Agreeing to Disagree

Robert J. Aumann

The Annals of Statistics, Vol. 4, No. 6 (Nov., 1976), 1236-1239.

Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28197611%294%3A6%3C1236%3AATD%3E2.0.CO%3B2-D

The Annals of Statistics is currently published by Institute of Mathematical Statistics.

The Annals of Statistics
1976, Vol. 4, No. 6, 1236–1239

AGREEING TO DISAGREE[1]

BY ROBERT J. AUMANN

Stanford University and the Hebrew University of Jerusalem

Two people, 1 and 2, are said to have *common knowledge* of an event E if both know it, 1 knows that 2 knows it, 2 knows that 1 knows is, 1 knows that 2 knows that 1 knows it, and so on.

THEOREM. *If two people have the same priors, and their posteriors for an event A are common knowledge, then these posteriors are equal.*

If two people have the same priors, and their posteriors for a given event A are common knowledge, then these posteriors must be equal. This is so even though they may base their posteriors on quite different information. In brief, people with the same priors *cannot agree to disagree*.

We publish this observation with some diffidence, since once one has the appropriate framework, it is mathematically trivial. Intuitively, though, it is not quite obvious; and it is of some interest in areas in which people's beliefs about each other's beliefs are of importance, such as game theory[2] and the economics of information.[3] A "concrete" illustration that may clarify matters (and that may be read at this point) is found at the end of the paper.

The key notion is that of "common knowledge." Call the two people 1 and 2. When we say that an event is "common knowledge," we mean more than just that both 1 and 2 know it; we require also that 1 knows that 2 knows it, 2 knows that 1 knows it, 1 knows that 2 knows that 1 knows it, and so on. For example, if 1 and 2 are both present when the event happens and see each other there, then the event becomes common knowledge. In our case, if 1 and 2 tell each other their posteriors and trust each other, then the posteriors are common knowledge. The result is not true if we merely assume that the persons know each other's posteriors.

Formally, let $(\Omega, \mathscr{B}, p)$ be a probability space, $\mathscr{P}_1$ and $\mathscr{P}_2$ partitions of Ω whose join[4] $\mathscr{P}_1 \vee \mathscr{P}_2$ consists of nonnull events.[5] In the interpretation, $(\Omega, \mathscr{B})$ is the space of states of the world, p the common prior of 1 and 2, and $\mathscr{P}_i$ the

Received November 1975.

[1] This work was supported by National Science Foundation Grant SOC74-11446 at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

[2] Cf. Harsanyi (1967-1968); also Aumann (1974), especially Section 9j (page 92), in which the question answered here was originally raised.

[3] Cf. e.g., Radner (1968) and (1972); also the review by Grossman and Stiglitz (1976) and the papers quoted there.

[4] Coarsest common refinement of $\mathscr{P}_1$ and $\mathscr{P}_2$.

[5] Events whose (prior) probability does not vanish.

AMS 1970 *subject classifications.* Primary 62A15, 62C05; Secondary 90A05, 90D35.

Key words and phrases. Information, subjective probability, posterior, statistics, game theory, revising probabilities, concensus, Harsanyi doctrine.

information partition of i; that is, if the true state of the world is ω, then i is informed of that element $\mathbf{P}_i(\omega)$ of $\mathscr{P}_i$ that contains ω. Given ω in Ω, an event E is called *common knowledge at* ω if E includes that member of the meet[6] $\mathscr{P}_1 \wedge \mathscr{P}_2$ that contains ω. We will show below that this definition is equivalent to the informal description given above.

Let A be an event, and let $\mathbf{q}_i$ denote the posterior probability $p(A \mid \mathscr{P}_i)$ of A given i's information; i.e., if $\omega \in \Omega$, then $\mathbf{q}_i(\omega) = p(A \cap \mathbf{P}_i(\omega))/p(\mathbf{P}_i(\omega))$.

PROPOSITION. *Let $\omega \in \Omega$, and let q_1 and q_2 be numbers. If it is common knowledge at ω that $\mathbf{q}_1 = q_1$ and $\mathbf{q}_2 = q_2$, then $q_1 = q_2$.*

PROOF. Let P be the member of $\mathscr{P}_1 \wedge \mathscr{P}_2$ that contains ω. Write $P = \bigcup_j P^j$, where the P^j are disjoint members of $\mathscr{P}_1$. Since $\mathbf{q}_1 = q_1$ throughout P, we have $p(A \cap P^j)/p(P^j) = q_1$ for all j; hence $p(A \cap P^j) = q_1 p(P^j)$, and so by summing over j we get $p(A \cap P) = q_1 p(P)$. Similarly $p(A \cap P) = q_2 p(P)$, and so $q_1 = q_2$. This completes the proof.

To see that the formal definition of "common knowledge" is equivalent to the informal description, let $\omega \in \Omega$, and call a member ω' of Ω *reachable from* ω if there is a sequence $P^1, P^2, \cdots, P^k$ such that $\omega \in P^1$, $\omega' \in P^k$, and consecutive P^j intersect and belong alternatively to $\mathscr{P}_1$ and $\mathscr{P}_2$. Suppose now that ω is the true state of the world, $P^1 = \mathbf{P}_1(\omega)$, and E is an event. To say that 1 "knows" E means that E includes P^1. To say that 1 knows that 2 knows E means that E includes all P^2 in $\mathscr{P}_2$ that intersect P^1. To say that 1 knows that 2 knows that 1 knows E means that E includes all P^3 in $\mathscr{P}_1$ that intersect P^2 in $\mathscr{P}_2$ that intersect P^1. And so on. Thus all sentences of the form "i knows that i' knows that i knows$\cdots E$" (where $i' = 3 - i$) are true if and only if E contains all ω' reachable from ω. But the set of all ω' reachable from ω is a member of $\mathscr{P}_1 \wedge \mathscr{P}_2$; so the desired equivalence is established.

