Looking Backwards, Looking Inwards: Priors and Introspection

by

Dov Samet

August 1996

ABSTRACT

The three notions studied here are Bayesian priors, invariant priors and introspection. A prior for an agent is Bayesian, if it agrees with the agent's posterior beliefs when conditioned on them. A prior is invariant, if it is the average, with respect to itself, of the posterior beliefs. Finally, an agent is introspective, if he is certain of his own beliefs. We show that a prior is Bayesian, if and only if it is invariant, and the agent is almost surely introspective. We show how to edogenize priors, and how to express the events that an agent has a Bayesian or invariant prior. Finally, we study properties of the endogenized common prior.

§1. Introduction

We study the relations between three properties of probabilistic beliefs. The first one is the property of being *Bayesian*, which means that agents form their beliefs by "looking back" at their previous, or prior, beliefs, and revise them, by conditioning on the information they have acquired. Another one is *introspection*, which means that agents can "look inward" and tell for sure what their beliefs are. We show that Bayesian beliefs are necessarily introspective. To wit, only those who can look inward are capable of looking back and properly using their past beliefs.

Another property of prior beliefs is *invariance*. A prior belief is invariant if it is the average, with respect to the prior itself, of the posterior beliefs. We prove that posterior beliefs are Bayesian with respect to a certain prior, if and only if the agent is almost surely introspective, with respect to the prior, and the prior in invariant.

Belief spaces, also known as type spaces, are the most suitable, and the most frequently used models for studying relations between beliefs. In section 2 we present the standard model of belief space. In this model, each agent has a *type function* over the space that associates [1] with each point of the space (called a *state*), his *type* at the state, which is a probability measure on the space. The model we present is a measure theoretic one, in the spirit of Heifetz and Samet (1996).

Other than imposing a straightforward measurability condition on the type functions, we make no assumption on these functions, that is, on the way types are associated with states. The reason is that we want to investigate the relations between properties of type functions and various assumption on priors.

In section 3 we define Bayesian priors. Here we give an informal explanation of our definition. The idea that posterior beliefs are derived, by conditioning, from prior beliefs is roughly expressed by,

posterior beliefs = prior beliefs, *given* the information the agent received.

But the information the agent has received is just what led him to revise his prior beliefs, by conditioning, to the posterior ones. Thus we can write,

posterior beliefs = prior beliefs, *given* the information that made the agent change his prior to his posterior beliefs.

But the case that the agent has information that makes him change his prior beliefs to the given posterior beliefs is precisely the case that he holds these posterior beliefs. Thus we can re-formulate the relationship between prior and posterior beliefs as,

posterior beliefs = prior beliefs, *given* the posterior beliefs.

To illustrate this principle, consider the following question we pose to John.

[1] Usually, one also needs to associate with each state a *state of nature*. For our purposes this is not required.

> "Given that *tomorrow* you will believe that the probability of Clinton being elected for a second term to be at least .6, and the probability of Saddam Hussein continuing to reign until the year 2000 to be at least .8, what do you think *today* about the probabilities of these events?"

If John is a Bayesian, we can expect no answer other than

> "These probabilities are at least .6 and .8, respectively".

The definition we give to Bayesian priors reflects and formalizes this dialogue. Note that this definition does not *presume* introspection, nor does it take for granted that there is any non-trivial event about which the agent is certain. Nevertheless, the existence of a Bayesian prior *implies* that the agent is certain about his own beliefs.

In section 3 we also introduce the notion of an invariant prior, which is closely related to the notion of a Bayesian prior. A prior is invariant for agent i if it satisfies,

> prior beliefs = expectation, with respect to the prior, of i's posterior beliefs.

Common invariant priors have been studied under various names by Mertens-Zamir (1985), Feinberg (1996) and Bonnano (1995) and Samet (1996).

We show that Bayesian priors are always invariant, but demonstrate, by a simple example, that the invariance of a prior does not imply that it is a Bayesian one. An important feature of this example is that the agent is not *introspective*, which leads us to define and study introspection in section 4. An agent is introspective if he is certain of his own beliefs. In terms of type functions, the agent is introspective in a given state if his type, at this state, ascribes probability one to the set of all states in which he has the same type. We show that the set of all states in which an agent is introspective is a measurable set, by showing that it can be expressed in terms of events of the form "If agent i ascribes to E a probability of at least p, then he is certain of this assessment", which are measurable by definition.

