols among more than two individuals in which values of functions are communicated *privately* through *messages*: when agent A communicates with agent B, other agents are not informed about the content of the message. They show that the agents may fail to reach agreement even with reasonable protocols. Krasucki (1996) takes this line of inquiry a step further by identifying

[5] A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is *stochastically monotone* if it can be written in the form $\phi(x_1, \ldots, x_n) = \phi_1(x_1) + \ldots + \phi_n(x_n)$ where each $\phi_i : \mathbb{R} \to \mathbb{R}$ is strictly increasing (this definition differs from, but is equivalent to, the one used by McKelvey and Page, 1986).

restrictions on protocols which guarantee that agreement is reached. Heifetz (1996) clarifies the relationship between consensus and common knowledge in this context.

V. Errors in information processing. Geanakoplos (1989) and Samet (1990) extend Aumann's (1976) result in a different direction. They ask what conditions on the individuals' information functions (or "possibility correspondences") are sufficient for the impossibility of agreeing to disagree. They assume the existence of a common prior and consider posterior beliefs obtained by updating the common prior on *non-partitional* possibility correspondences which represent how individuals process information. In particular, Samet generalizes Aumann's result from the case where the information function of each individual is reflexive and euclidean (and hence transitive) to the case where it is reflexive and transitive. In other words, he drops the Negative Introspection axiom for individual beliefs (see Section 7). Thus he takes a "bounded rationality" approach. Geanakoplos (1989) also focuses on "environments where information processing is subject to error" and finds even weaker conditions on individual beliefs that ensure the absence of speculation and of "agreement to disagree".[6] Geanakoplos, however, goes a step further by providing also a necessary condition, which he calls "positive balancedness".

[6] Like Samet, he assumes reflexivity of the information functions, but replaces transitivity with a property which he calls "(positive) balancedness". Thus the Truth Axiom (that is, reflexivity of the information function of each individual) plays a crucial role in Geanakoplos' and Samet's analysis.

5. Non-Bayesian or "Qualitative" generalizations of Aumann's result

Cave (1983) and Bacharach (1985) extended Aumann's result from the Bayesian setting (that is, from the conditional probability of an event given a common prior) to the case of like-minded individuals who follow a common decision procedure that satisfies the Sure Thing Principle. Roughly speaking, they showed that once two like-minded agents reach common knowledge of the actions each of them intends to perform, they will perform identical actions. This is illustrated in the following story, which Aumann (1989) attributes to Bacharach.

> A murder has been committed. To increase the chances of a conviction, the chief of police puts two detectives on the case, with strict instructions to work independently and exchange no information. The two, Alice and Bob, went to the same police school so given the same clues, they would reach the same conclusions. But as they will work independently, they will, presumably, not get the same clues. At the end of thirty days, each is to decide whom to arrest (possibly nobody). On the night before the thirtieth day, they happen to meet in the locker room at headquarters, and get to talking about the case. True to their instructions, they exchange no substantive information, no clues, but both are self-confident individuals, and feel that there is no harm in telling each other whom they plan to arrest. Thus when they leave the locker room, it is common knowledge between them whom Alice will arrest, and it is common knowledge between them whom Bob will arrest. Conclusion: They arrest the same people; and this, in spite of knowing nothing about each other's clues.

Let Ω be a set of states and denote by 2^{Ω} the set of events. Let A be a finite set of *actions.* A *decision procedure* is a function $D : 2^{\Omega}\backslash\varnothing \to A$. The interpretation is that the decision procedure D recommends action $D(I) \in A$ to an individual whose information (the set of states he considers possible) is I. What one learns in police school is a decision procedure: if you know such and such, then you should do so and so. The decision procedure $D : 2^{\Omega}\backslash\varnothing \to A$ satisfies the *Sure Thing Principle*, if and only if, for every event E, for every partition $\{P_1, ..., P_m\}$ of E and every $a \in A$,

$$D(P_i) = a \text{ for every } i = 1,...,m, \quad \text{implies } D(E) = a$$

Intuitively, suppose that if you knew which of the mutually exclusive events P_i happened, you would choose action *a* (which is the same for all P_i). Then you will take the same action *a* if you only know that some P_i happened, without knowing which one. Thus if Alice would arrest the butler if a certain blood stain is of type A, B, AB, or O, (perhaps for different reasons in each case), then she should arrest the butler without bothering to send the stain to the police laboratory.

Given decision procedure D and individuals i = 1,2 with information partitions $\mathcal{I}_i$ define *action functions* $d_i : \Omega \to A$ by $d_i(\omega) = D(I_i(\omega))$; in words, $d_i(\omega)$ is i's action at state ω. For every $a \in A$, we write $\| d_i = a \|$ for the event $\{\omega \in \Omega : d_i(\omega) = a\}$.

Generalized Agreement Theorem (Cave, 1983, Bacharach, 1985). Consider two individuals who follow the same decision procedure, which satisfies the Sure Thing Principle. If, at some state, it is common knowledge that individual 1 plans to take action *a* and individual 2 plans to take action *b*, then they must be planning to take the same action. Formally.

$$\text{if } \; K_*\left(\| d_1 = a \| \cap \| d_2 = b \|\right) \neq \emptyset \quad \text{then} \quad a = b.$$

REMARK. Aumann's Agreement Theorem is a corollary of the above: for a fixed event E, define the decision procedure D_E by $D_E(F) = \mu(E \mid F)$. By Lemma 1 of Section 3, this decision procedure satisfies the Sure Thing Principle. (Similarly, the expectation of a random variable satisfies the Sure Thing Principle.)

Moses and Nachum (1990) point out that Bacharach's technical definition of the Sure Thing Principle is considerably stronger than is suggested by the blood type example given above. Indeed, "it requires the decision procedure to be defined in a manner satisfying certain consistency properties at what amount to impossible situations" (Moses and Nachum, 1990, p. 152). They provide the following "counterexample" to the Generalized Agreement Theorem. A murder was committed and it is known that one of three suspects, A, B and C is the culprit. Two police officers are put on the case and are instructed to act independently and adhere to the following decision procedure:

1. If you know who the culprit is, indict him;

2. If you know that exactly one of them is *not* the culprit, of the other two arrest (for further interrogation) the one who comes first in alphabetical order;

3. If you cannot rule out any of the three as the culprit, do not arrest anybody.

This could be expressed formally as follows. For every $i \in \{A, B, C\}$ let i denote the state where individual i is the one who committed the murder. Then, for example, {A} represents the state of information of a detective who has established that A is the culprit, and {B, C} is the state of information of a detective who has established only that A is not the culprit. Then the above decision procedure can be expressed as follows:

