



# Computer Science and Artificial Intelligence Laboratory

## Technical Report

MIT-CSAIL-TR-2005-054  
AIM-2005-024

August 16, 2005

---

### Collective Choice with Uncertain Domain Models

Whitman Richards

| Report Documentation Page                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                    |                                     |                            | Form Approved<br>OMB No. 0704-0188                  |                                 |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|-------------------------------------|----------------------------|-----------------------------------------------------|---------------------------------|
| Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. |                                    |                                     |                            |                                                     |                                 |
| 1. REPORT DATE<br><b>16 AUG 2005</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                    | 2. REPORT TYPE                      |                            | 3. DATES COVERED<br><b>00-00-2005 to 00-00-2005</b> |                                 |
| 4. TITLE AND SUBTITLE<br><b>Collective Choice with Uncertain Domain Moldels</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                    |                                     |                            | 5a. CONTRACT NUMBER                                 |                                 |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                    |                                     |                            | 5b. GRANT NUMBER                                    |                                 |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                    |                                     |                            | 5c. PROGRAM ELEMENT NUMBER                          |                                 |
| 6. AUTHOR(S)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                    |                                     |                            | 5d. PROJECT NUMBER                                  |                                 |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                    |                                     |                            | 5e. TASK NUMBER                                     |                                 |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                    |                                     |                            | 5f. WORK UNIT NUMBER                                |                                 |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br><b>Massachusetts Institute of Technology,Computer Science and Artificial Intelligence Laboratory (CSAIL),32 Vassar Street,Cambridge,MA,02139</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                    |                                     |                            | 8. PERFORMING ORGANIZATION REPORT NUMBER            |                                 |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                    |                                     |                            | 10. SPONSOR/MONITOR'S ACRONYM(S)                    |                                 |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                    |                                     |                            | 11. SPONSOR/MONITOR'S REPORT NUMBER(S)              |                                 |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT<br><b>Approved for public release; distribution unlimited</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                    |                                     |                            |                                                     |                                 |
| 13. SUPPLEMENTARY NOTES                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                    |                                     |                            |                                                     |                                 |
| 14. ABSTRACT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                    |                                     |                            |                                                     |                                 |
| 15. SUBJECT TERMS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                    |                                     |                            |                                                     |                                 |
| 16. SECURITY CLASSIFICATION OF:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                    |                                     | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES<br><b>19</b>                    | 19a. NAME OF RESPONSIBLE PERSON |
| a. REPORT<br><b>unclassified</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | b. ABSTRACT<br><b>unclassified</b> | c. THIS PAGE<br><b>unclassified</b> |                            |                                                     |                                 |

## Collective Choice with Uncertain Domain Models

Whitman Richards  
wrichards@mit.edu

CSAIL  
MIT 32-364, Cambridge, MA 02139

### Abstract

When groups of individuals make choices among several alternatives, the most compelling social outcome is the Condorcet winner, namely the alternative beating all others in a pair-wise contest. Obviously the Condorcet winner cannot be overturned if one sub-group proposes another alternative it happens to favor. However, in some cases, and especially with haphazard voting, there will be no clear unique winner, with the outcome consisting of a triple of pair-wise winners that each beat different subsets of the alternatives (i.e. a “top-cycle”.) We explore the sensitivity of Condorcet winners to various perturbations in the voting process that lead to top-cycles. Surprisingly, variations in the number of votes for each alternative is much less important than consistency in a voter’s view of how alternatives are related. As more and more voter’s preference orderings on alternatives depart from a shared model of the domain, then unique Condorcet outcomes become increasingly unlikely.

