Agreeing to Disagree

Topics in Epistemology
Oxford, 17/5/12

Daniel Rothschild
daniel.rothschild@philosophy.ox.ac.uk

1 Partition Models

The main alternative to the iterative method of discussing common knowledge, is the partition model first given by Aumann [1976].
Assume a set of worlds W which constitute logical space. We let lower case letters in general denote subsets of w, propositions. (Like Aumann we won't be concerned with defining a formal language here.)
We might associate with each agent i a partition k_i over W (we will assume W is finite, though this assumption is not crucial). A partition of W is a set of disjoint sets whose union is W. Now we say that in any world w what i knows in w is just the cell of the partition k_i which w is in (which we write $k_i(w)$). More formally we can say that the proposition that i knows a proposition p , which we write $K_i(p)$ is itself just:

$$K_i(p) = \{w \in W : k_i(w) \subseteq p\}$$

Associated with a partition is a more familiar philosophical tool a relation r_k over W. A partition is formed by an equivalence relation, and as those familiar with modal logic know, the **S5** axioms as modeled by Kripke semantics require an equivalence relation. So the partition model of knowledge that Aumann uses presupposes the **S5** epistemic logic for knowledge.
Heres one (overly long) version of axioms and rules governing **S5**:

N if p is a theorem, then Kp

K $K(p \to q) \to (Kp \to Kq)$

T $Kp \to p$ (factivity)

4 $Kp \to KKp$ (positive introspection, KK principle).

5 $\neg Kp \to K\neg Kp$ (negative introspection)

Terminology: **T** is everything up to **T**, **S4** is everything up to **4**, **S5** is all of them.
Using the partition method allows for an elegant definition of common knowledge (Aumann's). In particular the partition associated with what is common knowledge is just the finest coarsening of all the partitions of the group (the meet). (A coarsening of a partition is just another partition that can only contain unions of cells of the first partition as its cells.) Aumann shows how (in the **S5** model) this condition is equivalent to the iterative one.
His basic argument goes as follows: A world w' is EPISTEMICALLY REACHABLE for a set of agents I from w if there is a series of worlds $w_1 \dots w_n$ such that for $w_1 = w$ and $w_n = w'$ and for any natural number $k \leqslant n$, there is an i in I such that $w_k \in p_i(w_{k-1})$. If something is common knowledge at w among I, then its negation cannot be epistemically reachable from w. The set of all propositions whose negations are not epistemically reachable w is just the set of all propositions that contain the cell at w of the finest common coarsening of all the k_is.
From this it follows that $c(w)$ just is the cell of the finest common coarsening of k_1 and k_2 that w is in.
Here is an example, of partition with three worlds and two epistemic agents:

	i_1	i_2
w_1	a	a
w_2	b	a
w_3	c	b

In w_1 and w_2, $\{w_1, w_2\}$ is common knowledge. Note that the model itself is (implicitly) part of the knowledge of i_1 and i_2.
A proposition p is SELF-EVIDENT to an agent i iff whenever it occurs i knows it occurs. So it is self-evident iff $p \supset K_i p$ (where $\supset$ is material implication). Only unions of partition cells for i are self-evident. If a proposition is self-evident for both i and j then it will be common knowledge (as will everything entailed by it). In fact the converse holds as well: if p is common knowledge among I, then there is some self-evident proposition p' for each i in I such that $p' \supset p$.[1]

[1] This might require that W be finite, at least that's the only way I can think of proving it. Consider some proposition p that is common knowledge at w. There must be a proposition set $p' \subseteq p$ that is a strengthening of p that is also common knowledge

Note that how we describe the worlds in W is sort of implicit common knowledge. But not modeled directly.

2 Agreement theorem

We now have a definition of common knowledge, so we can define such things as common knowledge of subjective probabilities. We need to say something about what probabilities are. We will first define PRIOR probabilities in this framework.

