REACHING CONSENSUS ON DECISIONS

Paul J. Krasucki
Dept. of Mathematical Sciences
Rutgers University
Camden College of Arts and Sciences
Camden, NJ 08102
email: krasucki@cancer.rutgers.edu

ABSTRACT

We investigate how like-minded agents can reach consensus on their decisions even if they receive different information.

The model used here was introduced by Aumann, and subsequently refined by Geanakoplos and Polemarchakis, Bacharach, Cave, Parikh and Krasucki ([Aum76,GP82,Cav83,Bac85,PK]).

The main result is that when any number of like-minded agents communicate according to some *fair* protocol whether they want to trade or not, and their decision is based solely on whether the conditional probability of some fixed event exceeds some threshold value, they must reach consensus in a finite time.

We also investigate some necessary conditions which functions communicated have to satisfy in order to guarantee consensus in fair protocols.

1 INTRODUCTION

In order to investigate whether the difference in the kind of information received can justify the speculative trading between rational agents the following model was created ([Aum76,GP82, Cav83,Bac85,PK]).

Let W be a set of possible results of an experiment. There are two agents receiving some information about the result. Information is given by choosing one of the elements of the *ith* partition of W, P_i. It is assumed that P_i's are common knowledge among agents (so *type* of information available to agents is common knowledge). It is also assumed that agents always receive *true* information, if the actual state of the world is x, then for all i, $x \in P_i(x)$.

Both agents are interested in computing a probability of some fixed event, so they could make their decisions based on that. There is given some prior probability distribution on W, and it is shared by both agents. Without any additional information they would both have the same value: $p(E)$. But if an agent 1 learns that the result x is in $P_1(x)$, then he can compute a new probability as $p(E|P_1(x))$. This is his *posterior* probability of E. Similarly, an agent 2 can compute his posterior probability $p(E|P_2(x))$.

There is no a priori reason why $p(E|P_1(x)) = p(E|P_2(x))$, but surprisingly Robert Aumann [Aum76], has shown in 1976 that when the posteriors are common knowledge, then they must indeed be the same. Like-minded agents cannot "agree to disagree".

Aumann didn't address the question how the agents computed their posteriors and how could they become common knowledge. Geanakoplos and Polemarchakis [GP82] first investigated the procedure in which one agent computes his posterior, sends the value to the second agent, who in turn computes his new posterior excluding from his set of possible worlds all the worlds in which the first agent would have sent a different value. Subsequently he sends this new posterior back to the agent one, and process continues until (as they have proved), both agents' posteriors converge to the same value, which will happen after $k+l$ rounds where k, l are numbers of sets in information partitions of agents 1 and 2 respectively.

Cave [Cav83] and Bacharach [Bac85] noticed that the property of the conditional probability function that was crucial in both Aumann's and Geanakoplos-Polemarchakis' proofs is that if the conditional probability of some event is the same in two disjoint sets, then the conditional probability of this event in the union of the two sets will be the same as it was in both of them. This property was called *union consistency* by Cave and the *sure - thing principle* by Bacharach. Cave and subsequently Bacharach proved that if two agents communicate values of some function satisfying the sure-thing principle back and forth revising their sets of possible worlds and therefore recomputing values of the function, they will end up with the same values.

All these results generalise to the case when there are $n > 2$ agents and all the agents in turn communicate their values of a function so all of them hear all these values (and there is common knowledge among them of this fact). This was also investigated by Cave.

The situation where every announcement is public, corresponds in computer science to the case of a broadcast from a reliable source. In economics it corresponds to an auction.

However, prices are not always set up via auctions. Communication in the system is not always synchronous n-cast. There are also private communications between agents, inaccessible to others.

In fact, the situation when there are more than two agents involved, and agents communicate using person-to-person links inaccessible to other agents (binary channels) is much more difficult to analyse. This is because even if there are 3 agents, and their information partitions are common knowledge among them, if they communicate values of f in a ring (1 to 2, 2 to 3, 3 to 1 and so on) then, since 3 doesn't know what 2 has heard from 1 and he can usually imagine more then one scenario which made 2 send the value he has sent, so it is more difficult for him to exclude worlds.