The meaning and justification of introspection, or lack of it, for probabilistic beliefs, is a subject of debate and dispute, (see a survey in Skyrms (1984)). Our contribute to this issue is the following equivalence theorem, which we prove in section 5:

> A prior for an agent is Bayesian, if and only if it is invariant, and the agent is almost surely introspective with respect to the prior.

Thus, in particular, the modeling of an agent as Bayesian requires that he be introspective. Hence, models of games with incomplete information, in which players have Bayesian priors (let alone a common one) must be partitional.

In section 5 we show how to handle priors and common priors *endogenously*, in much the same way as posterior beliefs are modeled. This enables us to represent, within the model itself, statements about beliefs concerning priors, and give, as we do in section 6, conditions under which a common prior is a common certainty among the agents.

Making priors endogenous to the model does not require adding any special features to belief spaces. Priors are introduced into the model by designating one of the agents *observer*, and his beliefs, or types, as the prior for another agent's beliefs or for several of them.

The nature of this observer is open to different interpretations. It can be the same agent in a previous period of time, or a real outside observer, or just the embodiment of the mathematical object — the prior. In any case, identifying the prior with the beliefs of an agent enables us to incorporate the prior in a standard uni-temporal model of multi-agent differential information. We assign the role of priors to the beliefs of one agent only, the observer, for the sake of simplicity. It is possible to have several observers, one for each agent, or even to have observers for the observers, thus modeling multi period beliefs.

Endogenizing Bayesian and invariant priors does not require the introduction of a binary operator for conditional probability into our model. We describe the event that an agent is Bayesian, or that he has an invariant prior, only in terms of events of the form "Agent i ascribes to E a probability of at least p." Events of this kind, which are formed using belief operators (see Monderer and Samet (1989)) fully describe the agents' beliefs in type space. Indeed, as is shown in Heifetz and Samet (1996), such events can be used to construct universal type spaces rather than the hierarchical construction of Mertens and Zamir (1985). Alternatively, one can use a formal language to describe such events, and using a deductive system define consistent sets of descriptions, as in Fagin and Halpern (1988) and Aumann (1989). Although we do not develop here a syntactic model of prior and posterior beliefs, our results provide the required axioms for the deductive system of such a model, in a language, whose only modal operators are the belief operators.

We discuss common priors in section 6. Common priors are related to common certainty (the probabilistic version of common knowledge) in a simple way.

> If the observer is certain of some event, and his type is a common invariant prior to all other agents, then he is certain that this event is common certainty among the other agents.

Using this observation we explore another relationship between introspection and priors.

> If the observer is introspective and his type is a common invariant prior to to the other agents, then he is certain that his type, namely the common prior, is common certainty.

Finally, we formulate and prove Aumann's (1976) agreement theorem, in our setup, which requires certain subtleties that are not required in the set theoretic model, with introspection in all states, for which it was originally proved.

§2. Belief spaces

Fix a set $I = \{0, 1, \ldots, n\}$ of *agents*.

Definition 1. *A type space is a triplet* $(\Omega,\ \Sigma,\ (t_i)_{i\in I})$ *where,*

(i) Ω *is a measurable space with a* σ*-field* Σ*, generated by a countable field* Σ_0*,*

(ii) for each $i \in I$*,* t_i *is a map from* Ω *to* $\Delta(\Omega)$ *— the set of all* σ*-additive probability measures on* Ω *— such that for each* $E \in \Sigma$*,* $t_i(\cdot)(E)$ *is a measurable real function.*

The elements of Ω are called *states*, and we refer to the measurable sets in Σ as *events* in Ω. The complement of an event E is denoted by $\neg E$. The maps t_i are called the *type maps*, and for each state $\omega \in \Omega$, the probability measure $t_i(\omega)$ is *i's type at* ω.

The event

$$B_i^p(E) = \{\omega \mid t_i(\omega)(E) \geq p\},$$

is the event that i *ascribes to* E *a probability of at least* p. By the measurability condition on t_i, this is indeed an event. The functions $B_i^p(\cdot)$, from Σ into itself, are called *belief operators*. These operators are monotonic, that is, if $E \subseteq F$ then for each p, $B_i^p(E) \subseteq B_i^p(F)$. The following continuity properties of belief operators can be easily established. If p_k is an increasing sequence converging to p, then $B_i^{p_k}(E) \downarrow B_i^p(E)$ for each E. If $E_k \downarrow E$, then $B_i^p(E_k) \downarrow B_i^p(E)$ for each p.

