The Communication Cost of Agreement

Scott Aaronson

February 3, 2004

What I want to talk about today has nothing whatsoever to do with quantum computing (pause for gasps). It was inspired by visiting the home page of this economist named Robin Hanson (`hanson.gmu.edu`). I don't agree with everything he says—which, as you'll see, is an extremely puzzling fact—but I like that he asks big questions. For example, he has a paper about how, even if you assign a small subjective probability to your living in a computer simulation, this should affect your behavior in interesting ways. The question I liked the most was, why do people disagree? Pick any two smart people, and there will be some topic they disagree on—the Israeli-Palestinian conflict, capitalism vs. socialism, the interpretation of quantum mechanics (oops! I said quantum), etc. The smarter they are, the easier it will be to find such a topic. And if they discuss it, chances are they won't reach consensus, but will just become more confirmed in their previous beliefs. Why?

Obviously, part of the explanation is that people are irrational, they're dishonest, they're self-serving, they only look for evidence that confirms what they want to believe, etc. Apparently 94% of professors think they're better than their average colleagues (which is logically possible, but I assume you'd get a similar response if you asked if they were better than their *median* colleagues). The question is, are these things the *whole* explanation, or is there also some reason why even honest, rational, trusting people would agree to disagree? You might say, well, they have different axioms. But were that the whole explanation, you'd expect people to say to each other, "I don't accept your axioms, but supposing I did, I agree that such-and-such would follow." And in my experience, people don't even agree to *that* extent in anything besides math.

In 1976, Aumann (a famous economist) proved the following simple theorem. Suppose Luca and Satish are honest, trusting Bayesians, and suppose they have the same prior probabilities for everything, but (of course) different life experiences. And suppose that conditioned on everything he knows, Luca assigns some posterior probability to theory lunch being cancelled next week. Satish also assigns some posterior probability. Then provided Luca and Satish are aware of each other's probabilities, those probabilities must be equal. This is so even if Luca has no idea on what sort of evidence Satish bases his opinion—since the sort of evidence can itself be considered a random variable, which is ultimately governed by this prior probability distribution that's the same for Luca and Satish. You might say, great, so they just have to announce their probabilities and then they'll agree. But suppose Luca says, I think there's a 5% probability and Satish says, I think there's a 10% probability. Then Luca might say, well, *given* that Satish thinks there's a 10% probability, actually I think there's an 8% probability, and Satish might also revise his opinion. But it's easy to see that, if they just keep on announcing their current posteriors, then provided the probability space is finite, the process will terminate with Luca and Satish having the same probability.

My reaction to this result as a complexity theorist was *duh.* I mean, if Luca and Satish exchanged everything they knew, then of course they'd agree! So the crucial question, which wasn't even touched on in the large economics literature on this subject, was how many rounds? And how many bits do they need to communicate? I conjectured on Lance Fortnow's weblog that one could prove a lower bound saying that in some cases, even to agree approximately they'd have to exchange a number of bits nearly equal to the total number of bits they know. That would then provide a reason why even honest, rational, trusting people could disagree: communication complexity!

My result is that this is not the case. More formally: let f be a function from $\{0,1\}^n \times \{0,1\}^n$ to $[0,1]$, and let $\mathcal{D}$ be a prior distribution over $(x,y) \in \{0,1\}^n \times \{0,1\}^n$ known to both Alice and Bob. As in standard communication complexity, Alice gets x, Bob gets y, and they exchange a sequence of messages. Then with probability at least $1-\delta$ over x, y, Alice and Bob can make their expectations of $f(x,y)$ agree to within

ε by exchanging only $O\left(1/\left(\delta\varepsilon^2\right)\right)$ bits of communication, independently of n. (I'll sometimes ignore the dependence on δ.) I conjecture this is tight, but so far I've proved that only for a "standard protocol" in which Alice and Bob just keep describing their current expectation of f. I also conjecture that the standard protocol is always the best protocol, but so far haven't proved that either.

Three remarks: first, if Alice and Bob want their expectations to agree *exactly*, then it's obvious that in the worst case they need $2n$ bits of communication: n from Alice and n from Bob. (Note the contrast with ordinary communication complexity, where n bits always suffice.) Second, the qualifier "with probability at least $1-\delta$ over x, y" is significant; it's easy to construct examples such that the standard protocol needs $\Omega(n)$ steps to produce even approximate agreement for some *particular* x, y pair. Third, my termination condition is that Alice and Bob should be able to end their conversation at a previously agreed-upon *random* time—or equivalently, they should agree to within ε for at least a $1-\delta$ fraction of their conversation. This condition lets me ignore awkward situations like the following:

> ALICE: Great news—I agree to within ε with the opinion you just announced! We're done!
>
> BOB: Wait a minute...*conditioned* on your wanting to end the conversation, I no longer agree with you!