This model was analysed first by Parikh and Krasucki in [PK]. They gave a general formula for updating agents knowledge in n-person case. They also defined a class of *fair* protocols, in which all the agents' information affect all other agents' knowledge. Of course if an agent doesn't communicate with anybody, we cannot expect consensus between this agent and the others.

Parikh and Krasucki showed that for $n > 2$ the sure-thing principle is not enough to guarantee consensus among agents communicating according to a fair protocol. They introduced stronger condition, *weak-convexity* which turned out to guarantee consensus in case of $n \leq 3$.

This condition was still not strong enough to guarantee consensus in case of $n > 3$ (an example was given in which four people communicate values of a weakly convex function according to a fair protocol without reaching consensus). They defined a property of *strong-convexity* which turned out to suffice to guarantee consensus for any number n of people communicating.

As a main result they proved that when any number of people communicate values of $f(X) = P(E|X)$ (expected value of some fixed event E) according to some fair protocol, they must reach consensus.

2 DECISIONS

In many cases agents base their decision whether to trade or not on the expected probability of a certain fixed event. If a probability of this event exceeds some value they will trade, otherwise not. Therefore the results asserting that in some cases agents always reach consensus on the value of conditional probability of an event are very important. They hint that the difference in available information might not be enough to justify trading between rational, like-minded agents.

All the consensus results obtained so far require that the agents send to each other their probabilities of the event E. In the market environment, where traders compete, it is unrealistic to assume that they will communicate to the others their assessment of probability. Instead they just communicate their *decision* (although the decision itself may be based on the assessment of probability). That's the case when a trader announces whether he is willing to trade without specifying what is his expected gain, or what kind of analysis lead him to his conclusions.

Our work shows that even if only decisions are communicated, in case where these decisions are based on a probability of a fixed event, consensus is guaranteed.

We talk about a model in which agents communicate values of "decision functions", functions with two-element range. We assume that the range is {0,1} (1 interpreted as "yes", 0 as "no").

The decision functions are the most difficult ones to analyse. They give least amount of information (except of constant functions, which give no information, but if everybody communicates value of the same constant function, consensus is trivially reached). The results of [PK] on strongly-convex functions are here of no help, since non-trivial two-valued functions cannot be strongly convex.

For two-valued functions weak-convexity is equivalent to union-consistency. By the result of [PK] consensus is guaranteed for $n \leq 3$ in such a case. We show that although in general union-consistency is not enough to guarantee consensus for $n > 3$, for the decision functions based on conditional probability consensus on their values in finitely many steps is guaranteed for any n and any fair protocol.

Formally, we prove consensus theorem for $f(X) = 1$ iff $P(E|X) > \alpha$ for some fixed event E and some constant α.

We also discuss some necessary conditions for guaranteeing consensus.

3 DEFINITIONS

All the notions defined here follow [PK].

W is the space of possible worlds. There is n participants with *finite* partitions P_i of W. P^+ is the common refinement (join) of the P_i (obviously P^+ is finite).

Let $cl(W, P^+)$ be the set of all subsets of W which are unions of elements of P^+ (we will call them *closed* sets).

We will consider functions $f : cl(W, P^+) \to \{0, 1\}$ (*decision functions*).

Let a *protocol* be a pair of functions $s(t), r(t)$ from the natural numbers (≥ 0) to the set {1,...,n} . t should be interpreted here as time and $s(t)$, $r(t)$ are, respectively, the sender and the recipient at time t. So $s(t) = i$, $r(t) = j$ means that at time t, i communicated with j (i sent a message which was received by j).

The protocol is *fair* iff every participant receives information from every other participant infinitely many times, possibly indirectly.

The simplest example of a fair protocol is a "round-robin" protocol, where the first person sends a value of f to the second person, second to third,...., and so on, until it reaches the last, nth person, who sends his value of f back to the first person. And this cycle is repeated forever. Formally, if participants are labeled from 0 to $n-1$, then this protocol can be expressed as $s(0) = r(0) = 0$ and for $t \geq 1$, $s(t) = r(t-1) = t \bmod n$.