### 1. Introduction

There are abundant theoretical results showing that for large numbers of voters each having random preferences over a large set of alternatives, there will almost surely be no stable agreement or unique outcome (e.g. Arrow 1963, Campbell & Tullock 1965, Kelly 1986, Saari 1994, Jones et al 1995.) To insure consensus, it has been clear for decades that some form of constraint must be introduced that prohibits voters from choosing alternatives haphazardly. One plausible constraint is that individual preference orders are consistent with a shared global model for relating alternatives (Runkel, 1956). In this case, the probability of the group reaching a stable agreement is over 90% (Richards et al., 1998, 2002.) For certain types of shared models, agreement is guaranteed regardless of the numbers of voters and their voting power. Simple examples of shared models relating alternatives include how presidential candidates are positioned along a liberal to conservative dimension, the organization of taste choices for soft (or alcoholic!) drinks, the perceived relation between landmarks

in a city, the democratic versus communist versus industrial wealth of nations, etc.

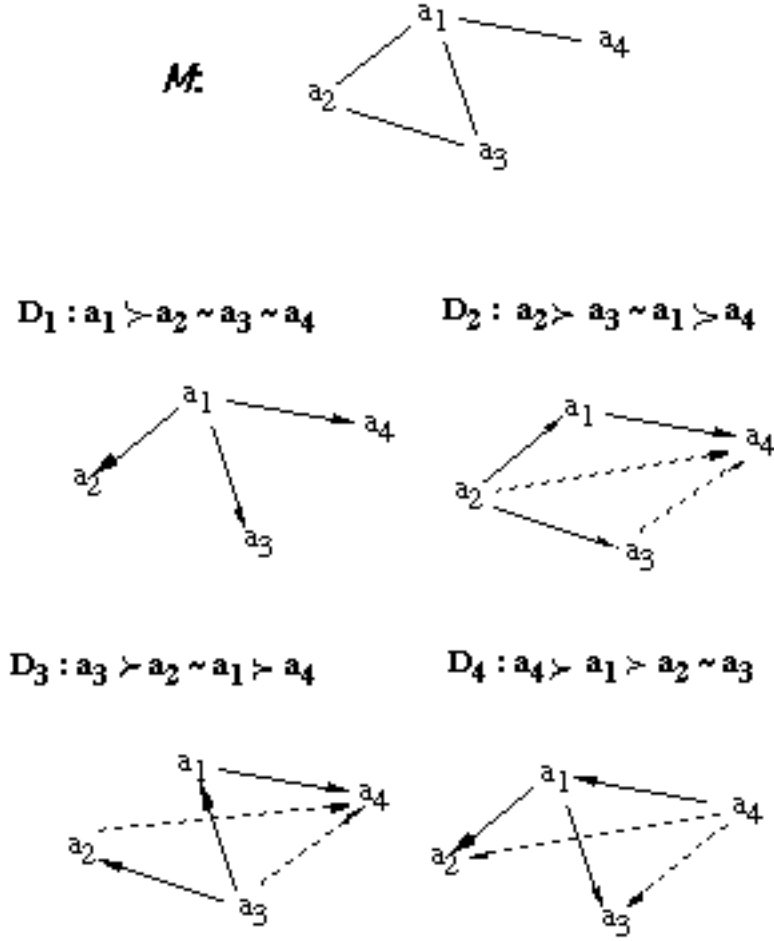
To insure consensus, important conditions include (i) that each individual votes faithfully, or not at all when "in doubt", (ii) that there is no uncertainty or external source of noise that perturbs a voter's ranking of alternatives, and (iii) that a voter's ranking is consistent with the shared global model. Any violations of these conditions will reduce the odds for consensus. Here, we explore the reduced likelihood of unique winners when a shared global model for relating alternatives is violated. The principal result will be that imperfect knowledge of a domain has small consequence if individuals vote faithfully, but haphazard preference orderings that are inconsistent with a shared domain model can create havoc.

## 2. The Shared Model Constraint

The main assumption is that alternatives are related by a labeled connected graph  $M(A,e)$  with a set of vertices  $A$  representing the alternative choice set  $\{a_1, \dots, a_n\}$  and a set of edges  $e$  that indicate a non-metric "similarity" relationship between the alternatives (Shepard 1980, Borg & Lingoes 1987.)