Priors For each agent i, i's PRIOR PROBABILITIES FUNCTION is a probability function p_i defined over $\mathcal{P}(W)$.[2]

Agents i and j have COMMON PRIORS if $p_i = p_j$. In any given world w the POSTERIOR probabilities of i in w is written p_{iw}. We assume that posteriors are formed by conditioning on i's knowledge in w:

Conditioning on Knowledge For all a in $\mathcal{P}(W)$, $p_{iw}(a) = p_i(a|k_i(w))$. (Where $p(a|b) = \frac{p(a \cap b)}{p(b)}$)

Given the priors of i, and a knowledge partition, we have thus determined i's posteriors across w. We can now define new propositions in W based on this. For example, there is the proposition that i's posteriors assign .5 to some propositions a. This is $\{w : p_{iw}(a) = .5\}$, and we'll write this as $P_i(a = .5)$.
Now we can state and prove Aumann's agreement theorem.

Agreement For two agents 1 and 2, with knowledge partitions k_1, k_2 and common priors, p, for any proposition a and real numbers x_1 and x_2 in $[0, 1]$, and world w, if it is common knowledge at w that $P_1(a = x_1)$ and that $P_2(a = x_2)$ then $x_1 = x_2$.

Proof. Call c the meet of k_1 and k_2. $K(w)$ is cell of K that includes w. By the common knowledge assumption we know that, $c(w) \subseteq p_1(a = x_1)$. Since K is the meet of k_1 and k_2, $c(w)$ can be built up out of a set of cells in k_1 (meets are made from a common coarsening of two partitions). Call this set S, where S is disjoint and the union of S is $c(w)$, for each s in S, $p(a|s) = x_1$ since at each w in $c(w)$, $p_{1w}(a) = x_1$. It follows that $p(a|c(w)) = x_1$. By analogous reasoning $p(a|c(w)) = x_2$. So $x_1 = x_2$. □

We can generalize Aumann's theorem considerably. A very abstract version of it goes as follows.

Abstract Agreement Let k_1 and k_2 be partitions and c be a coarsening of k_1 and k_2 (i.e. each cell of c is equal to a union of cells of k_1 as well as a union of cells of k_2) . Let π be a function with domain $\mathcal{P}(W)$ that satisfies the following condition for any two disjoint subsets s and s_1 of W, if $\pi(s) = \pi(s')$, then $\pi(s) = \pi(s \cup s')$ (this is often called the SURE-THING PRINCIPLE). For any world w, if there exists x_1 and x_2 such that for every world $w' \in c(w)$ is such that $\pi(k_1(w)) = x_1$ and $\pi(k_2(w)) = x_2$, then $x_1 = x_2$.[3]

Proof. Much as before: we show that $c(w)$ which is made up of cells of k_1 is such that $\pi(c(w)) = x_1$ by the sure-thing principle. Then, by symmetry, $\pi(c(w)) = x_2$ so, $x_1 = x_2$. □

Aumann's agreement theorem is a special case of this where $\pi(x)$ is the function $p(a|x)$ (for some proposition a), which satisfies the sure-thing principle. We might think that an expected utility function u also satisfies the sure-thing principle.

3 Consequences

- Agents who can and want to communicate with common priors should not disagree. How surprising is this? Consider consecutive announcements of probability. Aumann's coin example:
- No common knowledge of different actions. Note that expected utilities satisfy sure-thing principle. An Aumann example.

  > A murder has been committed. To increase the chances of conviction, the chief of police puts two detectives on the case, with strict instructions to work independently, to

at w. In any world $w' \in p'$, p' is common knowledge (argument otherwise then in w, $p' \backslash \{w'\}$ is common knowledge so a contradiction. So p' is self-evident to all agents.

[2] $p : P(W) \to [0, 1]$ satisfying (i) $p(W) = 1$ and (ii) for any disjoint subsets a and b of W, $p(a) + p(b) = p(a \cup b)$.

[3] The fact that the sure-thing principle that was all that is needed was independently noted by Cave [1983] and Bacharach [1985].

> 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.