Let $p = (s(t), r(t))$ be a protocol. We define by induction on t, the set of possible worlds for i at time t, given that the real world is x $C(x,i,t)$.

$$C(x,i,0) = P_i(x)$$

$$C(x,i,t+1) = \begin{cases} C(x,i,t) \cap f(C(x,s(t),t)) & \text{iff } i = r(t) \\ C(x,i,t) & \text{otherwise} \end{cases}$$

According to this definition, if i receives a message at t, then in order to obtain his set of possible worlds at $t+1$, he excludes from his set of possible worlds at t all the worlds incompatible with the received message.

We assume here for simplicity that communications are instantaneous and that always $s(t+1) = r(t)$. However both these assumptions are not necessary.

4 OVERVIEW OF THE USEFUL RESULTS

All the facts marked (P-K) are from [PK].

Fact 1 (P-K) : If all P_i are finite, then there is T s.t.

$$\forall x, i \, \forall t, t' \geq T \; C(x,i,t) = C(x,i,t')$$

so for every i we have some limiting partition of W (moreover this partition is finite). □

We use $C(x,i,\infty)$ to denote the limiting sets $C(x,i,T)$. Value $f(C(x,i,\infty))$ we call a *limiting value* of f for i at x. P^i, the set of limiting values for i is the set $\{f(C(x,i,\infty)) | x \in W\}$.

Let's define (following Bacharach) the *sure-thing principle* (Cave called the same condition *union consistency*):

Definition: Function $f : cl(W, P^+) \rightarrow D$ satisfies the *sure-thing principle* iff for every pair of disjoint sets $A, B \subseteq W$:

$$f(A) = f(B) \Rightarrow f(A \cup B) = f(A)$$

Fact 2 [Geanakoplos-Polemarchakis, Cave, Bacharach]: If two agents communicate values of a function satisfying the sure-thing principle, then they must reach consensus on the value of this function. □

Fact 3 (P-K) : If f is a decision function satisfying the sure-thing principle and the protocol P is fair then if 3 participants communicate values of f according to P then consensus on the value of f must be reached. □

Fact 4 (P-K) : If there is a space W, finite partitions $P_1,...,P_n$ of W and a decision function f s.t. for some fair protocol Pr, the sets of possible limiting values are $P^1,...,P^n$, then there exist finite partitions $P'_1,\ldots, P'_n$ of W and protocol Pr' with the same graph as Pr such that executing Pr', we get the *same* set of limiting values, but no one gains any knowledge during the execution of Pr'. I.e. $C(x,i,1) = C(x,i,\infty)$ for all i,x. □

In particular, if no consensus was reached in the first case, then in the second case there will be no consensus, and moreover, no learning will take place.

Fact 5 (P-K) : There is a decision function f satisfying the sure-thing principle s.t. values of f are communicated among 4 agents according to some fair protocol and no consensus is reached. □

5 MAIN RESULT

We prove that in a case when a decision based on the conditional probability of a fixed event is communicated in a fair protocol between n agents, consensus is guaranteed. We state it as the following theorem:

Theorem 1 : If α is a rational number, then the decision function d_E

$$d_E(A) = \begin{cases} 1 & \text{iff } p(E|A) > \alpha \\ 0 & \text{otherwise} \end{cases}$$

based on conditional probability of E brings consensus when communicated according to a fair protocol.

Proof: First we prove that when W is finite and every point has some weight assigned to it, then if values of the decision function based on weighed average are communicated in a fair protocol, then consensus is reached. We state it as the following:

Main Lemma : If w is some weight distribution on finite W ($w : W \to R^+$) and if a decision function f defined by:

$$f(A) = \begin{cases} 1 & \text{iff } [\Sigma_{a_i \in A} w(a_i)]/|A| > \alpha \\ 0 & \text{otherwise} \end{cases}$$

is communicated in a fair protocol, then consensus is reached (for any fixed rational α).

Proof: Let's suppose that consensus is not reached. When we shift w, to get $w' : W \to R^+$, $w'(A) = w(A) - \alpha$, we can express f as:

$$f(A) = \begin{cases} 1 & \text{iff } \Sigma_{a \in A} w'(a) > 0 \\ 0 & \text{iff } \Sigma_{a \in A} w'(a) < 0 \end{cases}$$

If consensus is not reached in some fair protocol when values of f are communicated then there must be some *unstable* worlds, worlds x for which there are two agents i, j s.t. in a stabilised situation (when nothing more can be learned) i sends 0 and j sends 1.