Events of the form $B_i^1(E)$ play an important role and deserve a special name. We call $B_i^1(E)$ the event that i is *certain* of E.

§3. Bayesian priors and invariant priors

In this section we fix an agent i, refer to him as "the agent", and omit the subscript i from his belief operators and type. We define first a Bayesian prior for the agent to be a probability measure on the state space, which, conditioned on his beliefs, agrees with these beliefs. Formally:

Definition 2. *The probability measure* μ *on* Ω *is a Bayesian prior for the agent, if, for each* $m \geq 1$*, all events* $E_1, \ldots, E_m$*, and all numbers* $p_1, \ldots, p_m$ *in* $[0, 1]$*,*

$$\mu\Big(E_l \mid \bigcap_{k=1}^{m} B^{p_k}(E_k)\Big) \geq p_l, \qquad (1)$$

for $1 \leq l \leq m$*, whenever this conditional probability is defined.*

An *invariant prior* for the agent is a probability measure on the state space such that the probability of each event is the expected probability the agent ascribes to it, where expectation is taken with respect to the prior itself [2]. The property of being an invariant prior for all players is called by Harsanyi (1975) consistency. Formally:

---

[2] The type function t_i can be considered as the transition function of a Markov chain on Ω. An invariant prior is an invariant probability measure of this Markov chain. Samet (1996) uses this observation to derive a necessary and sufficient condition for the existence of a common prior.

Definition 3. *The probability measure* μ *on* Ω *is an invariant prior for the agent, if, for each event* E,

$$\mu(E) = \int t(\omega)(E)\ d\mu(\omega). \tag{2}$$

Theorem 1. *If the probability measure* μ *in* $\Delta(\Omega)$ *is a Bayesian prior for the agent, then it is an invariant prior for him.*

Proof: To simplify the notation, we introduce an abbreviation for the event that the agent's assessment of the probability of E is in the interval $[p, q)$, as follows:

$$B^{p,q}(E) = B^p(E) \cap \neg B^q(E).$$

Let μ be a Bayesian prior for the agent. We show that for any given $\varepsilon > 0$ the two sides of (2) cannot differ by more than ε.

Choose a sequence of numbers $0 = p_0 < p_1 < \cdots < p_m = 1$, such that $|p_k - p_{k-1}| \leq \varepsilon$ for $k = 1, \ldots, m$. Clearly, the events $B^{p_{k-1}, p_k}(E)$ for $k = 1, \ldots m$ form a partition of Ω and therefore,

$$\mu(E) = \sum_{k=1}^{m} \mu\big(E \cap B^{p_{k-1}, p_k}(E)\big), \tag{3}$$

and

$$\int t(\omega)(E)\ d\mu(\omega) = \sum_{k=1}^{m} \int_{B^{p_{k-1}, p_k}(E)} t(\omega)(E)\ d\mu(\omega). \tag{4}$$

To evaluate the k term in (3), observe that when r converges to p_k from below, the events $F^r = B^{p_{k-1}}(E) \cap B^{1-r}(\neg E)$ converge monotonically to $B^{p_{k-1}, p_k}(E)$. By (1), $\mu(E \cap F^r) \geq p_{k-1}\, \mu(F^r)$ and $\mu\left(\neg E \cap F^r\right) \geq (1-r)\, \mu(F^r)$. By the continuity of μ we have,

$$p_{k-1}\ \mu\big(B^{p_{k-1}, p_k}(E)\big) \leq \mu\big(E \cap B^{p_{k-1}, p_k}(E)\big) \leq p_k\ \mu\big(B^{p_{k-1}, p_k}(E)\big). \tag{5}$$

To evaluate the k term in (4) we note that for $\omega \in B^{p_{k-1}, p_k}(E)$, $t(\omega)(E) \in [p_{k-1}, p_k)$, and thus

$$p_{k-1}\ \mu\big(B^{p_{k-1}, p_k}(E)\big) \leq \int_{B^{p_{k-1}, p_k}(E)} t(\omega)(E)\ d\mu(\omega)\ \leq p_k\ \mu\big(B_i^{p_{k-1}, p_k}(E)\big). \tag{6}$$

Therefore, the k terms in (3) and (4) have the same bounds and they can differ by at most the difference between these bounds, which is $\varepsilon\mu\ \big(B^{p_{k-1}, p_k}(E)\big)$. Thus, by (3) and (4), the two sides of (2) can differ by ε at most. ∎

The following example shows that an invariant prior for the agent is not necessarily a Bayesian prior for him.

