ELSEVIER

Economics Letters 77 (2002) 365–369

economics letters

www.elsevier.com/locate/econbase

Disagreement is unpredictable

Robin Hanson*

Department of Economics, George Mason University, Carow Hall, MSN 1D3, 22030 Fairfax, VA, USA

Received 8 April 2001; received in revised form 12 February 2002; accepted 16 May 2002

Abstract

Given common priors, no agent can publicly estimate a non-zero sign for the difference between his estimate and another agent's future estimate. Thus rational agents cannot publicly anticipate the direction in which other agents will disagree with them.

Keywords: Consensus; Common knowledge; Agreeing to disagree

JEL classification: D82

1. Introduction

The 'agreeing to disagree' literature typically has two kinds of results: consensus and convergence.

Consensus results show how something like common knowledge of a property of agents' beliefs implies stronger agreement. For example, agents must have identical estimates of a real-valued random variable when they have common knowledge of their exact estimates (Aumann, 1976), of a separating point (Sebenius and Geanakoplos, 1983), of which agent has the highest estimate (Hanson, 1998), or of a sum of strictly monotonic functions of agent expectations (McKelvey and Page, 1986; Neilsen et al., 1990). Replacing common knowledge with common belief gives similar results (Monderer and Samet, 1989; Neeman, 1996; Sonsino, 1995).

Convergence results show how the commonality of belief required for these consensus results can arise from repeated information exchange. For example, announcing a property of agents' beliefs is typically assumed to produce common knowledge of that fact. If announcing informs agents, however, then this common knowledge is of what the property was, and not of what it became after the announcement. However, in a finite world with fixed private information, an infinite sequence of

*Tel.: +1-703-993-2326; fax: +1-703-993-2323.
E-mail addresses: rhanson@gmu.edu (R. Hanson), http://hanson.gmu.edu (R. Hanson).

0165-1765/02/$ – see front matter
PII: S0165-1765(02)00149-0

announcements eventually reaches a point where no one learns anything new. At that point the common knowledge produced by the announcement does apply after the announcement, and so in this case, repeated announcements can eventually produce common knowledge of a property of beliefs (Geanakoplos and Polemarchakis, 1982; Sebenius and Geanakoplos, 1983; McKelvey and Page, 1986; Neilsen et al., 1990; Hanson, 1998). Related results can apply in non-finite worlds (Nielsen, 1984).

The empirical relevance of these convergence results is unclear. Some belief properties, when announced, typically produce convergence in a single step, but can require enormous amounts of information in that single step (McKelvey and Page, 1986; Neilsen et al., 1990). Other properties require only a few bits of information per announcement, but may then require an enormous number of announcements (Hanson, 1998). Worse, if agents acquire new private information between announcements, all of these processes can fail to converge.

This letter, in contrast, describes an agreeing to disagree type result where convergence is certain after one announcement of zero bits of information about one agent. If we define a disagreement to be the difference between two agent's estimates of a random variable, then the consensus result of this letter is that common knowledge of the sign of one agent's estimate of a disagreement implies that estimate is zero. The convergence result of this letter is that common knowledge of this sign results from a single announcement of this sign, even when agents concurrently obtain other information. And since the sign must be zero, the announcement's content is completely anticipated, and so requires zero expected bits.

A natural interpretation of this result is that rational agents cannot publicly anticipate the direction in which other agents will disagree with them. This is another 'no agreeing to disagree' type result, except that instead of describing the mythical end of a conversation where opinions never change, it applies at any point during the conversation. It thus seems more applicable to realistic situations, and more susceptible to empirical tests (though such testing still seems far from easy).

2. Analysis

Let Ω be a finite set of states of the world ω. Let two agents, named one and two, have common non-degenerate prior beliefs $p_\omega > 0$, and receive private information according to partitions of Ω. Call agent one's partition I, and let $I(\omega)$ denote the element of I containing ω. Let agent two's partition similarly be J.