Let's call the set of unstable worlds U. U is not empty. We can express U in two ways: $U = U_0 = U_1$ where

$$U_0 = \bigcup_{i,x} \{C(i,x,\infty) | f(C(i,x,\infty)) = 0,\ f(C(i \ominus 1, x, \infty)) = 1\}$$

and

$$U_1 = \bigcup_{j,x'} \{C(j,x',\infty) | f(C(j,x',\infty)) = 1,\ f(C(j \ominus 1, x', \infty)) = 0\}$$

($\ominus$ is a subtraction mod n).

For U_0 we have a sequence of inequalities of the form

$$\sum_{a \in C(i,x,\infty)} w'(a) < 0$$

Adding up this sequence we'll get inequality

$$\beta_1 = \sum_{a \in U_0} k_a \cdot w'(a) < 0$$

where k_a is the number of sets $C(i,x,\infty) \subseteq U_0$ s.t. $a \in C(i,x,\infty)$. Similarly by looking at U_1 we'll get

$$\beta_2 = \sum_{a \in U_1} l_a \cdot w'(a) > 0$$

where l_a is the number of sets $C(j,x',\infty) \subseteq U_1$ s.t. $a \in C(j,x',\infty)$.

If during the execution of one round of the protocol for a certain world x value communicated changed l_x times from 0 to 1, then value communicated must have changed the same number of times from 1 to 0. So $k_a = l_a$ for every a. Thus $\beta_1 = \beta_2$ and we get a contradiction. □

Now we return to the proof of the theorem:

Let $|P^+| = n$. Let's order elements of P^+: $X_1, \dots, X_n$. Let $W' = \{x_1, \dots, x_n\}$. There is a mapping $m : P^+ \to W'$, $m(X_i) = x_i$. We can extend m to all the closed subsets of W in a natural way: $m(X) = \bigcup\{m(X_i) | X_i \in P^+, X_i \subseteq X\}$.

If values of the function f are communicated according to the same protocol in W with partitions P_i as values of f' in W' with partitions P_i' (where $P_i'^k = \{x | X_l \subseteq P_i^k, m(X_l) = x\}$ and $f'(m(X)) = f(X)$), then consensus on f' is reached iff consensus on f is reached. So without loss of generality we can assume that our space W is finite.

If all the worlds are equally likely, then if we take $p(E \cap \{x\})$ to be the weight of x, then $p(E|A) > \alpha|W|$ iff an average weight of A is greater than α, so our theorem follows directly from the main lemma.

If weights of the points are different, we can create another set W'', and a mapping $\varphi : W'' \to W$, s.t. weights of all points are the same in W'', and for all sets in P''^+ (defined as $P_i''^k = \{x | \varphi(x) \in P_i^k\}$), $p(E|A) = p(\varphi(E)|\varphi(A))$ (we extend φ to a mapping $cl(W'', P''^+) \to cl(W, P^+)$ in the obvious way). So if there is no consensus in general case, there would be no consensus in a case where all the points are equally likely. This ends the proof of the theorem. □

Remark 1: The decision functions based on conditional probability satisfy the sure-thing principle. Unfortunately we cannot use it to get a contradiction by looking at the sets U_0 and U_1 since these sets represented as unions of sets of the form $C(i,x,\infty)$ are not necessarily disjoint unions. □

6 NECESSARY CONDITIONS FOR CONSENSUS

We proved that some decision functions always bring consensus among agents communicating their values according to fair protocols. We would like to find what conditions functions must satisfy so we could ever hope for consensus on their value.

Definition : Condition Φ is a *necessary condition* for guaranteeing consensus on the value of $f : cl(W, P^+) \to \{0, 1\}$ in a fair protocol iff for every f which doesn't satisfy Φ there are some partitions $P_i' \subseteq P^+$ s.t. for some $x \in W$ and for some fair protocol there is no consensus on the value of f in x in this protocol.