However, the $\Omega\left(1/\varepsilon^2\right)$ lower bound will apply even if a wizard slaughters Alice and Bob the moment they agree to within ε (so quickly that they don't have time to revise their opinions conditioned on the wizard's appearance).

I'll now prove the $O\left(1/\varepsilon^2\right)$ upper bound. For simplicity, first suppose Alice and Bob just take turns announcing their current expectations, however many bits that requires. A crucial observation made by Hanson is that throughout the protocol, Alice's expectation of Bob's expectation of f equals Bob's expectation of Alice's expectation. This is completely unlike real life—where maybe you know that I think I'm great, and I know that you think I suck—but it follows trivially since in both cases, we're averaging over *all* inputs of Alice and *all* inputs of Bob that are consistent with the sequence of messages exchanged so far. A corollary is that Alice can never predict the direction in which Bob will disagree with her: for if she's just sent a message to Bob, then her expectation of f equals Bob's expectation of her expectation, therefore her expectation equals her expectation of Bob's expectation.

Suppose a third party Eve doesn't know x or y, but she knows the prior distribution $\mathcal{D}$, and she gets to see the first t messages that Alice and Bob exchange. Let p_t be the expectation she'd then assign to $f(x, y)$. The intuition is that, so long as Alice and Bob disagree by more than ε with high probability, if we look at p_1, p_2, p_3, etc., these values follow a random walk with step size roughly ε. Furthermore, this walk has two absorbing barriers at 0 and 1—since if $p_t \approx 0$ or $p_t \approx 1$, then Alice and Bob *must* probably approximately agree. And we expect a random walk with step size ε to hit a barrier after $O\left(1/\varepsilon^2\right)$ steps.

To make this intuition rigorous, all we need to do is look at the expectation, not of p_t, but of p_t^2. Let $p_{t+1} = p_t + \Delta_t$. Then

$$\mathrm{EX}\left[\left(p_t + \Delta_t\right)^2 |\mathcal{M}_t\right] = p_t^2 + 2p_t \,\mathrm{EX}\left[\Delta_t|\mathcal{M}_t\right] + \mathrm{EX}\left[\Delta_t^2|\mathcal{M}_t\right]$$

where $\mathcal{M}_t$ consists of the first t messages. We've already argued that $\mathrm{EX}\left[\Delta_t|\mathcal{M}_t\right] = 0$, so

$$\mathrm{EX}\left[p_{t+1}^2|\mathcal{M}_t\right] = p_t^2 + \mathrm{EX}\left[\Delta_t^2|\mathcal{M}_t\right].$$

Now suppose without loss of generality that Bob sent the t^{th} message; then $\mathrm{EX}\left[\Delta_t^2|\mathcal{M}_t\right]$ is just the variance of Alice's opinion from Eve's point of view. Let a_t be Alice's expectation of f after t messages and let b_t be Bob's expectation; then $b_t = p_t$. So if $|a_t - b_t| > \varepsilon$ with probability γ (again, conditioned on $\mathcal{M}_t$), then $|a_t - p_t| > \varepsilon$ with probability γ, which implies that the variance $\mathrm{EX}\left[\Delta_t^2|\mathcal{M}_t\right]$ is more than $\gamma\varepsilon^2$. Let δ_t be the marginal probability that $|a_t - b_t| > \varepsilon$. By linearity of expectation, if $\delta_1 + \cdots + \delta_T > 1/\varepsilon^2$, then

$$\mathrm{EX}\left[p_T^2\right] \geq \delta_1\varepsilon^2 + \cdots + \delta_T\varepsilon^2 > 1,$$

which is absurd. So if Alice and Bob set $T = 1/\left(\delta\varepsilon^2\right)$, and end their conversation after t steps for t chosen uniformly at random from $\{1, \ldots, T\}$, then

$$\Pr\left[|a_t - p_t| > \varepsilon\right] \leq \frac{1}{T\varepsilon^2} = \delta.$$

To discretize the protocol, we can have Alice send a '1' bit if $a_t \geq p_t$ and a '0' bit otherwise, and *simultaneously* have Bob send a '1' if $b_t \geq p_t$ and a '0' otherwise. (Even if the model doesn't allow simultaneous messages, it's easy to simulate their effect—Bob simply "forgets" Alice's most recent message before sending his own.) It's still the case that $\mathrm{EX}\left[\Delta_t | \mathcal{M}_t\right] = 0$; all that changes in the above analysis is that Eve can no longer infer Alice's and Bob's expectations *exactly* from their most recent messages. All that's needed, however, is that if $|a_t - b_t| > \varepsilon$ with probability at least δ, then $\mathrm{EX}\left[\Delta_t^2 | \mathcal{M}_t\right] = \Omega\left(\delta\varepsilon^2\right)$. And it's clear that this is still the case.