Example. Let $\Omega = \{\omega_1, \omega_2\}$. The agent's type function is: $t(\omega_1) = (1/3, 2/3)$, and $t(\omega_2) = (2/3, 1/3)$. Then, the measure $\mu = (1/2, 1/2)$ is an invariant prior for the agent, because for each E,

$$\mu(E) = \sum_{k=1}^{2} t(\omega_k)(E)\mu\big(\{\omega_k\}\big).$$

But for $E = \{\omega_2\}$,

$$\mu\big(E \mid B_1^{2/3}(E)\big) = \mu\big(E \mid \neg E\big) = 0,$$

which violates (1).

In this example, the agent has different types in each state, and each of his types ascribes positive probability to the other type. This is, as we will see, an essential feature of the example. To understand it, we turn now to the notion of introspection.

§4. Introspection

The agent is *introspective* when he is certain of his type. In order to define this statement formally, we denote by $T(\omega)$ the set of states in which the agent's type is $t(\omega)$, that is,

$$T(\omega) = \{\omega' \mid t(\omega') = t(\omega)\}.$$

The following proposition claims that $T(\omega)$ is an event, i.e., a measurable set.

Proposition 1. *For each state ω,*

$$T(\omega) = \bigcap_{\{p,E \;\mid\; \omega \in B^{p}(E)\}} B^p(E), \tag{7}$$

where the intersection is taken for rational p in $[0,1]$ and $E \in \Sigma_0$. Therefore, as a countable intersection of events, $T(\omega)$ is an event.

Proof: Suppose $\omega' \in T(\omega)$. Then $\omega \in B^p(E)$ iff $\omega' \in B^p(E)$ and therefore ω' is in the set on the right hand side of (7). Conversely, suppose ω' is in the set on the right hand side of (7). Fix $E \in \Sigma_0$. Then, ω' is in $B^p(E) \cap B^q(\neg E)$ for all rational $p < q$ whenever ω is in this event. Thus $t(\omega)(E)$ and $t(\omega')(E)$ belong to the same intervals $[p, q)$ with rational ends. But this implies that $t(\omega)(E) = t(\omega')(E)$. Hence, $t(\omega)$ and $t(\omega')$ are two measures that agree on Σ_0 and therefore the same measure. Hence $\omega' \in T(\omega)$. ∎

Definition 4. *The agent is introspective at ω if*

$$t(\omega)\big(T(\omega)\big) = 1. \tag{8}$$

The following proposition gives a description of introspection in terms of the agent's belief operators, thereby showing, that that being introspective is an event. The measurability of this set is required in the next section.

Proposition 2. *The set of all states in which the agent is introspective is the event*

$$\bigcap_{p,E} \neg B^p(E) \cup B^1\big(B^p(E)\big), \tag{9}$$

where the intersection is taken over rational p in $[0,1]$ and $E \in \Sigma_0$.

Proof: A state ω is in (9), iff it satisfies for all rational p and $E \in \Sigma_0$:

$$\text{If } \omega \in B^p(E), \text{ then } \omega \in B^1\big(B^p(E)\big).$$

But this condition holds, iff

$$\omega \in \bigcap_{\{p,E \mid \omega \in B^p(E)\}} B^1\big(B^p(E)\big),$$

or equivalently, iff

$$\omega \in B^1\Big(\bigcap_{\{p,E \mid \omega \in B^p(E)\}} B^p(E)\Big).$$

But, by (7), this last condition is exactly what (8) says. ∎

§5. An equivalence theorem

As the example in section 3 shows, an invariant prior is not necessarily a Bayesian prior. But the agent in this example is not introspective. Indeed, it turns out that when the agent is introspective almost everywhere with respect to the invariant prior, then it is indeed a Bayesian prior. Moreover, a Bayesian prior for the agent always gives probability 1 to the event that the agent is introspective. Thus, introspection is not a prerequisite for the definition of Bayesian agents; rather it is implied by Bayesianism. Summarizing this gives the following equivalence theorem.