Figure 1 shows a simple example of a graph  $M(A,e)$  representing parameter changes between four alternatives,  $a_1, a_2, a_3, a_4$ . Each edge in  $M(A,e)$  connecting two alternatives indicates that those alternatives differ in a single attribute. To illustrate, if the alternatives are choices among alcoholic drinks,  $a_1$  and  $a_4$  may be two brands of a scotch whiskey such as Glenlivet ( $a_1$ ) versus a scotch liquor like Southern Comfort ( $a_4$ ). Likewise,  $a_2$  and  $a_3$  may differ from  $a_1$  by two different blends of bourbon, such as Jack Daniels or Jim Bean.

In the ideal case, we assume that each voter's preferences are consistent with the relationships specified by the shared model  $M(A,e)$ , with each voter having a unique most-preferred alternative, called the ideal point. Lower-ranked preferences follow from the set of transitive paths of the graph, with the second-most preferred alternatives being vertices adjacent to the ideal point, third-most the next set, etc. An ordering that jumps around, violating the relationships among alternatives in  $M(A,e)$  is not allowed (except later when noise or uncertainty is explicitly introduced.) Let  $D/i$  be the partial order induced from  $M(A,e)$  beginning at the ideal point  $a_i$ . Then  $D/i = (A,P)$  is an asymmetric and transitive directed graph where  $(a_j, a_k) \in P$  iff the number of edges on the shortest path from  $a_i$  to  $a_j$  is less than the number of edges



**Figure 1:** Set of Partial Orders Induced from ***M***.  
(No arc means alternatives are indifferent; dashed lines are implied by transitivity over preferred or indifferent alternatives.)

on the shortest path from  $a_i$  to  $a_k$ . If  $(a_j, a_k) \in P$  then we say that  $a_j$  is preferred to  $a_k$ , denoted  $a_j \succ a_k$ . If the number of edges on the shortest path from  $a_i$  to  $a_j$  is equal to the number of edges on the shortest path from  $a_i$  to  $a_k$  then we say that  $a_j$  and  $a_k$  are non-comparable (or indifferent) for those voters with ideal point  $a_i$ , denoted as  $a_j \sim a_k$ . Figure 1 gives an example of the set of four directed graphs induced from a model ***M*** and the corresponding sets of feasible preferences over the set of these alternatives.

### 3. Definitions and Notation

Let  $w = (w_1 \dots w_n)$  be the normalized weights over the  $n$  preference types -- i.e.,  $w_i$  is the proportion of voters with ideal point  $a_i$  and thus the proportion of voters with the partial order  $D_i$  over the set of alternatives  $A$ . Let  $|a_j \succ a_k|$  denote the number of voters for whom  $a_j$  is preferred to  $a_k$ . Then an alternative  $a_j$

$a_j \in A$  is the alternative most preferred by the group if for all  $a_k \in A$ ,  $a_k \neq a_j$ ,  $|a_j > a_k| > |a_k > a_j|$ . Hence,  $a_j$  is the top-ranked alternative or, more simply, "the winner". The Condorcet tally method, which evaluates all pairs of alternatives, is used to find this winner (Condorcet 1785.)

Very often in noisy contests, there will not be a Condorcet winner. Rather, one alternative  $a_j$  may beat  $a_k$  in a pair-wise comparison, but  $a_k$  is beaten by  $a_i$ , which in turn beats  $a_j$ . If either  $a_i$ ,  $a_j$ , or  $a_k$  also beat all remaining  $n-3$  alternatives, then there is a top-cycle and no winner. We call such outcomes unstable.

*Stability* (or conversely, the instability of an outcome): For a fixed set of alternatives and model  $M_n$ , the stability of an outcome is the probability that there will not be a top-cycle, or, equivalently, that there will be a unique Condorcet winner (excluding ties.)