The idea here is that if our necessary condition is not satisfied, then without increasing the total amount of information in the system (join of new partitions is no finer than old P^+) we can change the *distribution* of information among agents so that in some world there will be no consensus.

Notice that the sure-thing principle is a necessary condition for guaranteeing consensus among any number of agents.

We will try to find some stronger necessary conditions for the case of more then two agents communicating.

Definition : Let's define a *domination relation* ($\succ_f$) generated by f on $cl(W, P^+)$ as $A \succ_f B$ iff A,B disjoint, f(A)≠f(B) and $f(A \cup B) = f(A)$

Definition: The relation $\succ_f$ is *cyclic* iff there is a sequence of sets of worlds $A_1, ..., A_k$ s.t.

$$A_1 \succ_f A_2 \succ_f ... \succ_f A_k \succ_f A_1$$

This sequence we call a *cycle* of length k.

If $\succ_f$ is not cyclic we call it *acyclic*.

Note that the fact that $\succ_f$ is acyclic is equivalent to the fact that $< cl(W, P^+), \succ_f >$ is embeddable into $< R, > >$.

Fact 6 : All the cycles in $\succ_f$ are of even length. □

Definition : A set of sets of worlds $\mathcal{A} \subseteq cl(W, P^+)$ has *pairwise disjoint dominators with respect to* f iff

$$\forall A_1, A_2 \in \mathcal{A} \; f(A_1 - A_2) = f(A_1)$$

Intuitively, we don't want two sets in our collection to share some "strong" (dominating) subset of worlds. The other way of expressing this condition is to say that

$$\forall A_1, A_2 \in \mathcal{A} \; A_1 \cap A_2 \not\succ_f A_1 - A_2$$

Theorem 2: If $\succ_f$ is cyclic and there is a cycle $A_1 \succ_f A_2 \succ_f ... \succ_f A_n \succ_f A_1$ which has pairwise disjoint dominators with respect to f, then there exist partitions $P_1, ..., P_n$ of W s.t., when values of f are communicated in any world $x \in \bigcup A_i$ in an n-person round-robin protocol then no consensus is reached.

Proof: Suppose that we have a cycle in $\succ_f$ of length $n = 2m$ (see fact 6). $A_1 \succ_f A_2 \succ_f ... \succ_f A_n$, and $f(A_i - A_j) = f(A_i)$ for $i \neq j$. Let's assume that $f(A_{2k}) = 0$ and $f(A_{2k-1}) = 1$ for $k = 1, ..., n/2$. Now we construct partitions $P_1, ..., P_n$. In P_i we define A_j' in the following way: $A_j' = A_j - (A_i \cup A_{i+1})$:
$P_1 = \{A_1 \cup A_2, A_3', ..., A_n', W - \bigcup A_i\}$

$P_2 = \{A_2 \cup A_3, A'_1, A'_4, ..., A'_n, W - \bigcup A_i\}$
$P_3 = \{A_3 \cup A_4, A'_1, A'_2, A'_5, ..., A'_n, W - \bigcup A_i\}$
...
$P_n = \{A_n \cup A_1, A'_2, A'_3, ..., A'_{n-1}, W - \bigcup A_i\}$
In any world $x \in A_i$, all the people except the person $i-1$ send $f(A_i)$ while the person $i-1$ sends the opposite value. □

Corollary 1 : : If $\succ_f$ is cyclic, there is a cycle in $\succ_f$ built up of pairwise disjoint sets, then there exist partitions $P_1, ..., P_n$ of W s.t., when values of f are communicated in any world x in any of the sets forming the cycle in an n-person round-robin protocol then no consensus is reached. □

The question is whether cyclicity of $\succ_f$ is a sufficient condition for existence of partitions that might lead to a situation with no consensus reached. As we see on the following example, the answer to that is negative, therefore the additional condition (sets in a cycle have pairwise disjoint dominators), turns out to be essential.