If agent one is Bayesian, then given any real-valued random variable V_ω, if the true state is ω^* he will have a conditional expectation of V given by

$$X_{\omega^*} = \mathrm{E}[V|I(\omega^*)] = \frac{\sum_{\omega \in I(\omega^*)} p_\omega V_\omega}{\sum_{\omega \in I(\omega^*)} p_\omega}. \tag{1}$$

Bayesian agent two will similarly have an expectation $Y_\omega = \mathrm{E}[V|J(\omega)]$. Agent one's expectation of agent two's expectation is then $Z_\omega = \mathrm{E}[Y|I(\omega)]$. Thus $Z_\omega - X_\omega$ is agent one's estimate of how agent two disagrees with agent one in estimating V_ω.

For any true state ω, we say it is *common knowledge* among agents one and two that the true state is in $(I \wedge J)(\omega)$, where $I \wedge J$ is the meet (or finest common coarsening) of the partitions I and J. We call an event $E \subset \Omega$ common knowledge at ω if $(I \wedge J)(\omega) \subset E$, and call a random variable V common knowledge if the partition made of its variable value events $E_v = \{\omega \in \Omega | V_\omega = v\}$ is a coarsening of the common knowledge partition $I \wedge J$.

We can now prove that if it is common knowledge that one agent's expectation of another agent's expectation is no less than his own expectation, then these two expectations of the first agent are equal.

Theorem 1. *If it is common knowledge at* ω *that* $Z_\omega \geq X_\omega$ *(or that* $Z_\omega \leq X_\omega$*), then* $Z_\omega = X_\omega$*, and this fact is common knowledge.*

Proof. Since $X_\omega = X_{\omega*}$ for all $\omega \in I(\omega*)$, we can rearrange Eq. (1) as

$$\sum_{\omega \in I(\omega*)} X_\omega p_\omega = \sum_{\omega \in I(\omega*)} V_\omega p_\omega.$$

Summing this over the partition elements $I(\omega*) \subset (I \wedge J)(\omega*)$, we get

$$\sum_{\omega \in (I \wedge J)(\omega*)} X_\omega p_\omega = \sum_{\omega \in (I \wedge J)(\omega*)} V_\omega p_\omega$$

The equations similar to Eq. (1) for Y and Z, treated similarly, give

$$\sum_{\omega \in (I \wedge J)(\omega*)} Y_\omega p_\omega = \sum_{\omega \in (I \wedge J)(\omega*)} V_\omega p_\omega,$$

$$\sum_{\omega \in (I \wedge J)(\omega*)} Z_\omega p_\omega = \sum_{\omega \in (I \wedge J)(\omega*)} Y_\omega p_\omega.$$

We see that these last three equations are all equal to each other, and so

$$\sum_{\omega \in (I \wedge J)(\omega*)} (Z_\omega - X_\omega) p_\omega = 0 \tag{2}$$

If it is common knowledge at $\omega*$ that $Z_{\omega*} \geq X_{\omega*}$, then this is true for all $\omega \in (I \wedge J)(\omega*)$, and so Eq. (2) becomes a sum of non-negative terms set equal to zero, which can only be if each term in the sum is zero. Thus for all $\omega \in (I \wedge J)(\omega*)$ we have $Z_\omega = X_\omega$, and so this equality is common knowledge. The result for $Z_\omega \leq X_\omega$ follows by applying the just proved result to the variable $-V$. □

Agent information partitions can change with time. If at time t a rational agent has information K^t, and then obtains new information described by a partition A^t, his time $t+1$ information should be the join (or coarsest common refinement) of these partitions $K^{t+1} = K^t \vee A^t$. This implies $K^{t+1}(\omega) \subset A^t(\omega)$. Such a changing agent can be thought of as a sequence of *agent selves* with differing information partitions K^t. We can thus speak of common knowledge between two agent selves I^s and J^t, even when $s \neq t$. That is, we can say that it is common knowledge between those selves that the true state is in $(I^s \wedge J^t)(\omega)$.

Let us say that agent self t, with information K^t, has been *reliably informed* about the information in partition A if $K^t(\omega) \subset A(\omega)$ for all ω, and say this self is reliably informed about a random variable if he is reliably informed about that variable's value partition. When two agent selves are reliably informed of a variable V, then V is common knowledge between them, since V's value partition is then a coarsening of both agent partitions.