1. D({A}) = indict A, D({B}) = indict B, D({C}) = indict C;

2. D({A, B}) = D({A, C}) = arrest A, D({B, C}) = arrest B;

3. D({A, B, C}) = do not arrest anybody.

This decision procedure satisfies the Sure Thing Principle trivially. Now imagine that detective Maigret has not collected any clues, while detective Columbo has established that (and only that) C is not the culprit. Columbo therefore initially intends to suggest that A be arrested, while Maigret would suggest that no arrest be made. Columbo communicates his intentions to Maigret, but Maigret cannot use this information to rule out any suspects and therefore insists on suggesting that no arrest be made. In the end it becomes common knowledge between them that Columbo intends to arrest A, while Maigret intends to suggest that no arrest be made. Hence they agree to disagree. The situation after their initial communication of intentions can be represented using a state space Ω that contains four points: α, β, γ and δ. At both α and γ suspect A is the murderer, at β the murderer is B, and at δ the murderer is C. Columbo's information partition is $\{\{\alpha, \beta\}, \{\alpha, \gamma\}\}$ while Maigret's information partition is $\{\Omega\}$.This is illustrated in Figure 6.

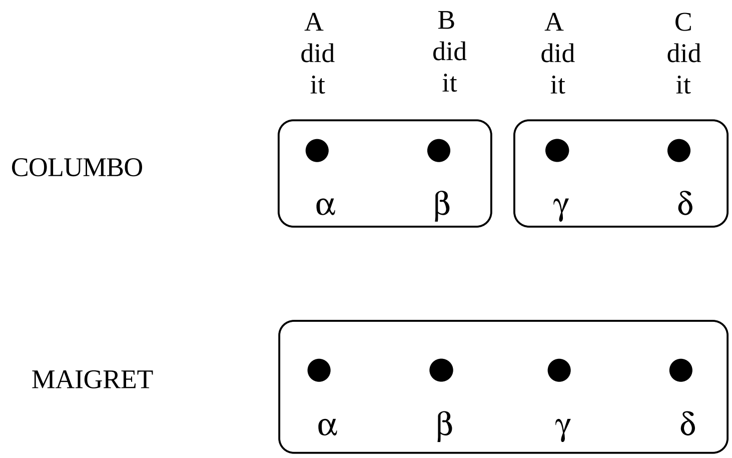

Figure 6

By the above decision procedure, $D(\{\alpha,\beta\}) = D(\{\gamma,\delta\})$ = "arrest A" and $D(\Omega)$ = "no arrest", thus violating Bacharach's Sure Thing Principle. Moses and Nachum point out that Bacharach's Sure Thing Principle in this case does not capture the intuition which is normally associated with it. Indeed, it makes no sense for Columbo to ask himself what he would do if he had Maigret's information, for the following reason. Maigret's information is that (1) either Columbo knows that C is not guilty or Columbo knows that B is not guilty, (2) he (Maigret) considers it possible that C is guilty and considers it possible that B is guilty. If Columbo were to know what Maigret knows (if he had the information Maigret has) then he would find himself believing contradictory propositions. Formally, let $E = \{\alpha, \beta\}$ and suppose that the true state is α. Then $\alpha \in K_M(K_C E \cup K_C \neg E)$ and $\alpha \in \neg K_M E \cap \neg K_M \neg E$. For Columbo to be in the same state of information at α as Maigret, it would have to be true that $\alpha \in K_C(K_C E \cup K_C \neg E)$ and $\alpha \in \neg K_C E \cap \neg K_C \neg E$. By Negative Introspection, $\neg K_C E \subseteq K_C \neg K_C E$ and $\neg K_C \neg E \subseteq K_C \neg K_C \neg E$. Thus $\alpha \in K_C \neg K_C E \cap K_C \neg K_C \neg E = K_C(\neg K_C E \cap \neg K_C \neg E) = K_C \neg(K_C E \cup K_C \neg E)$. Thus $\alpha \in K_C F \cap K_C \neg F$, where $F = K_C E \cup K_C \neg E$, contradicting consistency of knowledge. As Moses and Nachum (1990, p. 156) point out, "taking the union of states of knowledge in which an agent has differing knowledge does not result in a state of knowledge in which the agent is more ignorant; it simply does not result in a state of knowledge at all!".

Moses and Nachum's criticism of Bacharach's Sure Thing Principle is similar to Gul's (1996) and Dekel and Gul's (1997) criticism of the notion of a common prior in situations of incomplete information (see Section 6).

In their paper, Moses and Nachum go on to propose a weakening of the Sure Thing Principle and find conditions under which the weaker notion yields the impossibility of agreeing to disagree.

6. Common Prior and Agreement in situations of incomplete information.

The assumption of a common prior is central to Aumann's result on agreeing to disagree and related results, such as the no-trade theorem (Milgrom and Stokey, 1982). The Common Prior Assumption (CPA) plays an important role also in game theory: it is the basic assumption behind decision-theoretic justifications of equilibrium reasoning in games (Aumann, 1987, Aumann and Brandenburger, 1995).[7] Not surprisingly, the CPA has attracted its share of criticism. In models of *asymmetric information* (where there is an *ex ante* stage at which the individuals have identical information and subsequently update their beliefs in response to private signals), the controversy focuses on the plausibility or appropriateness of assuming *commonness* of the prior beliefs (see Morris, 1995). In this section we want to focus on situations of *incomplete information*, where there is no *ex ante* stage and where the primitives of the model are the individuals' beliefs about the external world (their first-order beliefs), their beliefs about the other individuals' beliefs (second-order beliefs), etc., i.e. their hierarchies of beliefs. In this context, the CPA is a mathematical property whose conceptual content is not clear. This has given rise to a novel and, in a way, more radical, criticism of the CPA, one that questions its very *meaningfulness* in situations of incomplete information (Dekel and Gul, 1997, Gul, 1996, Lipman, 1995).

---

[7] For an introduction to the epistemic foundations of solution concepts in game theory see Bonanno and Nehring (1997c).

The skepticism concerning the CPA in situations of incomplete information can be developed along the following lines. As Mertens and Zamir (1985) showed in their classic paper, the description of the “actual world” in terms of belief hierarchies generates a collection of “possible worlds”, one of which is the actual world. This set of possible worlds, or states, gives rise to a formal similarity between situations of asymmetric information and those of incomplete information. However, while a state in the former represents a real contingency, in the latter it is “a fictitious construct, used to clarify our understanding of the real world” (Lipman, 1995, p.2), “a notational device for representing the profile of infinite hierarchies of beliefs” (Gul, 1996, p. 3). As a result, notions such as that of a common prior, “seem to be based on giving the artificially constructed states more meaning than they have” (Dekel and Gul, 1997, p.42). Thus an essential step in providing a justification for, say, correlated equilibrium under incomplete information is to provide an interpretation of the “common prior” based on “assumptions that do not refer to the constructed state space, but rather are assumed to hold in the true state”, that is, assumptions “that only use the artificially constructed states the way they originated − namely as elements in a hierarchy of belief” (Dekel and Gul, 1997, p.116).