Example 1 : This is a decision function with cyclic domination relation s.t. in every world, in every partition consensus on the value of f must be reached in every fair protocol.
$W = \{a, b, c, d, e\}$
$f(a) = f(b) = f(c) = f(d) = 1$
$f(e) = 0$ and $\forall i \in \{a, b, c, d\}$ $f(\{e, i\}) = 0$
$f(\{b, d, e\}) = f(\{a, c, e\}) = 0$
for all other sets A, $f(A) = 1$.

Clearly $\succ_f$ is cyclic, since

$$\{a\} \succ_f \{b, e\} \succ_f \{d\} \succ_f \{c, e\} \succ_f \{a\}$$

f has the following property: for any $B \subseteq A$, $\{e\} \succ_f A \rightarrow \{e\} \succ_f B$ (this is because e is the only element on which f gives value 0).

We prove that f guarantees consensus; suppose otherwise, then there are some agents i and $i+1$ s.t. i sends 0 and $i+1$ sends 1 in x. If i sends 0 in x, then $e \in C(i, x, \infty)$. Nothing is learned, so

$$C(i+1, x, \infty) \subseteq \bigcup\{C(i, x, \infty) | f(C(i, x, \infty) = 0\} = C(i, x, \infty)$$

But then $e \in C(i+1, x, \infty)$, and by the property of f mentioned earlier, $f(C(i+1, x, \infty)) = 0$, a contradiction. □

Remark 2: Not every decision function with an acyclic domination relation can be expressed as a decision based on weighed average. For example, if we have

$$\{a\} \succ_f \{b\} \text{ and } \{a\} \succ_f \{c\} \text{ but } \{b, c\} \succ_f \{a\}$$

then we cannot assign weights to a, b, c so that f would be based on averages of weights. □

7 SUMMARY

We proved that for any number of participants there is always consensus on the decision based on probabilities of some fixed event, provided that these decisions are communicated in some fair protocol.

We are not able at that point to establish complexity (how long it takes to reach consensus) as a function of partition sizes. Complexity here would have to be expressed in terms of *rounds* where one round is a sequence of communications s.t. everyone has a chance to contribute to everyone else's information (so protocols as: "exchange information between $n-1$ agents (in a circle) every time unit, and consult the remaining agent every 2^t time units (t is time used so far)" would not give us automatically high complexity).

Geanakoplos and Polemarchakis showed that in case of two agents communicating the complexity is linear. Their proof was based on the observation that if an agent removes a world from his set of possible worlds, he must remove the whole equivalence class (element of the other agents' partition), and this fact is common knowledge between the agents.

In the case of n agents it is no longer true. It seems that an agent may remove an arbitrary closed set. If all the partitions have n elements, there is 2^n closed sets, and we don't know if consensus is always guaranteed in polynomial number of rounds.

There are other natural examples of decision functions, which when communicated in a fair protocol must create consensus among participants. One such example is a decision based on maximum weight (in a case of finite W). We state it as the following fact:

Fact 7 : Every function $f: W \to \{0,1\}$ (where W - finite) of the form:

$$f(A) = \begin{cases} 1 & \text{iff } max(\{f(a)|a \in A\}) > \alpha \\ 0 & \text{otherwise} \end{cases}$$

for some fixed α, brings consensus in a fair protocol (the same is true for f defined as based on minimum). □

We introduced necessary condition decision functions must satisfy so that consensus is possible.

It remains open to find complete characterisation of functions which always create consensus in fair protocols.

Acknowledgements

I would like to thank prof. Rohit Parikh who introduced me to the consensus problems. Prof. Parikh who was my Ph.D. adviser gave me enormous help and encouragement while I was working on my Ph.D. thesis, which contains the results from this paper. Also, I would like to thank prof. Melvin Fitting for his valuable suggestions.

References

[Aum76] R. Aumann. Agreeing to disagree. *Annals of Statistics*, 4:1236–1239, 1976.

[Bac85] M. Bacharach. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory*, 37:167–190, 1985.

[Cav83] J. Cave. Learning to agree. *Economics Letters*, 12:147–152, 1983.

[GP82] J. Geanakoplos and H. Polemarchakis. We can't disagree forever. *Journal of Economic Theory*, 28:192–200, 1982.

[PK] R. Parikh and P. Krasucki. Communication, consensus and knowledge. *to appear in Journal of Economic Theory.*