OK—the lower bound. For "ordinary" functions, $O\left(\log 1/\varepsilon\right)$ steps are usually enough—but I can custom-design a bizarre function for which the $1/\varepsilon^2$ random walk behavior actually occurs. Specifically, let $\mathcal{D}$ be the uniform distribution over $x = x_1 \ldots x_n$ and $y = y_1 \ldots y_n$ in $\{-1, 1\}^n$, and let $n \geq 1/\varepsilon^2$. Then I can show that the discretized simultaneous-message protocol needs $\Omega\left(1/\varepsilon^2\right)$ steps for the following function:

$$f\left(x, y\right) = \begin{cases} w\left(x, y\right) & \text{if } w\left(x, y\right) \in [0, 1] \\ 0 & \text{if } w\left(x, y\right) < 0 \\ 1 & \text{if } w\left(x, y\right) > 1 \end{cases}$$

where

$$w\left(x, y\right) = \frac{1}{2} + \varepsilon \sum_{i=1}^{n} x_{i-1} y_{i-1} \left(x_i + 2y_i\right)$$

and $x_0 = y_0 = 1$. To understand this function, it might help to look at the "communication matrix" of w in the $n = 1$ and $n = 2$ cases:

$x \setminus y$	-1	1
-1	$1/2 - 3\varepsilon$	$1/2 + \varepsilon$
1	$1/2 - \varepsilon$	$1/2 + 3\varepsilon$

$x \setminus y$	$-1, -1$	$-1, 1$	$1, -1$	$1, 1$
$-1, -1$	$1/2 - 6\varepsilon$	$1/2 - 2\varepsilon$	$1/2 + 4\varepsilon$	$1/2$
$-1, 1$	$1/2 - 4\varepsilon$	$1/2$	$1/2 + 2\varepsilon$	$1/2 - 2\varepsilon$
$1, -1$	$1/2 + 2\varepsilon$	$1/2 - 2\varepsilon$	$1/2$	$1/2 + 4\varepsilon$
$1, 1$	$1/2$	$1/2 - 4\varepsilon$	$1/2 + 2\varepsilon$	$1/2 + 6\varepsilon$

At the beginning, Alice's expectation a_0 is either $1/2 - \varepsilon$ or $1/2 + \varepsilon$, and Bob's expectation b_0 is either $1/2 - 2\varepsilon$ or $1/2 + 2\varepsilon$. Thus $|a_0 - b_0| \geq \varepsilon$. Moreover, w has a fractal-like property that ensures that $|a_t - b_t| \geq \varepsilon$ for all t. To see this: initially $a_0 = 1/2 + \varepsilon x_1$ and $b_0 = 1/2 + 2\varepsilon y_1$. Most of the terms in the sum defining $w\left(x, y\right)$ just average to 0 for both players, since Alice doesn't know the y_i's and Bob doesn't know the x_i's. In the first step, however, the bit Alice sends to Bob reveals x_1 to him, and the bit Bob sends to Alice reveals y_1 to her. These bits "unlock" the term $x_1 y_1 \left(x_2 + 2y_2\right)$, which now causes Alice's and Bob's expectations to differ by at least ε again. Then in the second step, Alice reveals x_2 and Bob reveals y_2, thereby unlocking the term $x_2 y_2 \left(x_3 + 2y_3\right)$, and so on. As this process continues, Eve's expectation p_t follows an unbiased random walk with starting point $1/2$ and step size ε. We expect such a walk to hit 0 or 1 only after $\Omega\left(1/\varepsilon^2\right)$ steps. There are extremely small corrections when we replace $w\left(x, y\right)$ by $f\left(x, y\right)$, but we can ignore those with the help of a Chernoff bound.

The above argument *should* still work for the variant of the discretized protocol in which Alice and Bob take turns instead of sending their messages simultaneously. However, for the continuous protocol I only get a lower bound of $\Omega\left(1/\varepsilon\right)$, not $\Omega\left(1/\varepsilon^2\right)$, since I can no longer replace $w\left(x, y\right)$ by $f\left(x, y\right)$ so cavalierly. I don't know whether this is a technical problem or something fundamental.

Anyway, the conclusion is that, whenever two people disagree about anything, there must be a reason besides their not having enough time to communicate with each other.