- No speculative trades. Milgrom and Stokey [1982]:

  > Our central result is that, regardless of the institutional structure, if the initial allocation is ex ante Pareto-optimal (as occurs, for example, when it is the outcome of a prior round of trading on complete, competitive markets), then the receipt of private information cannot create any incentives to trade.

  Idea: trade requires common knowledge of willingness of both parties. But any private information about best action will need to be identical by version of Agreeing to Disagree result.

4 Assumptions

4.1 Conditionalizing

Two issues

1. Do we update by conditionalizing?
2. Do we condition on knowledge?

While conditionalizing is widely accepted, most take *subjective* probabilities to be based on *beliefs* rather than knowledge.

4.2 Common Priors

The common priors assumption is extremely strong and, of course, completely necessary for this. Note that it's just not common priors, but really common knowledge of common priors (at least on the intuitive reading of the theorem).

4.3 Partition Model of Knowledge

As we said this assume **S4** (positive introspection) and **S5** (negative introspection) in addition to **T** (factivity). We might think this is too much, after all Williamson and other have given rather strong arguments against **S4** and **S5**.

5 Non-partition models of knowledge

Suppose we weaken our model of knowledge so that knowledge no longer partitions out logical space. If we use a model of knowledge in **T**, the theorem will not hold. Here is a model showing this: we have three worlds w_1, w_2, w_3. For agent 1, there is just the trivial a knowledge partition $\{w_1, w_2, w_3\}$. For agent 2, in w_1, the proposition known is $\{w_1, w_2\}$, in w_2 and w_3 it is $\{w_2, w_3\}$. Suppose there are even priors. Let $a = \{w_2\}$. In any world w, $p_{1w}(a) = \frac{1}{3}$, but $p_{2w}(a) = \frac{1}{2}$. Anything true in all worlds is common knowledge (trivially) so we can have different common knowledge of posteriors despite the same priors in **T**.

6 Belief versions

Suppose we weakened our system by assuming that agents conditionalize on belief rather than knowledge. In this case, the crucial issue is whether under the common priors assumptions we can have common belief in a subjective probabilities without having those probabilities.
Collins [1997] argues that the answer is 'no'. Here is a very simple model to make the point. Suppose there are two worlds w_1, w_2, and even common priors over them.

Let us assume a similar model of belief as Aumann had of knowledge. The function b takes us from world w to all the worlds compatible with beliefs at w. We no longer assume that $w \in b(w)$, but we do assume that if w' is in $b(w)$ then $b(w) = b(w')$ (this gets us positive and negative introspection of belief).

Suppose that in w_1 and w_2 agent 1 believes $\{w_1\}$, whereas in w_1 and w_2 agent 2 believes that $\{w_1, w_2\}$. If posteriors condition on belief, then for all w, $p_{1w}(\{w_1\}) = 1$, whereas $p_{2w}(\{w_1\}) = \frac{1}{2}$. Anything believed in all worlds is common belief (on iterative or any other sense).

Note that here we seem to have a weird notion of priors: after all 1 assigns a prior of 2 to w_2 but in no case assigns it any posterior. This is weird. Even adding more worlds will not rescue this problem if we require that if w_1 is compatible with beliefs at some world, then w_1 must be be compatible with beliefs at w_1.

Collin calls worlds like w_1 epistemic BLIND SPOTS since they are never believed possible. He argues that to make sense of priors we should assign them a zero prior. Now if we assume that agents 1 and 2 each assign zero priors to all their blind spots, *then* we can prove an analogue of the agree to disagree theorem (relatively trivially). But, he argues that while we should assign 0 to our own blindspots there is no reason to do so to others' blindspots. So, in the end, he argues the common prior assumption is unsustainable.

References

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

Michael Bacharach. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory37*, 167–190, 1985.

Jonathan A. Cave. Learning to agree. *Economic Letters*, 12:147–152, 1983.

John Collins. How can we agree to disagree? manuscript, Columbia University, 1997.

Paul Milgrom and Nancy Stokey. Information, trade and common knowledge. *Journal of Economic Theory*, 26:17–27, 1982.