When the beliefs of the individuals can be viewed *as if* they were obtained by updating a common prior on some information, they are called *Harsanyi consistent.* Harsanyi consistency is a well-defined mathematical property, but, due to the “artificial nature” of the states in situations of incomplete information, “we do not know what it is that we would be accepting if we were to accept the common prior assumption” (Gul, 1996, p.5).

In this section we show that the existence of a common prior can be understood as a generalized form of Agreement, which we call Comprehensive Agreement. In order to motivate this notion, we take as point of departure the observation that, in some special cases, it is easy to find an interpretation of Harsanyi consistency that does not involve an *ex ante* stage. In particular, in situations of *complete* information (characterized by the fact that the beliefs of each individual are commonly known) Harsanyi consistency amounts to identity of beliefs across individuals. It thus seems natural, in situations of *incomplete* information, to think of Harsanyi consistency as likewise amounting to equality of those *aspects* of beliefs *that are commonly known*. For instance, one can take as an aspect of beliefs the subjective probability of an event E, in which case Agreement reduces to the notion introduced by Aumann (1976). Subjective probabilities of events are rather special aspects of beliefs and are not rich enough to fully capture the conceptual content of Harsanyi consistency. Thus one needs a more general notion of *Comprehensive Agreement* as the absence of *any* “agreement to disagree” about aspects of beliefs (or belief indices) in an appropriately defined general class.

In general situations of incomplete information where some individuals might have false beliefs (i.e. in non-partitional models), the relationship between Comprehensive Agreement and the existence of a common prior is somewhat complex. To see this, consider the following example, illustrated in Figure 7. Individual 2 is an economist who knows the correct spelling of his name (Mas-Colell). Individual 1 mistakenly believes that the spelling is Mas-Collel. She even believes this spelling to be common belief between them. These beliefs are represented by state τ in Figure 7. Note, in particular, that 1’s mistaken beliefs are represented with the help of an “artificial state” β which she believes to obtain for sure at τ (hence the arrow from τ to β)[8].

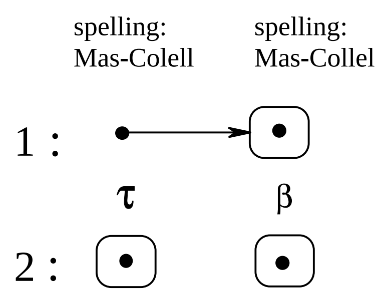

Figure 7

In this example Comprehensive Agreement is satisfied at the true state τ (and also at β). To see this, think of a *belief index,* or aspect of belief, as a function whose domain is the set of probability distributions over $\{\tau, \beta\}$. Let f be any such belief index. Then individual 1’s value of f is (the same, hence) common belief at every state, in particular at τ. Call this value x. Now, individual 2’s value of f at β must also be x (since they have the same beliefs there). Thus if 2’s index is common belief at τ it must be x. Hence at τ (and at β), there cannot be agreement to disagree (i.e. commonly known disagreement) about the aspect of beliefs captured by f.

Are the beliefs represented by state τ in Figure 7 Harsanyi consistent? This question requires clarification, since in an incomplete information setting properties need to be stated *locally* as pertaining to a particular profile of belief hierarchies − represented by the true state τ − rather than globally as pertaining to the model as a whole. In a weak sense the question can be answered affirmatively: the true state τ could be thought of as the *ex interim* stage of an asymmetric information model with a common prior that assigns probability 0 to τ and probability 1 to β. We define a corresponding weak notion of the CPA (*Harsanyi Quasi Consistency*), which in Figure 7 is satisfied at τ, and in Proposition 1 below we show it to be equivalent to Comprehensive Agreement. However, Harsanyi Quasi Consistency allows the "common prior" to give zero probability to the true state even if *every* individual assigns positive probability to it (see Figure 9). In such a case the true beliefs are compatible with the common prior largely due to the lack of restrictions associated with updating on zero probability events. As a result, the beliefs at the true state may be accounted for only incompletely by the common prior. A considerably stronger notion (*Strong Harsanyi Consistency*) requires the common prior to assign positive probability to the true state (in Figure 7, this requirement is not met). In Proposition 2 we provide the following characterization: Strong Harsanyi Consistency is equivalent to the conjunction of Comprehensive Agreement, "Truth *of* common belief" (what is actually commonly believed is true) and common belief in "Truth *about* common belief" (if somebody believes that E is commonly believed, then E is indeed commonly believed).

Requiring the "prior" to assign positive probability to the true state (that is, requiring Strong Harsanyi Consistency) is what is needed in order to translate to situations of incomplete information probability 1 results based on the Common Prior Assumption obtained in an asymmetric information context, such as Aumann's (1987) characterization of correlated equilibrium (see Bonanno and Nehring, 1997c).

---

[8] The state β is *defined* as the following conjunction of facts about the world and individuals' beliefs: "the correct spelling is MasCollel and it is commonly (and correctly) believed to be MasCollel". For more details on the

We now turn to the formal analysis.[9]

DEFINITION 1. An *interactive Bayesian frame* (or Bayesian frame, for short)[10] is a tuple

$$\mathcal{B} = \langle N, \Omega, \tau, \{p_i\}_{i \in N} \rangle$$

where

- $N = \{1, ..., n\}$ is a finite set of *individuals*.
- Ω is a finite set of *states* (or possible worlds)[11]. The subsets of Ω are called *events*.
- $\tau \in \Omega$ is the "true" or "actual" state[12].
- for every individual $i \in N$, $p_i : \Omega \to \Delta(\Omega)$ (where $\Delta(\Omega)$ denotes the set of probability distributions over Ω) is a function that specifies her *probabilistic beliefs*, satisfying the following property [we use the notation $p_{i,\alpha}$ rather than $p_i(\alpha)$]: $\forall \alpha, \beta \in \Omega$,

$$\text{if} \quad p_{i,\alpha}(\beta) > 0 \quad \text{then} \quad p_{i,\beta} = p_{i,\alpha} \tag{1}$$

Thus $p_{i,\alpha} \in \Delta(\Omega)$ is individual i's subjective probability distribution at state α and condition (1) says that every individual knows her own beliefs. We denote by $||p_i = p_{i,\alpha}||$ the event $\{\omega \in \Omega : p_{i,\omega} = p_{i,\alpha}\}$. It is clear that the set $\{ ||p_i = p_{i,\omega}|| : \omega \in \Omega \}$ is a partition of Ω; it is often referred to as individual i's *type partition*.

"state space" representation of belief hierarchies, see below.

[9] For a more detailed introduction to the semantics of belief and common belief see Bonanno and Nehring (1997b).