Theorem 2. *The probability measure μ in $\Delta(\Omega)$ is a Bayesian prior for the agent, iff it is an invariant prior for him, and the agent is introspective μ-almost everywhere.*

Proof: Suppose that μ is a Bayesian prior for the agent. By Theorem 1 it is an invariant prior. We show now that μ-a.e. the agent is introspective. Assume, to the contrary, that he is not. Then by (9),

$$\mu\Big(\neg \bigcap_{p,E} \neg B^p(E) \cup B^1\big(B^p(E)\big)\Big) > 0,$$

where the intersection is countable. Therefore, for some p and E,

$$\mu\Big(B^p(E) \cap \neg B^1\big(B^p(E)\big)\Big) > 0.$$

But $\neg B^1\big(B^p(E)\big) = \cup_{r>0} B^r\left(\neg B^p(E)\right)$, where r is rational. Hence for some $r > 0$,

$$\mu\Big(B^p(E) \cap B^r\left(\neg B^p(E)\right)\Big) > 0. \tag{10}$$

Applying (1), as the conditional probability is defined by (10), yields

$$\mu\Big(\neg B^p(E)\big) \mid B^p(E) \cap B^r\left(\neg B^p(E)\right)\Big) > r > 0.$$

But this is a contradiction, since $\neg B^{\ p}(E)$ and the conditioning event are disjoint.

Conversely, suppose that μ is an invariant prior and the agent is introspective μ-almost everywhere. Applying (2) to $\cap_{k=1}^{m} B^{p_k}(E_k) \cap E_l$, we have

$$\mu\Big(\bigcap_{k=1}^{m} B^{p_k}(E_k) \cap E_l\Big) = \int t(\omega)\Big(\bigcap_{k=1}^{m} B^{p_k}(E_k) \cap E_l\Big)\ d\mu(\omega). \tag{11}$$

Now, if $\omega \in \cap_{k=1}^{m} B_i^{p_k}(E_k)$, then $T(\omega) \subseteq \cap_{k=1}^{m} B_i^{p_k}(E_k)$, and otherwise, $T(\omega) \subseteq \neg\cap_{k=1}^{m} B_i^{p_k}(E_k)$. Hence, if the agent is introspective at ω, then $t(\omega)\ \left(\cap_{k=1}^{m} B^{p_k}(E_k)\right)$ is 1 in the first case, and 0 in the second. Therefore, $t(\omega)\ \left(\cap_{k=1}^{m} B^{p_k}(E_k) \cap E_l\right) = t_i(\omega)(E_l) \geq p_l$, in the first case, and $t(\omega)\ \left(\cap_{k=1}^{m} B^{p_k}(E_k) \cap E_l\right) = 0$, in the other.

As the agent is μ-almost everywhere introspective, we can rewrite (11) as,

$$\mu\Big(\bigcap_{k=1}^{m} B^{p_k}(E_k) \cap E_l\Big) = \int_{\bigcap_{k=1}^{m} B^{p_k}(E_k) \cap E_l} t(\omega)(E_l)\ d\mu(\omega).$$

Since the integrand is no less than $p_{\ l}$, we conclude that,

$$\mu\Big(\bigcap_{k=1}^{m} B_i^{p_k}(E_k) \cap E_l\Big) \geq p_l\, \mu\Big(\bigcap_{k=1}^{m} B_i^{p_k}(E_k)\Big),$$

which is precisely (1). ∎

§5. Priors endogenized

In the previous sections Bayesian and invariant priors differ from agents' beliefs in the following way. The latter are defined in each state and therefore are themselves the subject of beliefs. In this sense they are endogenous. Priors were not defined per state and therefore they are given exogenously. In this section they become part of the model such that it becomes meaningful to talk about the event that a certain agent has a prior. We express such events purely in terms of belief operators, much the same way the event that an agent is introspective is expressed in (9).

We call agent 0 the observer and assign his beliefs the role of being a prior for some, or all, agents. Thus, we say that the observer type at ω is a Bayesian prior (or an invariant prior) for agent i if the probability measure $t_{\ 0}(\omega)$ is a Bayesian prior (or an invariant prior) for i. We denote by [i-Bayesian] the set of all states in which the observer is a Bayesian prior for i, and by [i-invariant] the set of all states in which he is an invariant prior for i.

Proposition 3. *The set* [i-Bayesian] *is the event*

$$\bigcap \neg B_0^q\big(\cap_{k=1}^m B_i^{p_k}(E_k)\big) \;\cup\; B_0^{qp_l}\big(\cap_{k=1}^m B_i^{p_k}(E_k)\cap E_l\big), \tag{12}$$

where the intersection is over all $m \geq 1$, all events $E_1,\ldots,E_m$, in Σ_0 and all rational numbers $p_1,\ldots,p_m$ and q in $[0,1]$.