Not to be confused with the stability is the robustness of an outcome. For example, an outcome may not include top-cycles, but still be very sensitive to the choice of weights, or to the particular form of the model  $M_n$ .

*Robustness*: The robustness of an outcome is the likelihood that perturbations in the edge set for model  $M_n$ , or fluctuations in the weights on alternatives will lead to a different winner.

Note that stability measures the ease with which an outcome can be overturned by another alternative, whereas robustness tests whether or not the same outcome will be reached following some perturbation. Following a brief aside on relevant aspects of robustness, we focus on stability.

#### **4. Robustness**

Robustness impacts stability analysis in two ways: (i) the choice of tally procedure and (ii) the relative roles of model  $M_n$  compared with weight variations on alternatives.

The simplest and most common method for choosing winners among a set of alternatives is simple Plurality, i.e. a winner-take-all. This procedure ignores any model relating alternatives, because the outcome is that alternative with the maximum number of votes (or here, equivalently, the maximum weight node in the graphical version of  $M_n$ .) The plurality winner need not be a majority winner, and in extreme cases will garner only as few percent of the total votes. Not surprisingly, this winner will be very easy to overturn, and hence is not robust. In

contrast, the Condorcet and Borda procedures favored here are quite robust to variations in voting strengths if there is some modicum of relationships among alternatives (Condorcet 1785, Borda 1786, Young 1986.) These two procedures are highly correlated (>90%) with the most likely winner being that alternative receiving the most support from many similar alternatives (Richards & Seung 2004.) Hence variations in voting strengths for one alternative become diluted with much less impact. Appendix 1 defines and compares data for these three tally procedures, and shows the striking advantage of the Condorcet and Borda winners.

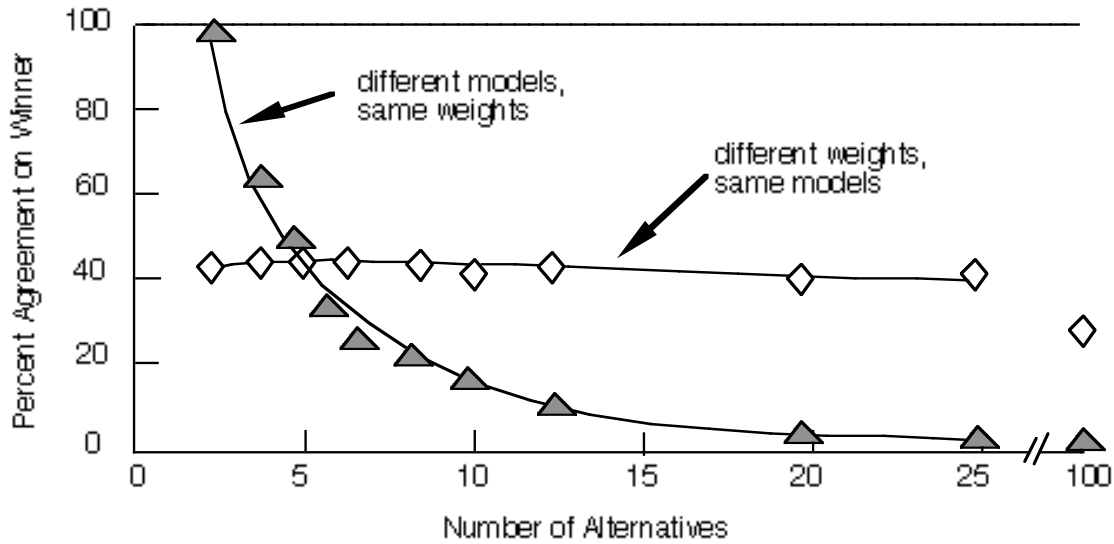


Figure 2: Robustness of winners to perturbations in either weights on nodes (open diamonds) or to the structure of model (gray triangles.) Models are random graphs; weights are taken from a uniform distribution.

