Aggregating Beliefs: Consensus Versus Compromise

Mark Colyvan

Sydney Centre for the Foundations of Science
University of Sydney
Australian Centre of Excellence for Risk Analysis

SRA Workshop: Philosophical Reflections on Risk Management and Complex Systems
29th September 2008

Outline

What is Consensus?

- Consensus is the situation where all agents in a multi-agent decision situation agree on the preferred action.
- More stringently, consensus is achieved when all agents agree on both the utility functions and the probability distributions (i.e. the agents can be represented as a single agent).
- We are mostly interested in consensus in relation to estimates of a particular parameter (e.g. criteria weights).

What is Compromise?

- Compromise is where the agents persist with their own choices of the preferred act (and these choices are not the same for all agents) so some compromise action is chosen.
- Similarly for utilities and probabilities.
- Similarly for compromise with respect to a given parameter estimate.

Why Consensus?

- With consensus everyone gets what they want; everyone goes home happy.
- With compromise there is no guarantee that anyone gets what they want; they might all go home unhappy.
- Consensus, in effect, delivers a single agent, so there's no need to worry about Arrow's theorem, discursive dilemmas and the like.
- Sometimes legislation requires consensus.

Assigning Weights to Conservation Priorities

- The problem here is that of determining a series of weights for various conservation priorities (purchasing, with a fixed budget, various land packages for wildlife reserves): *criteria weights*.
- Each committee member a_i assigns their initial series of (normalised) criteria weights p_i^0.
- Each committee member also provides a *weight of respect* for the other committee members, where a_i's respect for a_j is represented as w_{ij}.
- The central idea here is that each agent will modify their criteria weights based on their weights of respect (or lack thereof) for the other n committee members according to: $p_i^1 = w_{i1}p_1^0 + w_{i2}p_2^0 + \ldots + w_{in}p_n^0$.

A Bit of Linear Algebra

- We can make this all easier by compiling all the weights of respect for all the agents into a single n by n matrix and all agents' criteria weights into a single n-vector.
- The updating is then simply the matrix multiplication:

$$WP = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \begin{bmatrix} p_1^0 \\ p_2^0 \\ \vdots \\ p_n^0 \end{bmatrix}$$

A Bit More Linear Algebra

- Of course we can iterate this procedure to obtain the vector of criteria weights obtained after m iterations, $W^m P$. (The matrix of weights of respect are fixed, so this is well defined.)
- The crucial result is a theorem that ensures convergence of $W^m P$, as m approaches infinity.
- All we need to assume here is that the weights of respect have a few fairly plausible properties, the most notable of which is that they are non-trivial.

A Practical Problem and a Pragmatic Fix

- In practice, eliciting criteria weights is hard enough; eliciting weights of respect is near impossible.
- The modified Lehrer-Wagner model (modification due to Regan) measures weight of respect in terms of distance from and individual's criterion weights.
- That is,

$$w_{ij} = \frac{1 - |p_i^0 - p_j^0|}{\sum_{j=1}^{n} 1 - |p_i^0 - p_j^0|}$$

Future Work on the Model

- Allowing non-trivial weights of respect for yourself.
- Allowing non-constant weights of respect.
- Study rate and nature of convergence.

Why Trust a Mathematical Model?

- It's repeatable.
- It guarantees a result.
- Every agent's views are taken on board.
- There are some safeguards against manipulation and bullying:
  - Deliberate over- or under-bidding as an attempt to skew the results runs the risk of getting lower respect weightings, so can be counterproductive.
  - Bullies are not give the opportunity, and initial attempts to bully run the risk of lowering respect weightings and could be counterproductive.

But is it Really Consensus?

- Yes!
- Of course this answer requires a particular interpretation of both the model and the results.
- But this is a very plausible interpretation of the model.
- It is not unlike seeing posterior probabilities as the probabilities that an agent ought to hold in light of new evidence (i.e. as opposed to merely numbers spat out by Bayes's formula).
- Or Nash equilibria as the solutions players in a game ought to shoot for.

Moore's Paradox and Consensus

- There is something (Moore) paradoxical about asserting "P but my guru says $\neg P$, and my guru is always right about such things".
- Similarly, it is (Moore) paradoxical to assert that "my preferred criterion weight is p but someone whom I trust on such matters disagrees with me".
- What should an agent do who finds them self disagreeing with someone whom they see as an expert?
- They ought to genuinely revise either their criteria weights or change their weight of respect for the expert.
- The model strongly suggests that the convergence is to a consensus not a compromise.

Good Decisions versus Happy Decision Makers

- A good decision may not be popular, and a popular decision may not be good.
- Here, I take a good decision to be one that is, in some sense, best for the group, irrespective of what they believe to be in their best interest.
- “Happy decision makers” is a situation where the decision makers are satisfied with the process and happy with the decision, even if it is not what they may have personally wanted (they may have wanted to simply “have a voice”, for instance).
- It might seem obvious that it is the former, not the latter (if we must choose between the two).
- But sometimes there may be no correct decision so the whole exercise is about stake-holder participation and procedural justice.

Evaluating the Model in these Terms

- The Modified Lehrer-Wagner model of consensus has several nice features:
  - Transparent and difficult-to-manipulate mechanism for arriving at consensus.
  - All agents' views are incorporated.
  - Guarantees a result.
- It thus seems to satisfy many features required for happy decision makers.
- But does it deliver good decisions?
- That will depend on the weights of respect and the initial criteria weights, but it is hard to see how the proposed model can be held accountable for poor choices with these.
- It is at least capable of delivering good decisions under suitable conditions. Isn't that enough?

Selected Bibliography I

Lehrer, K., and Wagner, C. 1981. *Rational Consensus in Science and Society*, Reidel, Dordrecht.

Regan, H.M., Colyvan, M. and Markovchick-Nicholls, L. 2006. 'A Formal Model of Consensus and Negotiation in Environmental Management', *Journal of Environmental Management*, 80(2): 167–176.

Steele, K., Regan, H.M. Colyvan, M. and Burgman, M.A. 2007. 'Right Decisions or Happy Decision Makers?', *Social Epistemology*, 21(4):349–368.