Proof: Consider a weaker definition of a Bayesian prior, in which the measure μ is required to satisfy (1) only when the numbers $p_1,\ldots,p_m$ are rational, and the events $E_1,\ldots,E_m$ are in Σ_0. It is easy to see that even for this weaker definition, the proof that μ is an invariant prior (Theorem 1), and that the agent is μ-a.e. introspective (Theorem 2) go through with almost no change. In the proof of Theorem 1, the numbers $p_0,\ldots,p_m$ should be chosen to be rational, as well as the number r defining the events F^r. We have also to assume that E is in Σ_0. The proof shows, then, that the two sides of (2) coincide for each E in Σ_0. But as each side of (2) is a probability measure in $\Delta(\Omega)$, they must coincide.

The proof that the agent is μ-a.e. introspective, in Theorem 2, does not require any changes, as p and E in this proof are taken from (9), where p is rational and E is in Σ_0.

Thus, the weaker definition of a Bayesian prior implies that μ is an invariant prior and the agent is μ-a.e. introspective. But these two conditions imply, by Theorem 2, the stronger definition of a Bayesian prior, and thus the stronger and weaker definitions are equivalent.

Now, let $\mu = t_0(\omega)$. Let F be the intersection in (12), where $m \geq 1$, the events $E_1,\ldots,E_m$, and the numbers $p_1,\ldots,p_m$, are all fixed, and only q varies over rationals in $[0,1]$. Then, $\omega \in F$, iff, whenever $\mu\,\big(\cap_{k=1}^m B_i^{p_k}(E_k)\big) \geq q$, it is also the case that $\mu\big(\cap_{k=1}^m B_i^{p_k}(E_k)\cap E_l\big) \geq qp_l$. But this is equivalent to (1). Thus μ is in the set in (12) iff it satisfies the weaker definition of a Bayesian prior. ∎

Proposition 4. *The set* [i-invariant] *of all states in which i is introspective is the event*

$$\bigcap \neg \cap_{k=1}^m B_0^{r_k,s_k}\big(B_i^{p_{k-1},p_k}(E)\big) \;\cup\; B_0^{\sum_{k=1}^m r_k p_{k-1},\sum_{k=1}^m s_k p_k}(E). \tag{13}$$

where the intersection is over all events E in Σ_0, all $m \geq 1$, and all rational numbers $0 = p_0 < p_1 < \cdots < p_m = 1$, $r_1,\ldots,r_m$, $s_1,\ldots,s_m$, in $[0,1]$, such that $r_k < s_k$ for $k = 1,\ldots,m$.

Proof: Denote $\mu = t_0(\omega_0)$. Observe, first, that ω_0 is in (13), iff, for any events and numbers as described in the proposition, whenever

$$r_k \leq \mu\big(B_i^{p_{k-1},p_k}(E)\big) < s_k, \tag{14}$$

for $k = 1,\ldots m$, then

$$\sum_{k=1}^m r_k p_{k-1} \leq \mu(E) < \sum_{k=1}^m s_k p_k. \tag{15}$$

Assume, now, that μ is an invariant prior for i. We show that (14) implies (15). To see this, add up all the m inequalities described in (6), and substitute for $\mu\left(B_i^{p_{k-1},p_k}(E)\right)$ its lower and upper bounds from (14). Also substitute, by (2), $\mu(E)$ for the integral. This gives (15).

Conversely, suppose that (14) implies (15). We show that for any $\varepsilon > 0$ the two sides of (2), when $E \in \Sigma_0$, cannot differ by more than ε. This implies that the two sides of (2) are the same for each event E. Choose m and rational numbers in $[0,1]$, p_k, such that $|p_k - p_{k-1}| \leq \varepsilon/2$. Choose also rational r_k, s_k which satisfy (14) and such that $0 < s_k - r_k < \varepsilon/2$ for $k = 1,\dots,m$. Then (15) holds. But by (6), $\int t_i(\omega)(E)\, d\mu(\omega)$ has the same bounds as $\mu(E)$ in (15). Thus, the difference between them cannot exceed the difference between these two bounds, which is at most ε. ∎

Denoting the event that agent i is introspective by [i-introspective] we can now re-formulate Theorem 2, as follows.

Theorem 2. *For each agent i,*

$$[i\text{-Bayesian}] = [i\text{-invariant}] \cap B_0^1\big([i\text{-introspective}]\big).$$

§6. Common priors

Let $J = \{1,\dots,n\}$ be the set of all agents other than the observer.