To further reinforce the importance of model  $M_n$  in a choice domain, rather than weights on alternatives, consider Figure 2. In this figure, the two curves differ in whether the structure of the domain model is altered, or whether the weights on vertices (alternatives) are changed. Again, as will be inferred unless otherwise noted, weights that voters place on vertices in  $M_n$  are chosen from a uniform distribution, and the graphical model  $M_n$  with  $n$  vertices is a random graph with all edges bi-directional with edge probability of one-half. The directed graphs  $D_i$  governing a voter's preference orders are limited to the ideal point and its neighbors in  $M_n$ , with all lower ranked preferences taken as equivalent (i.e. indifferent.) The open diamonds show that when the domain model is held fixed, but a second set of weights on alternatives are chosen from a uniform distribution, there is little change in the percent of agreement in outcomes, which remains

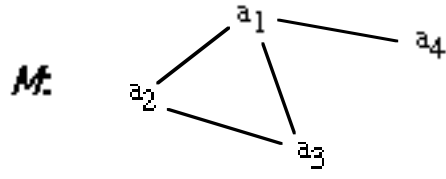
roughly constant at 40% for  $n < 30$  and  $1/2 < p < 2/3$ . In contrast, when the weights are held fixed, but applied to two different random models for  $M_n$ , there is a dramatic fall in agreement between the two winners (gray triangles.) Elsewhere we have shown that for  $n > 10$  the expected agreement in outcomes when  $M_n$  is revised is roughly  $(n-k)/n$ , where  $k$  is the number of vertices in  $M_n$  that have been altered (Richards, 2005.) The sections that follow detail the result highlighted in Fig. 2, showing that both the structure of the model  $M_n$  and the extent to which voter's preferences adhere to this model are the major sources of instability in outcomes.

## 5. Methods

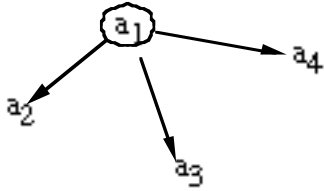
Our results are largely based on Monte Carlo simulations. The procedure is to construct a connected graph  $a$  with  $n$  vertices and edge probability  $p$ . (For most of these simulations,  $p = 1/2$ .) In the ideal case, with no "noise" and faithful voting, the random graph  $a$  determines the set of  $n$  feasible preference orders, with each preference order assigned a weight  $w_i$ ,  $i = 1, \dots, n$ , drawn uniformly from the interval  $[0, 1000]$ . These weights create an  $n$ -tuple  $w_i$  representing the distribution of voters over feasible preferences. We then evaluate all a pairs of alternatives to determine whether one alternative beats all others using the Condorcet tally. The number of trials varied between 200 and 500 depending upon the probability of no-winner. Because of the high correlation between the Borda and Condorcet winners, the presence of Condorcet top-cycles gives a good indication of the likelihood that a Borda winner can be overturned. The maximum average error in the results is about 3 percent.

## 6. Uncertainty in Voting

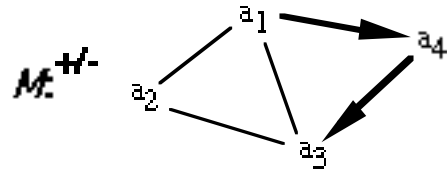
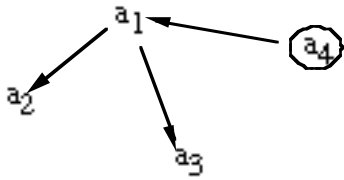
We assume that each voter's ideal point (first choice) is respected, and only second or lower ranked preferences are subject to uncertainty. In the extreme case, there is no regard for any shared model  $M_n$  relating alternative choices, and all votes are cast haphazardly, excepting the voter's first choice. Backing away from this extreme, we can group voters by their ideal points, and let the haphazard voting take place only for small segments of the voting population, and only for alternatives that do not correspond to the ideal points. Figure 3 is an example. For this particular model  $M_4$ , the directed edge from  $a_1$  to  $a_4$  and the new directed edge from  $a_4$  to  $a_3$  shows that now  $a_3$  (rather than  $a_1$ ) is a second choice preference for  $a_4$ , whereas  $a_4$  still remains one of three second-choices for  $a_1$ . The result of these changes is a new preference ordering  $D_4$  for  $a_4$ , but NO change