Let us apply Theorem 1 to two particular agent selves, agent one self s with information I^s and agent two self t with information J^t, for $s<t$. Thus $X_\omega = \mathrm{E}[V|I^s(\omega)]$, $Y_\omega = \mathrm{E}[V|J^t(\omega)]$, and $Z_\omega = \mathrm{E}[Y|I^s(\omega)]$. Let us define the variable *positivity* P of $Z-X$ to be

$$P_\omega = 1 \text{ if } Z_\omega - X_\omega > 0, \text{ and 0 otherwise.}$$

Let the variable *negativity* N of $Z-X$ be defined similarly, replacing $>$ with $<$.

We can now show that no agent can tell any other agent about the direction of their disagreement. That is, no agent can reliably inform a second agent of a non-zero sign of his estimate of the direction in which, at a particular future time, that second agent's opinion will differ from his own current opinion.

Theorem 2. *If agent two self t is reliably informed of P (or N), then* $Z_\omega = X_\omega$ *(and* $P=N=0$*), and all this is common knowledge.*

Proof. Variable P is defined in terms of Z and X, both of which are constant across each $I(\omega)$. Thus P must be constant as well, making agent one self s reliably informed about P. Since by assumption agent two self t is also reliably informed about P, variable P must be common knowledge. When $P_\omega = 1$, it is thus common knowledge that $Z_\omega > X_\omega$, which implies $Z_\omega \geq X_\omega$, and so by Theorem 1, we have $Z_\omega = X_\omega$, which is a contradiction. When $P_\omega = 0$, it is common knowledge that $Z_\omega \leq X_\omega$, which by Theorem 1 implies $Z_\omega = X_\omega$, and common knowledge of this fact, which in this case is not a contradiction. Announcing instead the negativity N of $Z-X$ only changes which of these two cases is a contradiction. □

Note that the result of Theorem 2 is compatible with either agent acquiring new private information as P or N is announced. Note also that since agent two knows with certainty that he will hear $P=0$ or $N=0$, the expected number of bits required to communicate this information to him is zero. The actual telling has no effect; it seems to be the commitment of agent one to tell the sign of the disagreement that eliminates the disagreement.

3. Conclusion

This letter shows that, given common priors, no agent can ever tell another agent the direction in which that other agent will, at some future time, disagree with his current opinion.

Most agreeing to disagree type results either assume common knowledge or belief, or show that such commonality can arise from an indefinite sequence of announcements. The arrival of further private information can prevent such convergence, however, and we know of no bounds on the information that must be communicated in such processes. These features make these results harder to test empirically.

In contrast, this new agreeing to disagree type result requires only a single announcement of zero expected bits, and is robust to the concurrent arrival of new private information. These features of this result should make it easier to test empirically, although far from easy, as there still remain many other challenges to such testing.

Acknowledgements

For their comments, I thank Ron Heiner and an anonymous referee.

References

Aumann, R., 1976. Agreeing to disagree. The Annals of Statistics 4 (6), 1236–1239.

Geanakoplos, J., Polemarchakis, H., 1982. We can't disagree forever. Journal of Economic Theory 28, 192–200.

Hanson, R., 1998. Consensus by identifying extremists. Theory and Decision 44 (3), 293–301.

McKelvey, R., Page, T., 1986. Common knowledge, consensus and aggregate information. Econometrica 54 (1), 109–127.

Monderer, D., Samet, D., 1989. Approximating common knowledge with common beliefs. Games and Economic Behavior 1, 170–190.

Neeman, Z., 1996. Approximating agreeing to disagree results with common p-beliefs. Games and Economic Behavior 12, 162–164.

Neilsen, L.T., Brandenburger, A., Geanakoplos, J., McKelvey, R., Page, T., 1990. Common knowledge of an aggregate of expectations. Econometrica 58 (5), 1235–1239.

Nielsen, L.T., 1984. Common knowledge, communication, and convergence of beliefs. Mathematical Social Sciences 8, 1–14.

Sebenius, J., Geanakoplos, J., 1983. Don't bet on it: contingent agreements with asymmetric information. Journal of the American Statistical Association 78 (382), 424–426.

Sonsino, D., 1995. 'Impossibility of speculation' theorems with noisy information. Games and Economic Behavior 8, 406–423.