[10] For a similar definition see, for example, Aumann and Brandenburger (1995), Dekel and Gul (1997) and Stalnaker (1994, 1996).

[11] Finiteness of Ω is a common assumption in the literature (cf. Aumann, 1987, Aumann and Brandenburger, 1995, Dekel and Gul, 1997, Morris, 1994, Stalnaker, 1994, 1996).

[12] We have included the true state in the definition of an interactive Bayesian model in order to stress the interpretation of the model as a representation of a particular profile of hierarchies of beliefs.

Given a Bayesian model $\mathcal{B}$, its *qualitative interactive belief frame* (or frame, for short) is the tuple $\mathcal{Q} = \langle N, \ \Omega, \ \tau, \ \{P_i\}_{i\in N} \rangle$ where N, Ω, and τ are as in Definition 1 and

- for every individual $i\in N$, $P_i : \Omega \rightarrow 2^{\Omega}\backslash\varnothing$ is i's *possibility correspondence*, derived from i's probabilistic beliefs as follows: [13]

$$P_i(\alpha) = \text{supp}(p_{i,\alpha}).$$

Thus, for every $\alpha\in\Omega$, $P_i(\alpha)$ is the set of states that individual i considers possible at α.

REMARK 1. It follows from condition (1) of Definition 1 that the possibility correspondence of every individual i satisfies the following properties (whose interpretation is given in Footnote 16): $\forall\, \alpha,\beta\in\Omega$,

Seriality (or *non-empty-valuedness*):	$P_i(\alpha) \neq \varnothing$,
Transitivity:	if $\beta\in P_i(\alpha)$ then $P_i(\beta)\subseteq P_i(\alpha)$,
Euclideanness:	if $\beta\in P_i(\alpha)$ then $P_i(\alpha)\subseteq P_i(\beta)$.

REMARK 2 (Graphical representation). A non-empty-valued and transitive possibility correspondence $P : \Omega \rightarrow 2^{\Omega}$ can be uniquely represented (see Figures 6-10) as an asymmetric directed graph[14] whose vertex set consists of disjoint events (called *cells* and represented as rounded rectangles) and states (represented as points), and each arrow goes from, or points to, either a cell or a state that does not belong to a cell. In such a directed graph, $\omega' \in P(\omega)$ if and only if either ω and ω' belong to the same cell or there is an arrow from ω, or the cell containing ω, to ω', or the cell containing ω'. Conversely, given a transitive directed graph in the above class such that each state either belongs to a cell or has an arrow out of it,

[13] If $\mu\in\Delta(\Omega)$, supp(μ) denotes the support of μ, that is, the set of states that are assigned positive probability by μ.

[14] A directed graph is *asymmetric* if, whenever there is an arrow from vertex v to vertex v′ then there is no arrow from v′ to v.

there exists a unique non-empty-valued, transitive possibility correspondence which is represented by the directed graph.

The possibility correspondence is euclidean if and only if all arrows connect states to cells and no state is connected by an arrow to more than one cell (for an example of a non-euclidean possibility correspondence see the common possibility correspondence P_* of Figure 8 below).

Finally, if − in addition − the possibility correspondence is reflexive, then one obtains a partition model where each state is contained in a cell and there are no arrows between cells (as is the case, for example, in Figure 6).

Given a frame and an individual i, i's *belief operator* $B_i : 2^{\Omega} \to 2^{\Omega}$ is defined as follows: $\forall E \subseteq \Omega$, $B_i E = \{\omega \in \Omega : P_i(\omega) \subseteq E\}$. $B_i E$ can be interpreted as the event that (i.e. the set of states at which) individual i *believes for sure* that event E has occurred (i.e. attaches probability 1 to E).[15]

Notice that we have allowed for false beliefs by not assuming reflexivity of the possibility correspondences ($\forall \alpha \in \Omega$, $\alpha \in P_i(\alpha)$), which −as is well known (Chellas, 1984, p. 164) − is equivalent to the *Truth Axiom* (if the individual believes E then E is indeed true): $\forall E \subseteq \Omega$, $B_i E \subseteq E$[16].

[15] Thus Condition (1) of Definition 1 can be stated as follows: $\forall i \in N$, $\forall \alpha \in \Omega$, $\|p_i = p_{i,\alpha}\| = B_i \|p_i = p_{i,\alpha}\|$.

[16] It is well known (see Chellas, 1984, p. 164) that non-empty-valuedness of the possibility correspondence is equivalent to *consistency* of beliefs (an individual cannot simultaneously believe E and not E): $\forall E \subseteq \Omega$, $B_i E \subseteq \neg B_i \neg E$ (where, for every event F, $\neg F$ denotes the complement of F). Transitivity of the possibility correspondence is equivalent to *positive introspection* of beliefs (if the individual believes E then she believes that she believes E): $\forall E \subseteq \Omega$, $B_i E \subseteq B_i B_i E$. Finally, euclideanness of the possibility correspondence is equivalent to *negative introspection* of beliefs (if the individual does not believe E, then she believes that she does not believe E): $\forall E \subseteq \Omega$, $\neg B_i E \subseteq B_i \neg B_i E$.

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

$$B_* E = B_e E \cap B_e B_e E \cap B_e B_e B_e E \cap \ldots$$

The corresponding possibility correspondence P_* is then defined as follows: for every $\alpha \in \Omega$, $P_*(\alpha) = \{\omega \in \Omega : \alpha \in \neg B_* \neg\{\omega\}\}$. It is well known that P_* can be characterized as the *transitive closure* of $\bigcup_{i \in N} P_i$, that is,

> $\forall \alpha, \beta \in \Omega$, $\beta \in P_*(\alpha)$ if and only if there is a sequence $\langle i_1, \ldots i_m \rangle$ in N (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 $k = 0, \ldots, m-1$, $\eta_{k+1} \in P_{i_{k+1}}(\eta_k)$.

Note that, although P_* is always non-empty-valued and transitive, in general it need not be euclidean (despite the fact that the individual possibility correspondences are: for an example see Figure 8; recall that −cf. Footnote 16 − P_* is euclidean if and only if B_* satisfies Negative Introspection).