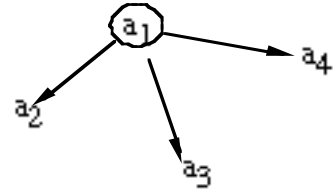
$$D_1 : a_1 \succ a_2 \sim a_3 \sim a_4$$



$$D_4 : a_4 \succ a_1 \succ a_2 \sim a_3$$



$$D_1 : a_1 \succ a_2 \sim a_3 \sim a_4$$



$$D_4 : a_4 \succ a_3 \succ a_1 \sim a_2$$

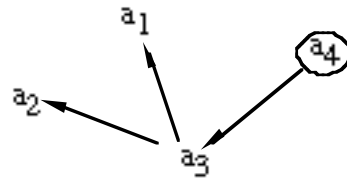


Figure 3: Preference orders for  $a_1, a_2, a_3$  remain the same, but preference order for  $a_4$  is altered.

in  $D_1$  for  $a_1$ . The most extreme case of such rearrangements will be when edges of  $M_n$  are chosen randomly from a list of all directed edges among the alternatives. This case will be referred to as a “random directed graph”, to be contrasted with the ideal  $M_n$  whose edges are all bidirectional.

### 6.1 Haphazard Voting

The simplest perturbation to describe is the most extreme: excepting the ideal point, all individuals make choices haphazardly when choosing between the two alternatives being compared in each Condorcet trial. In other words, model  $M_n$  becomes irrelevant, and furthermore, each voter’s preference orders

are changing widely from one Condorcet test pair to the next. Figure 4 shows the result: the probability of a top-cycle, and hence no unique winner, rises rapidly toward 100%, already reaching 90% for 10 alternatives. In contrast, as shown by the lowest curve, if voters respect the

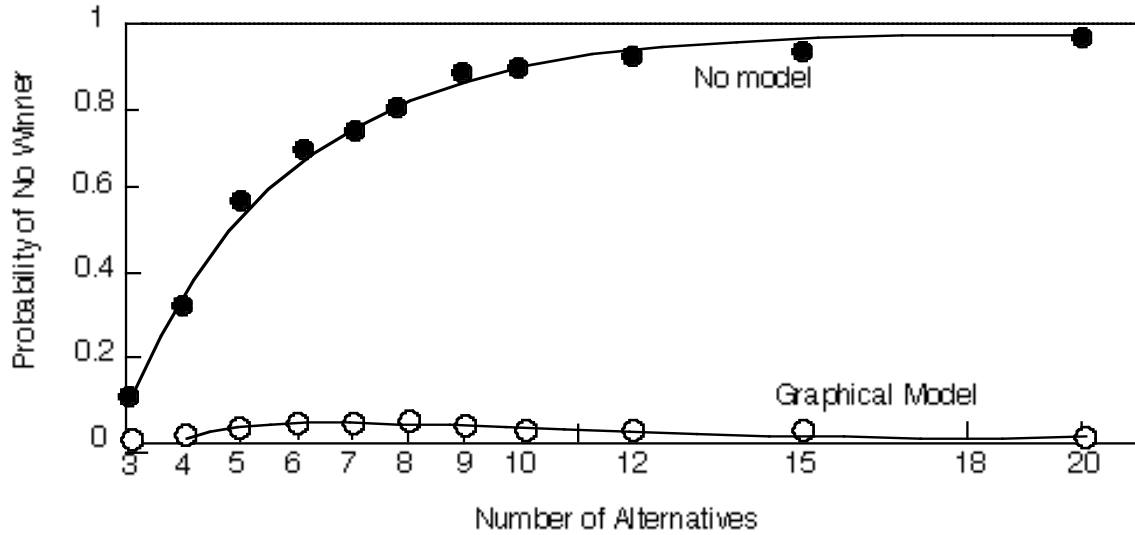


Figure 4: Top curve: random preference orders. Bottom curve: preference orders of voters respect a shared domain model relating alternatives.