The result fails when people merely know each other's posteriors. Suppose Ω has 4 elements $\alpha, \beta, \gamma, \delta$ of equal (prior) probability, $\mathscr{P}_1 = \{\alpha\beta, \gamma\delta\}$, $\mathscr{P}_2 = \{\alpha\beta\gamma, \delta\}$, $A = \alpha\delta$, and $\omega = \alpha$. Then 1 knows that $\mathbf{q}_2$ is $\frac{1}{3}$, and 2 knows that $\mathbf{q}_1$ is $\frac{1}{2}$; but 2 thinks that 1 may not know what $\mathbf{q}_2$ is ($\frac{1}{3}$ or 1).

Worthy of note is the implicit assumption that the information partitions $\mathscr{P}_1$ and $\mathscr{P}_2$ are themselves common knowledge. Actually, this constitutes no loss of generality. Included in the full description of a state ω of the world is the manner in which information is imparted to the two persons. This implies that the information sets $\mathbf{P}_1(\omega)$ and $\mathbf{P}_2(\omega)$ are indeed defined unambiguously as functions of ω, and that these functions are known to both players.

Consider next the assumption of equal priors for different people. John Harsanyi (1968) has argued eloquently that differences in subjective probabilities should be traced exclusively to differences in information—that there is no rational basis for people who have always been fed precisely the same information to maintain different subjective probabilities. This, of course, is equivalent to

[6] Finest common coarsening of $\mathscr{P}_1$ and $\mathscr{P}_2$.

the assumption of equal priors. The result of this paper might be considered evidence against this view, as there are in fact people who respect each other's opinions and nevertheless disagree heartily about subjective probabilities. But this evidence is not conclusive: even people who respect each other's acumen may ascribe to each other errors in calculating posteriors. Of course we do not mean simple arithmetical mistakes, but rather systematic biases such as those discussed by Tversky and Kahnemann (1974). In private conversation, Tversky has suggested that people may also be biased because of psychological factors, that may make them disregard information that is unpleasant or does not conform to previously formed notions.

There is a considerable literature about reaching agreement on subjective probabilities; a recent paper is DeGroot (1974), where a bibliography on the subject may be found. A "practical" method is the Delphi technique (see, e.g., Dalkey (1972)). It seems to me that the Harsanyi doctrine is implicit in much of this literature; reconciling subjective probabilities makes sense if it is a question of implicitly exchanging information, but not if we are talking about "innate" differences in priors. The result of this paper might be considered a theoretical foundation for the reconciliation of subjective probabilities.

As an illustration, suppose 1 and 2 have a uniform prior on the parameter of a coin, and let A be the event that the coin will come up H (heads) on the next toss. Suppose that each person is permitted to make one previous toss, and that these tosses come up H and T (tails) respectively. If each one's information consists precisely of the outcome of his toss, then the posteriors for A will be $\frac{2}{3}$ and $\frac{1}{3}$ respectively. If each one then informs the other one of his posterior, then they will both conclude that the previous tosses came up once H and once T, so that both posteriors will be revised to $\frac{1}{2}$.

Suppose now that each person is permitted to make several previous tosses, but that neither one knows how many tosses are allowed the other one. For example, perhaps both make 4 tosses, which come up $HHHT$ for 1, and $HTTT$ for 2. They then inform each other that their posteriors are $\frac{3}{5}$ and $\frac{1}{3}$ respectively. Now these posteriors may result from a single observation, from 4 observations, or from more. Since neither one knows on what observations the other's posterior is based, he may be inclined to give more weight to his own observations. Some revision of posteriors would certainly be called for even in such a case; but it does not seem clear that it would necessarily lead to equal posteriors.

Presumably, such a revision would take into account each person's prior on the number of tosses available to him and to the other person. By assumption these two priors are the same, but each person gets additional private information—namely, the actual number of tosses he is allotted. By use of the prior and the information that the posteriors are, respectively, $\frac{3}{5}$ and $\frac{1}{3}$, new posteriors may be calculated. If the players inform each other of these new posteriors, further revision may be called for. Our result implies that the process of exchanging information on the posteriors for A will continue until these posteriors are equal.

REFERENCES

[1] AUMANN, R. J. (1974). Subjectivity and correlation in randomized strategies. *J. Math. Econom.* **1** 67–96.

[2] DALKEY, N. C. (1972). *Studies in the Quality of Life.* Lexington Books, Lexington, Mass.

[3] DEGROOT, M. H. (1974). Reaching a consensus. *J. Amer. Statist. Assoc.* **69** 118–121.

[4] GROSSMAN, S. J. and J. E. Stiglitz (1976). Information and competitive price systems. *Amer. Econom. Rev.* **46** 246–253.

[5] HARSANYI, J. (1967–1968). Games of incomplete information played by Bayesian players, Parts I–III. *Management Sci.* **14** 159–182, 320–334, 486–502.

[6] RADNER, R. (1968). Competitive equilibrium under uncertainty. *Econometrica* **36** 21–58.

[7] RADNER, R. (1972). Existence of equilibrium plans, prices, and price expectations in a sequence of markets. *Econometrica* **40** 289–304.

[8] TVERSKY, A. and D. KAHNEMAN (1974). Judgment under uncertainty: Heuristics and biases. *Science* **185** 1124–1131.

THE INSTITUTE FOR MATHEMATICAL STUDIES
IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305