To give contents to the beliefs of the individuals, one needs to add to a frame a specification of the "facts of Nature" that are true at every state. By doing so one obtains a *model* based on the given frame. A state in a model determines, for each individual, her beliefs about the external world (her first-order beliefs), her beliefs about the other individuals' beliefs about the external world (her second-order beliefs), her beliefs about their beliefs about her beliefs (her third-order beliefs), and so on, *ad infinitum*. An entire hierarchy of beliefs about beliefs about beliefs ... about the relevant facts is thus encoded in each state of a model. For example, consider the following model, which is illustrated in Figure 8 according to the convention established in Remark 2: $N = \{1, 2\}$, $\Omega = \{\tau, \beta, \gamma, \delta\}$, $p_{1,\tau} = p_{1,\gamma} = p_{1,\delta} = \begin{pmatrix} \beta & \tau & \gamma & \delta \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$, $p_{1,\beta} = p_{2,\beta} = p_{2,\tau} = \begin{pmatrix} \beta & \tau & \gamma & \delta \\ 1 & 0 & 0 & 0 \end{pmatrix}$, $p_{2,\gamma} = p_{2,\delta} = \begin{pmatrix} \beta & \tau & \gamma & \delta \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. Here the event $\{\beta, \tau, \gamma\}$ represents the proposition "it is sunny" and event $\{\delta\}$ the proposition "it is cloudy". The true state τ describes a world where in fact it is sunny, individual 2 believes that it is sunny and believes that 1 also believes it is sunny (indeed he believes that this is common belief), but in fact 1 believes that it is sunny with probability $\frac{1}{3}$ and cloudy with probability $\frac{2}{3}$ and believes that also 2 is uncertain as to whether it is sunny or cloudy (and attaches equal probability to both), etc.

Conversely, given any profile of infinite hierarchies of beliefs (one for each individual) satisfying minimal coherency requirements, one can construct an interactive Bayesian model $\mathcal{B}$ such that *at the true state* τ the beliefs of each individual $i \in N$ fully capture i's original infinite hierarchy of beliefs (see Boege and Eisele, 1979, Brandenburger and Dekel, 1993, Mertens and Zamir, 1985, and Battigalli, 1997)[17].

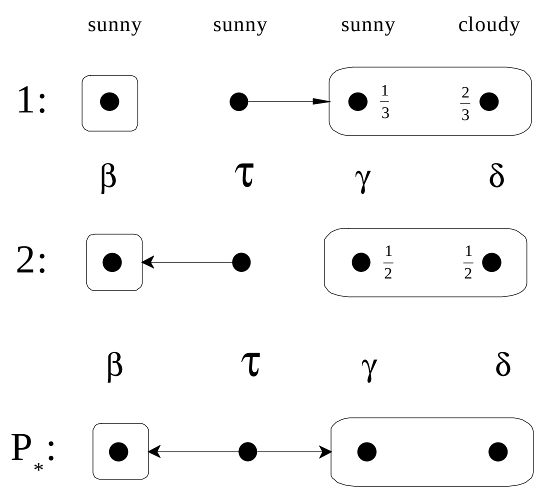

Figure 8

It is not obvious what a proper *local* formulation of the existence of a "common prior" ought to be. Below we suggest two definitions. The first turns out to be equivalent to a

[17] Finiteness of Ω, however, cannot be guaranteed in general.

generalized notion of Agreement, but is too weak in some respects. The second, stronger, definition is more appealing but is no longer equivalent to Agreement.

DEFINITION 2. For every $\mu \in \Delta(\Omega)$, let $\mathbf{HQC}_\mu$ (for Harsanyi Quasi Consistency with respect to the "prior" μ) be the following event: $\alpha \in \mathbf{HQC}_\mu$ if and only if

(1) $\forall i \in \mathrm{N}$, $\forall\, \omega, \omega' \in \mathrm{P}_*(\alpha)$, if $\mu(\|\mathrm{p_i} = \mathrm{p}_{\mathrm{i},\omega}\|) > 0$ then $\mathrm{p}_{\mathrm{i},\omega}(\omega') = \dfrac{\mu(\omega')}{\mu(\|\mathrm{p_i} = \mathrm{p}_{\mathrm{i},\omega}\|)}$ if $\omega' \in \|\mathrm{p_i} = \mathrm{p}_{\mathrm{i},\omega}\|$ and $\mathrm{p}_{\mathrm{i},\omega}(\omega') = 0$ otherwise (that is, $\mathrm{p}_{\mathrm{i},\omega}$ is obtained from μ by conditioning on $\|\mathrm{p_i} = \mathrm{p}_{\mathrm{i},\omega}\|$)[18], and

(2) $\mu(\mathrm{P}_*(\alpha)) > 0$.

If $\alpha \in \mathbf{HQC}_\mu$, μ is a *local common prior at* α. Furthermore, let $\mathbf{HQC} = \bigcup_{\mu \in \Delta(\Omega)} \mathbf{HQC}_\mu$.

We now define formally a general notion of agreement. Agreement as equality of beliefs is essentially a two-person property. Hence, for the remaining part of this section, we specialize to the case where N = {1,2}.

Let X be a set with at least two elements. A *belief index* is a function $f : \Delta(\Omega) \to X$ [19].

EXAMPLE 1. (i) Let $\mathrm{E} \subseteq \Omega$ be an arbitrary event, $X = [0,1]$ and f^{E} the following belief index: $f^{\mathrm{E}}(\mathrm{p}) = \mathrm{p(E)}$; thus $f^{\mathrm{E}}(\mathrm{p}_{\mathrm{i},\alpha})$ is individual i's subjective probability of event E at state α.

[18] Note that, for every $\omega \in \Omega$ and $i \in N$, $\omega \in \|\mathrm{p_i} = \mathrm{p}_{\mathrm{i},\omega}\|$. Thus $\mu(\omega) > 0$ implies $\mu(\|\mathrm{p_i} = \mathrm{p}_{\mathrm{i},\omega}\|) > 0$.

[19] It may seem that a belief index f depends on the set of states Ω. However, this is not so: one should think of f as being defined on the "universal belief space" (cf. Mertens and Zamir, 1985). Indeed, all that matters is the restriction of f to $\mathrm{P}_*(\tau)$.

(ii) Let $\mathrm{Y} : \Omega \to \mathbb{R}$ be a random variable, $X = \mathbb{R}$ and f_Y be the belief index given by $f_Y(p) = \sum_{\omega\in\Omega} Y(\omega)p(\omega)$.; thus $f_Y(p_{i,\alpha})$ is i's subjective expectation of Y at state α.

(iii) Let A be a set of actions, $X = 2^{\mathrm{A}}$ and $\mathrm{U} : \mathrm{A} \times \Omega \to \mathbb{R}$ a utility function. Define the belief index $f_{\mathrm{U}} : \Delta(\Omega) \to 2^{\mathrm{A}}$ as follows: $f_{\mathrm{U}}(\mathrm{p}) = \arg\max_{a\in\mathrm{A}} \sum_{\omega\in\Omega} \mathrm{U}(a,\omega)\,\mathrm{p}(\omega)$. Thus $f_{\mathrm{U}}(\mathrm{p}_{\mathrm{i},\alpha})$ is the set of actions that maximize individual i's expected utility at state α.