model  $M_n$  (here a random graph relating alternatives with  $p(e) = 1/2$ ), the chance of no winner is less than 5%. Hence model  $M_n$  provides enormous stability in outcomes, because the likelihood of no-winner is small.

## 6.2 Haphazard Preferences Orders

The disastrous case above is equivalent to all voters changing their preference orders for each Condorcet comparison. Let us then require that each voter's preference ordering on alternatives be fixed. Thus rather than constraining the preference orders  $D_i$  to be constrained by a shared model  $M_n$ , let the  $D_i$ 's be chosen at random (but as before, limited to three levels as in Fig. 1.) This perturbation can be effected by altering the bidirectional edges in  $M_n$ , specifically by choosing edges at random from a uniform distribution of all  ${}_nC_2$  edges.

Let  $e_{ik}$  be the directional edge linking  $v_i$  to  $v_k$ , and  $e_{ki}$  be the opposite directional edge from  $v_k$  to  $v_i$ . If voter (i) associated with vertex  $v_i$  drops a similarity relation to  $v_k$ , then edge  $e_{ik}$  is deleted and the remaining directional

edge from vertex  $v_k$  is  $e_{ki}$ . In this process, only vertex  $v_i$  has been altered, or, equivalently, only voter (i) has changed his preference orderings.