Definition 5. *The observer type is a common Bayesian prior at ω, if it is a Bayesian prior for each $i \in J$ at this state. The event that the observer is a common Bayesian prior is:* [common B-prior] $= \cap_{i\in J}[i\text{-Bayesian}]$. *The event* [common Invar-prior] *is defined mutatis mutandis.*

Common Bayesian and invariant priors are closely related to the notion of common certainty, which we define next. The event that all players are certain of E is,

$$B^1(E) = \bigcap_{i\in J} B_i^1(E).$$

The event that E is *common certainty* is,

$$C(E) = \bigcap_{k\geq 0} \left(B^1\right)^k(E),$$

where $\left(B^1\right)^1(E) = B^1(E)$ and for $k \geq 1$, $\left(B^1\right)^{k+1}(E) = B^1\big(\left(B^1\right)^k(E)\big)$.

Proposition 5. *If the observer type is a common invariant prior, then if he is certain of some event, he is also certain that it is common certainty. That is, for each E,*

$$[\text{common Invar-prior}] \cap B_0^1(E) \subseteq B_0^1\big(C(E)\big).$$

Proof: Suppose $\omega_0 \in$ [common Invar-prior]$\cap B_0^1(E)$, and let $\mu = t_0(\omega_0)$. Then, $\mu(E) = 1$ and therefore, by (8), for each $i \in J$,

$$\int t_i(\omega)(E)\ d\mu(\omega) = 1.$$

But as the integrand is bounded by 1, $\mu\left(\{\omega \mid t_i(\omega)(E) = 1\}\right) = 1$. That is $\omega_0 \in B_0^1\big(B_i^1(E)\big)$. As this is true for each $i \in J$, it follows that $\omega_0 \in B_0^1\big(B^1(E)\big)$. But now we can conclude by induction that $\omega_0 \in B_0^1\big(\left(B^1\right)^k(E)\big)$, for all k and hence $\omega_0 \in B_0^1\big(C(E)\big)$. ∎

We are able now to formulate a condition on the observer type that guarantees that when this type is a common invariant prior, the type is common certainty. For this purpose we need to show that the set of states in which the observer type is common certainty is an event.

The observer's type at ω is common certainty if $\omega \in C\big(T_0(\omega)\big)$, where $T_0(\omega) = \{\omega' \mid t_0(\omega') = t_0(\omega)\}$.

Proposition 6. *The set of all states in which the observer's type is common certainty, denoted* [observer type cc] *is the event*

$$\bigcap_{k,p,E} \neg B_0^p(E) \cup \left(B^1\right)^k\big(B_0^p(E)\big), \tag{16}$$

where k is an integer, p a rational number in $[0,1]$ and $E \in \Sigma_o$.

Proof: A state ω is in (16), iff

$$\omega \in \bigcap_{k,\{p,E \mid \omega \in B^p(E)\}} \left(B^1\right)^k\big(B_0^p(E)\big),$$

or equivalently,

$$\omega \in \bigcap_k \left(B^1\right)^k\Big(\bigcap_{\{p,E \mid \omega \in B^p(E)\}} B_0^p(E)\Big).$$

By (7) this is equivalent to $\omega \in C\big(T_0(\omega)\big)$. ∎

Theorem 4. *If the observer type is a common invariant prior and the observer is introspective then he is certain that there is common certainty of his type. That is,*

$$[\text{common Invar-prior}] \cap [\text{0-introspective}] \subseteq B_0^1\big([\text{observer type cc}]\big).$$

Proof: If $\omega \in$ [0-introspective], then $\omega \in B_0^1\big(T_0(\omega)\big)$. The claim now follows from Proposition 5. ∎

The agreement theorem (Aumann (1985)) is formulated here in terms that are amenable to syntactic formulation. Consider an event E, and numbers $0 \le p_i, q_i \le 1$ such that $p_i + q_i \le 1$ for $i \in J$. The event $D = \cap_{i\in J} B_i^{p_i}(E) \cap B_i^{q_i}(\neg E)$ is a *disagreement* if $\cap_{i\in J}[p_i, 1-q_i] = \emptyset$.

Theorem 5. *If D is a disagreement, then*

$$[\text{common B-prior}] \subseteq B_0^1\big(\neg C(D)\big).$$

That is, when the observer is common Bayesian prior, then he is certain that there cannot be a common certainty of disagreement.