DEFINITION 3. A belief index is *proper* if and only if it satisfies the following property: $\forall\, \mathrm{p},\mathrm{q} \in \Delta(\Omega),\ \forall x\in X,\ \forall a\in[0,1]$,

$$\text{if } f(\mathrm{p}) = f(\mathrm{q}) = x \text{ then } f(a\mathrm{p} + (1-a)\mathrm{q}) = x.$$

Let $\mathcal{J}_\infty$ denote the class of proper belief indices.

It is easily verified that all the belief indices of Example 1 are proper.

Given a proper belief index $f : \Delta(\Omega) \to X$ and an individual $\mathrm{i}\in \mathrm{N}$, define $f_{\mathrm{i}} : \Omega \to X$ by $f_{\mathrm{i}}(\alpha) = f(\mathrm{p}_{\mathrm{i},\alpha})$. For every $x\in X$ denote the event $\{\alpha\in\Omega : f_{\mathrm{i}}(\alpha) = x\}$ by $||f_{\mathrm{i}} = x||$.

DEFINITION 4. Given a Bayesian model and a proper belief index $f: \Delta(\Omega) \to X$, at $\alpha \in \Omega$ there is *Agreement for f* or *f-Agreement* if and only if, for all $x_1, x_2\in X$,

$$\alpha \in \mathrm{B}_*\left(||f_1 = x_1|| \cap ||f_2 = x_2||\right) \Rightarrow x_1 = x_2. \qquad (2).$$

That is, if at α it is common belief that individual 1's belief index is x_1 and individual 2's index is x_2, then $x_1 = x_2$. Correspondingly, define the following event:[20]

$$\boldsymbol{f}\textbf{-Agree} \;=\; \bigcap_{\substack{x_1, x_2 \in \mathrm{X} \\ x_1 \neq x_2}} \neg \mathrm{B}_* \left(\|f_1 = x_1\| \cap \|f_2 = x_2\| \right). \qquad (3).$$

Given a Bayesian model and a set $\mathcal{F}$ of proper belief indices, at α there is *Agreement on* $\mathcal{F}$ or $\mathcal{F}$*-Agreement* if and only if , $\forall f \in \mathcal{F}$, $\alpha \in$ $\boldsymbol{f}$**-Agree**. Correspondingly, let

$$\mathcal{F}\textbf{-Agree} \;=\; \bigcap_{f \in \mathcal{F}} \boldsymbol{f}\textbf{-Agree} \qquad (4).$$

A general notion of agreement is given by the entire class of proper belief indices.

DEFINITION 5. Let **CA** (for Comprehensive Agreement) be the following event:

$$\mathbf{CA} \;=\; \mathcal{F}_{\infty}\textbf{-Agree.}$$

The following proposition (proved in Bonanno and Nehring, 1996) characterizes Comprehensive Agreement as equivalent to Harsanyi Quasi Consistency.

PROPOSITION 1. **CA** = **HQC.**

The notion of Harsanyi Quasi Consistency is rather weak: it allows the “common prior” to assign zero probability to the true beliefs of *all* the individuals (even if none of the individuals has false beliefs: Example 2) and it is compatible with some individuals believing that there is agreement to disagree (Example 3).

---

[20] Throughout the paper bold-face letters and expressions are used to denote events (subsets of Ω).

EXAMPLE 2. Consider the frame of Figure 9. Let $\mu \in \Delta(\Omega)$ be such that $\mu(\beta) = 1$. Then $\mathbf{HQC}_\mu = \Omega$. Thus at τ Harsanyi Quasi Consistency is satisfied even though the type (beliefs) of each individual is assigned zero probability by μ. Note that at τ *both individuals have correct beliefs* ($\tau \in P_1(\tau) \cap P_2(\tau)$).

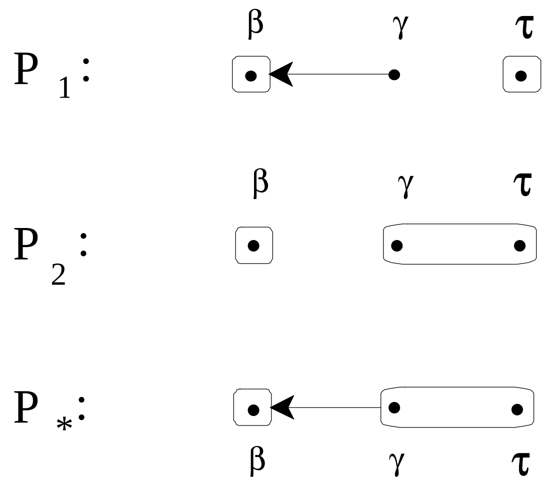

Figure 9

EXAMPLE 3. Consider the model of Figure 8. Let $\mu \in \Delta(\Omega)$ be such that $\mu(\beta) = 1$. Then $\mathbf{HQC}_\mu = \{\tau, \beta\}$. At τ Harsanyi Quasi Consistency is satisfied even though individual 1 believes that he and 2 "agree to disagree" about the probability that it is sunny (that is, 1 believes that it is common belief that he himself attaches probability 1/3 to the event $\{\beta, \tau, \gamma\}$ while 2 attaches probability 1/2 to it).

In view of the above examples, Harsanyi Quasi Consistency is too weak a notion to allow the translation to situations of incomplete information of results that are based on the Common Prior Assumption, such as Aumann's (1987) characterization of correlated equilibrium. In order to strengthen the notion of Harsanyi Quasi Consistency one needs to tighten the connection between the implied prior and the true beliefs/state. The following definition does so by requiring the prior to assign positive probability to the true state.

DEFINITION 6. For every $\mu \in \Delta(\Omega)$, let $\mathbf{SHC}_\mu$ (for Strong Harsanyi Consistency with respect to the "prior" μ) be the following event: $\alpha \in \mathbf{SHC}_\mu$ if and only if

(1) $\alpha \in \mathbf{HQC}_\mu$, and

(2) $\mu(\alpha) > 0$.

Furthermore, let $\mathbf{SHC} = \bigcup_{\mu \in \Delta(\Omega)} \mathbf{SHC}_\mu$.

To explore the gap between **HQC** and **SHC** we introduce the following events ($\mathbf{T}_{\mathrm{CB}}$ stands for Truth *about* common belief, $\mathbf{T}^*$ stands for Truth *of* common belief, while $\mathbf{NI}^*$ stands for Negative Introspection of common belief):

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

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

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

$\mathbf{T}_{\mathrm{CB}}$ captures the notion that individuals are correct in their beliefs about what is commonly believed: $\alpha \in \mathbf{T}_{\mathrm{CB}}$ if and only if, for every event E and individual i, if, at α, individual i believes that E is commonly believed, then, at α, E is indeed commonly believed (if $\alpha \in B_i B_* E$ then $\alpha \in B_* E$). On the other hand, $\alpha \in \mathbf{T}^*$ if and only if at α whatever is commonly believed is true (for every event E, if $\alpha \in B_* E$ then $\alpha \in E$)[21]. Finally, $\alpha \in \mathbf{NI}^*$ if and only if – for every event E –

[21] It is straightforward that $\alpha \in \mathbf{T}^*$ if and only if, $\alpha \in I_*(\alpha)$. Clearly, Truth of common belief is qualitatively weaker than Truth; since $B_* \mathbf{T}^* = \Omega$, $\mathbf{T}^*$ can be viewed as Truth shorn of any intersubjective implications.

whenever at α it is not common belief that E, then, at α, it is common belief that E is not commonly believed.[22]

The following proposition follows from results proved in Bonanno and Nehring (1996, 1997a)

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

7. Qualitative agreement

In this section we show that the qualitative counterpart to Harsanyi Quasi Consistency is the property of *Qualitative Agreement* denoted by **A**. First, let **T** (for Truth) be the following event:

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

Thus, for every $\alpha \in \Omega$, $\alpha \in \mathbf{T}$ if and only if no individual has any false beliefs at α (for every $i \in N$ and for every $E \subseteq \Omega$, if $\alpha \in B_i E$ then $\alpha \in E$)[23]. Let **A** (for Agreement) be the following event:

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

A captures the notion of Agreement on qualitative belief indices, as we now show. Among the proper belief indices defined in the previous section, of particular interest are the following special cases: *simple* indices, which take on only two values, 0 and 1, and *qualitative* indices,

[22] It is well known that $\alpha \in \mathbf{NI}^{*}$ if and only if $P_{*}(\alpha)$ satisfies the following property: $\forall \beta, \gamma \in P_{*}(\alpha)$, $\gamma \in P_{*}(\beta)$.

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

which depend only on the support of $p \in \Delta(\Omega)$. We denote the first class by $\mathcal{F}_2$ and the latter by $\mathcal{F}_Q$. Thus

$$\mathcal{F}_2 = \{\, f : \Delta(\Omega) \to X \ : \ \text{(i) } f \in \mathcal{F}_\infty, \ \text{(ii) } X = \{0, 1\} \text{ and (iii) } f^{-1}(1) \text{ is closed} \,\} \tag{5}$$

$$\mathcal{F}_Q = \{\ f \in \mathcal{F}_\infty \ : \forall\, p, q \in \Delta(\Omega), \text{ if } \mathrm{supp}(p) = \mathrm{supp}(q) \text{ then } f(p) = f(q) \,\}. \tag{6}$$

The following results are proved in Bonanno and Nehring (1996).

PROPOSITION 3. $f \in \mathcal{F}_2$ if and only if there exists a random variable $Y : \Omega \to \mathbb{R}$ such that, for all $p \in \Delta(\Omega)$, $f(p) = \begin{cases} 1 & \text{if } \sum_{\omega \in \Omega} Y(\omega)\, p(\omega) \geq 0 \\ 0 & \text{otherwise} \end{cases}$.

REMARK 3. A qualitative belief index can be written as $f = d_f \circ \mathrm{supp}$, with $d_f : 2^{\Omega} \setminus \varnothing \to X$ (such functions d_f have been studied in Rubinstein and Wolinsky, 1990). A qualitative belief index is proper if and only if d_f is *union consistent*, that is,

$$\forall\, m \geq 1,\ \forall E_1, \ldots, E_m \in 2^{\Omega},\ \forall\, x \in X, \quad \text{if } d_f(E_k) = x \text{ for all } k = 1, \ldots, m \text{ then } d_f\Big(\bigcup_{k=1}^{m} E_k\Big) = x. \tag{7}$$

Note that since the events $E_1, \ldots, E_m$ are *not* assumed to be pairwise disjoint, union consistency is a stronger property than the Sure Thing Principle defined by Bacharach (1985).

Fix an event $E \neq \varnothing$ and consider the following index: $f_E(p) = \begin{cases} 1 & \text{if } \mathrm{supp}(p) \subseteq E \\ 0 & \text{otherwise} \end{cases}$.
Thus, for every individual i and state α, $f_E(p_{i,\alpha}) = 1$ if and only if $\alpha \in B_i E$[24]. Let

[24] To represent f_E in the manner of Proposition 3, let $Y : \Omega \to \mathbb{R}$ be as follows: $Y = 1_E - 1$, where $1_E : \Omega \to \{0,1\}$ is the characteristic function of E: $1_E(\omega) = 1$ if and only if $\omega \in E$. Hence

$\mathcal{F}_S = \{ f_E : \Delta(\Omega) \to \{0,1\} : E \subseteq \Omega \}$ (the subscript "S" stands for "simple"). Clearly, $\mathcal{F}_S \subseteq \mathcal{F}_2 \cap \mathcal{F}_Q$. The following proposition shows that in fact $\mathcal{F}_S$ coincides with $\mathcal{F}_Q \cap \mathcal{F}_2$.

PROPOSITION 4. $\mathcal{F}_S = \mathcal{F}_Q \cap \mathcal{F}_2$.

Note that $\alpha \in \mathcal{F}_S$**-Agree** if and only if, for no event E, $\alpha \in B_*(B_1E \cap \neg B_2E)$, that is, there is no event about which the two individuals "agree to disagree":

$$\mathcal{F}_S\textbf{-Agree} = \bigcap_{i \in N} \bigcap_{j \in N} \bigcap_{E \in 2^\Omega} \neg B_*(B_iE \cap \neg B_jE).$$

LEMMA 1. $\forall \alpha \in \Omega$, $\alpha \in \mathcal{F}_S$**-Agree** if and only if

$$\forall i,j \in N,\ \exists \beta \in P_*(\alpha) \text{ such that } P_j(\beta) \subseteq \bigcup_{\omega \in P_*(\alpha)} P_i(\omega).$$

As a corollary to Lemma 1 we get that Qualitative Agreement rules out agreeing to disagree about events.

COROLLARY 1. $\mathbf{A} \subseteq \mathcal{F}_S$**-Agree**.

The converse to Corollary 1 does not hold. To see this, consider the frame illustrated in Figure 10. By Lemma 1, $\mathcal{F}_S$**-Agree** $= \Omega$; on the other hand, $\mathbf{A} = \varnothing$ (in fact, $\mathbf{T} = \{\tau, \beta\}$ and, therefore, $B_*\mathbf{T} = \varnothing$; thus $\neg B_*\neg B_*\mathbf{T} = \varnothing$).

$Y(\omega) = \begin{cases} 0 & \text{if } \omega \in E \\ -1 & \text{if } \omega \notin E \end{cases}$. Then $\sum_{\omega \in \Omega} Y(\omega)p(\omega) = \sum_{\omega \in \neg E} Y(\omega)p(\omega) = -\sum_{\omega \in \neg E} p(\omega) < 0$ if and only if $p(\omega) > 0$ for some $\omega \in \neg E$, if and only if $\sum_{\omega \in E} p(\omega) < 1$.