Proof: We need three preliminaries:

(a) $C(D) \subseteq B_i^1\big(C(D)\big)$. Indeed, by definition of B^1, for each $n \geq 1$, $(B^1)^{n+1}(D) \subseteq B_i^1\big((B^1)^n(D)\big)$. Intersecting the left and the right hand sides of these inclusions yields the desired inclusion.

(b) $[i\text{-introspective}] \subseteq B_i^1\big(C(D)\big) \cup B_i^1\big(\neg C(D)\big)$. To see this, note that if i is introspective at ω and $\omega \in \neg B_i^1\big(C(D)\big)$ then i must be certain of this event, i.e., $[i\text{-introspective}] \cap \neg B_i^1\big(C(D)\big) \subseteq B_i^1\big(\neg B_i^1\big(C(D)\big)\big)$. By monotonicity and (a) the latter event is a subset of $B_i^1\big(\neg C(D)\big)$.

(c) For each F, $[i\text{-Bayesian}] \subseteq B_0^1\big(\neg B_i^1(F) \cup F\big)$. Indeed, let $\omega_0 \in [i\text{-Bayesian}]$ and $\mu = t_0(\omega_0)$. Then, as μ is an invariant prior, $\mu\big(B_i^1(F) \cap \neg F\big) = \int t_i(\omega)\big(B_i^1(F) \cap \neg F\big)\, d\mu$. But i is introspective at ω, μ-a.e. and for such ω, $t_i(\omega)\big(B_i^1(F)\big)$ is 1 if $\omega \in B_i^1(F)$, and 0 otherwise. Thus the integral is $\int_{B_i^1(F)} t_i(\omega)(\neg F)\, d\mu$, which is 0.

Now, let $\omega_0 \in [\text{common B-prior}]$ and let $\mu = t_0(\omega_0)$. Then, for each $i \in J$,

$$\mu\big(E \cap C(D)\big) = \int t_i(\omega)\big(E \cap C(D)\big)\, d\mu.$$

By applying (c) to $F = C(D)$, the event $C(D) \cup \neg B_i^1\big(C(D)\big)$ has μ measure 1 and thus can be the domain of integration. We now examine the value of the integrand on this domain. If $\omega \in C(D)$, then by (a), $t_i(\omega)\big(E \cap C(D)\big) = t_i(\omega)(E)$. As i is μ-a.e. introspective, it follows by (b) that, μ-a.e., if $\omega \in \neg B_i^1\big(C(D)\big)$, then $\omega \in B_i^1\big(\neg C(D)\big)$, and hence $t_i(\omega)\big(E \cap C(D)\big) = 0$. Thus the last integral is $\int_{C(D)} t_i(\omega)(E)\, d\mu$. By definition, $C(D) \subseteq B_i^1(D)$ and by Proposition 3, μ-a.e., $B_i^1(D) \subseteq B_i^{p_i}(E) \cap B_i^{q_i}(\neg E)$. Thus, the integrand is in the interval $[p_i, 1 - q_i]$, and evaluating the integral we find

$$p_i \mu\big(C(D)\big) \leq \mu\big(E \cap C(D)\big) \leq (1 - q_i)\mu\big(C(D)\big).$$

Since this is true for all $i \in J$, and D is a disagreement, it must be the case that $\mu\big(C(D)\big) = 0$. ∎

References

Aumann, R. (1976), "Agreeing to Disagree," *The Annals of Statistics*, 4, pp 1236-1239.

Aumann, R. (1989), "Notes on Interactive Epistemology," Cowles Foundation for Research in Economics working paper (1989).

Fagin, R. and J. Y. Halpern (1988), "Reasoning about Knowledge and Probability," in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge,* Morgan Kaufmann, California.

Harsanyi, J.C. (1967-68), "Games with Incomplete Information Played by Bayesian Players," parts I,II, & III, *Man. Sc.* , 14, pp. 159-182, 320-334, 486-502.

Heifetz, A. and D. Samet (1996), "Topology-Free Typology of Beliefs," mimeo.

Mertens, J.F. and S. Zamir, (1985), "Formulation of Bayesian Analysis for Games with Incomplete Information," *Int. Jour. Game Theory* 14, pp. 1-29.

Monderer, D. and D. Samet (1989), "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior* 1, pp. 170-190.

Samet, D. (1996), "Common Prior and Markov Chains." mimeo.

Skyrms, B. (1984), "Pragmatics and Empiricism," Yale University Press, New Haven and London.