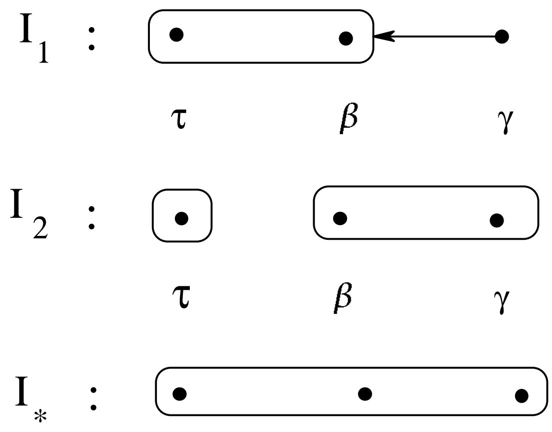

Figure 10

To obtain a full characterization of Qualitative Agreement one needs to consider the *entire* class of qualitative belief indices.

PROPOSITION 5. **A** = $\mathcal{I}_Q$-**Agree**.

It follows from Proposition 5 and the above example that $\mathcal{I}_S$-**Agree** $\neq$ $\mathcal{I}_Q$-**Agree**. Thus, in contrast to the case of general "quantitative" proper belief indices, for which simplicity can be assumed without loss of generality (i.e. $\mathcal{I}_\infty$-**Agree** = $\mathcal{I}_2$-**Agree**: see Bonanno and Nehring, 1996), simplicity *is* a restrictive assumption for qualitative belief indices.

References

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

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

Aumann, R. (1989), Notes on interactive epistemology, mimeo, Hebrew University of Jerusalem.

Aumann, R. and A. Brandenburger (1995), Epistemic conditions for Nash equilibrium, *Econometrica*, 63, 1161-1180.

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

Battigalli, P. (1997), Hierarchies of conditional beliefs and interactive epistemology in dynamic games, mimeo, Princeton University.

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

Boege, W. and T. Eisele (1979), On solutions of Bayesian games, *International Journal of Game Theory*, 8, 193-215.

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

Bonanno, G. and K. Nehring (1996), How to make sense of the Common Prior Assumption under incomplete information, Working paper, University of California Davis.

Bonanno, G. and K. Nehring (1997a), Assessing the Truth Axiom under incomplete information, Working paper, University of California Davis; forthcoming in *Mathematical Social Sciences*.

Bonanno, G. and K. Nehring (1997b), Introduction to the semantics of belief and common belief, Working paper, University of California Davis.

Bonanno, G. and K. Nehring (1997c), Introduction to the epistemic foundations of solution concepts in game theory, Working paper, University of California Davis.

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

Brandenburger, A. and E. Dekel (1993), Hierarchies of beliefs and common knowledge, *Journal of Economic Theory*, 59, 189-198.

Cave, J. (1983), Learning to agree, *Economics Letters*, 12, 147-152.

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

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 (eds.), *Advances in Economics and Econometrics*, vol. 1, Cambridge University Press.

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

Feinberg, Y. (1995), A converse to the Agreement Theorem, Discussion Paper # 83, Center for Rationality and Interactive Decision Theory, Jerusalem.

Feinberg, Y. (1996), Characterizing the existence of a common prior via the notion of disagreement, Ph. D. thesis, Hebrew University of Jerusalem.

Geanakoplos, J. (1989), Game theory without partitions and applications to speculation and consensus, Cowles Foundation Discussion Paper 914, Yale University.

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

Geanakoplos, J. (1994), Common knowledge, in R. J. Aumann and S. Hart (Eds), *Handbook of Game Theory*, Vol. 2, Elsevier.

Geanakoplos, J. and M. Polemarchakis (1982), We can't disagree forever, *Journal of Economic Theory*, 26, 363-390.

Gul, F. (1996), A comment on Aumann's Bayesian view, mimeo, Northwestern University [forthcoming in *Econometrica*].

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

Harsanyi, J. (1967-68), Games with incomplete information played by "Bayesian players", Parts I-III, *Management Science*, 8, 159-182, 320-334, 486-502.

Heifetz, A. (1996), Comment on consensus without common knowledge, *Journal of Economic Theory*, 70, 273-277.

Heifetz, A. (1996), Common belief in monotonic epistemic logic, *Mathematical Social Sciences*, 32, 109-123.

Kaneko, M. and T. Nagashima (1991), Final decisions, the Nash equilibrium and solvability in games with common knowledge of logical abilities, *Mathematical Social Sciences*, 22, 229-255.

Kaneko, M. and T. Nagashima (1997), Indefinability of the common knowledge concept in finitary logics, in M. Bacharach, L.A. Gérard-Varet, P. Mongin and H. Shin (Eds.), *Epistemic logic and the theory of games and decisions*, Kluwer Academic.

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

Lipman, B. (1995), Approximately common priors, mimeo, University of Western Ontario.

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

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

McKelvey, R. and T. Page (1986), Common knowledge, consensus and aggregate information, *Econometrica*, 54, 109-127.

Mertens, J-F. and S. Zamir (1985), Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory*, 14, 1-29.

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

Milgrom, P. and N. Stokey (1982), Information, trade and common knowledge, *Journal of Economic Theory*, 26, 17-27.

Morris, S. (1994), Trade with heterogeneous prior beliefs and asymmetric information, *Econometrica*, 62, 1327-1347.

Morris, S. (1995), The common prior assumption in economic theory, *Economics and Philosophy*, 11, 227-53.

Moses, Y. and G. Nachum (1990), Agreeing to disagree after all, in R. Parikh (Ed.), *Theoretical aspects of reasoning about knowledge: proceedings of the third conference*, Morgan Kaufmann, San Francisco.

Nielsen, L. (1984), Common knowledge, communication and convergence of beliefs, *Mathematical Social Sciences*, 8, 1-14.

Nielsen, L., A. Brandenburger, J. Geanakoplos, R. McKelvey and T. Page (1990), Common knowledge of an aggregate of expectations, *Econometrica*, 58, 1235-1239.

P. Krasucki (1996), Protocols forcing consensus, *Journal of Economic Theory*, 70, 266-272.

Parikh, R. and P. Krasucki (1990), Communication, consensus and knowledge, *Journal of Economic Theory*, 51, 184-193.

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

Samet, D. (1990), Ignoring ignorance and agreeing to disagree, *Journal of Economic Theory*, 52, 190-207.

Sebenius, J. and J. Geanakoplos (1983), Don’t bet on it: contingent agreements with asymmetric information, *Journal of the American Statistical Association*, 78:382, 424-26.

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

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