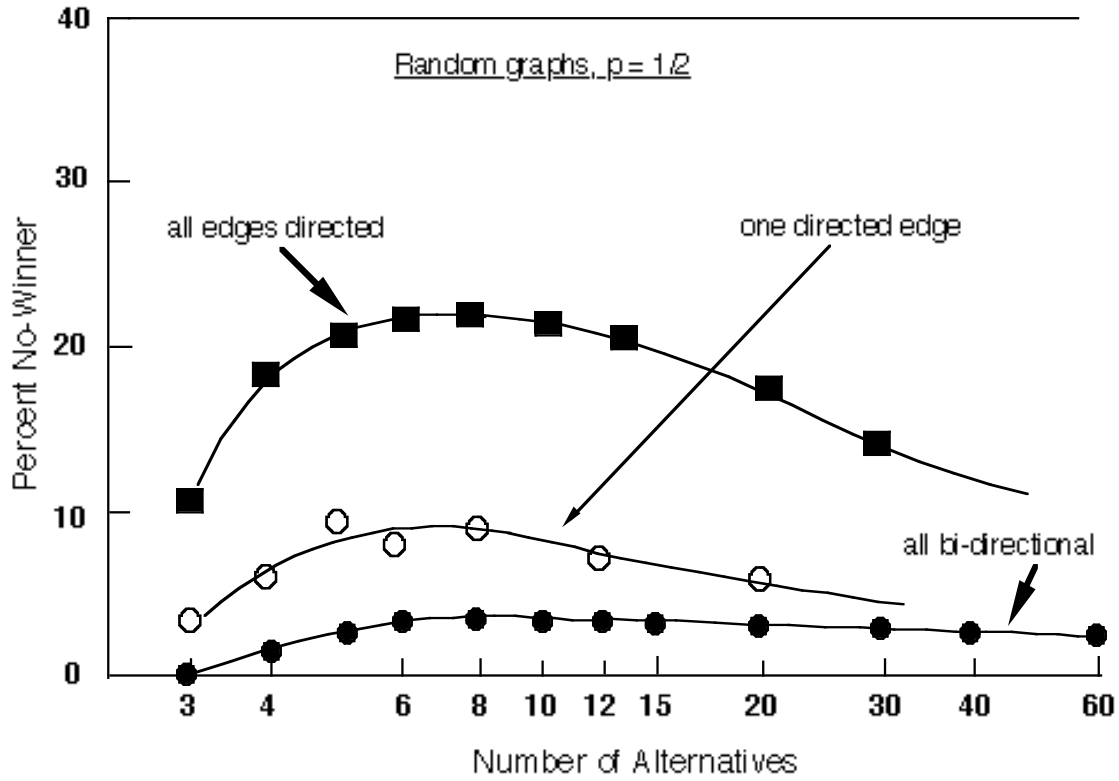


Figure 5: Probability of top-cycles (i.e. no winner) when the shared domain model is perturbed (top and middle) versus the ideal case where all preference orders respect the domain model (solid dots.)

The top curve in Fig. 5 (filled squares) shows the probability of top-cycles when all voters rearrange their edges in  $M_n$ , choosing new neighbors from a uniform distribution of  $(n-1)$  vertices. (Hence for  $p = 1/2$ , about one-half (0.4) of the links between vertices will be bi-directional.) For this condition, note the maximum of roughly 20% compared with only 4% of top cycle outcomes for the ideal bi-directional  $M_n$  (lowest curve.) Significantly, unlike random noise on alternative weights, as the number of alternatives becomes large, the odds for no unique winner become small.

Between these two cases of all bidirectional or mostly directed edges in  $M_n$  is shown another, much less extreme “miss-matched” condition where only one type of voter rearranges only one edge (open circles.) An intermediate miss-match is if all voters rearrange only one relationship in the global domain model  $M_n$ ; the result is similar and roughly intermediate between the solid squares and

open circles. In the complementary miss-match where only one type of voter rearranges *all* edges, again the result is also an intermediate curve with a maximum near 8 alternates. These results are surprising: even one type of voter with directed edges has a disturbing effect on the probability of consensus and the effect is roughly equivalent to all voters mismatching one relationship.

## 7. Partial Uncertainty

Here we explore further the condition where most of the population will agree on a model for the domain and vote accordingly, but a smaller segment will have beliefs and preference orders inconsistent with the shared model held by the majority. How detrimental to achieving consensus will be an aberrant set of voters?

### 7.1 Preliminaries

As before, the manipulation is for each individual to vote their first choice but otherwise choose alternatives arbitrarily during each tally, ignoring the shared model  $M_n$ . The fraction of haphazard votes cast will be the main independent variable. In a Condorcet tally with the “indifferent” option, there will be a one-third probability of choosing either one of the two alternatives being compared, or simply punting. Because punting will not disrupt a fair social outcome, the uncertainty or noise in this case will be taken as 67%, namely the two-thirds of the votes that are cast for one alternative or another. Obviously, as the number of haphazard votes increases, the probability of no-winner will also increase (see Fig.4.) We can increase the odds for such negative outcomes in two ways: (i) by adding more uncertain (or rogue) voters who always vote haphazardly, or (ii) by distributing the haphazard votes across all voters. As will be shown, one set of curves predicts the unsuccessful outcome in both cases.

### 7.2 Haphazard votes for all less preferred alternatives

The solid curves in Fig. 6 show the probability of no Condorcet winner when varying amounts of noise or uncertainty is distributed uniformly across all voters, for all choices other than their first choice. Each curve represents the result for different random graphs having vertices ranging from 3 to 100, with edge probability of one-half. These results are rather insensitive to whether the random graph is sparse or dense, specifically for edge probabilities ranging from 1/4 to 3/4. Note that the slope of the curves is about one over most of the range, with the percent no-winner proportional to the uncertainty for a random

graph of known size  $n$ . As the size,  $n$ , of these graphs increases, so does the effects of uncertainty or noise in the aggregation process. The translation from one curve to another is approximately  $O(n^2)$  as  $n$  increases. Note that even a

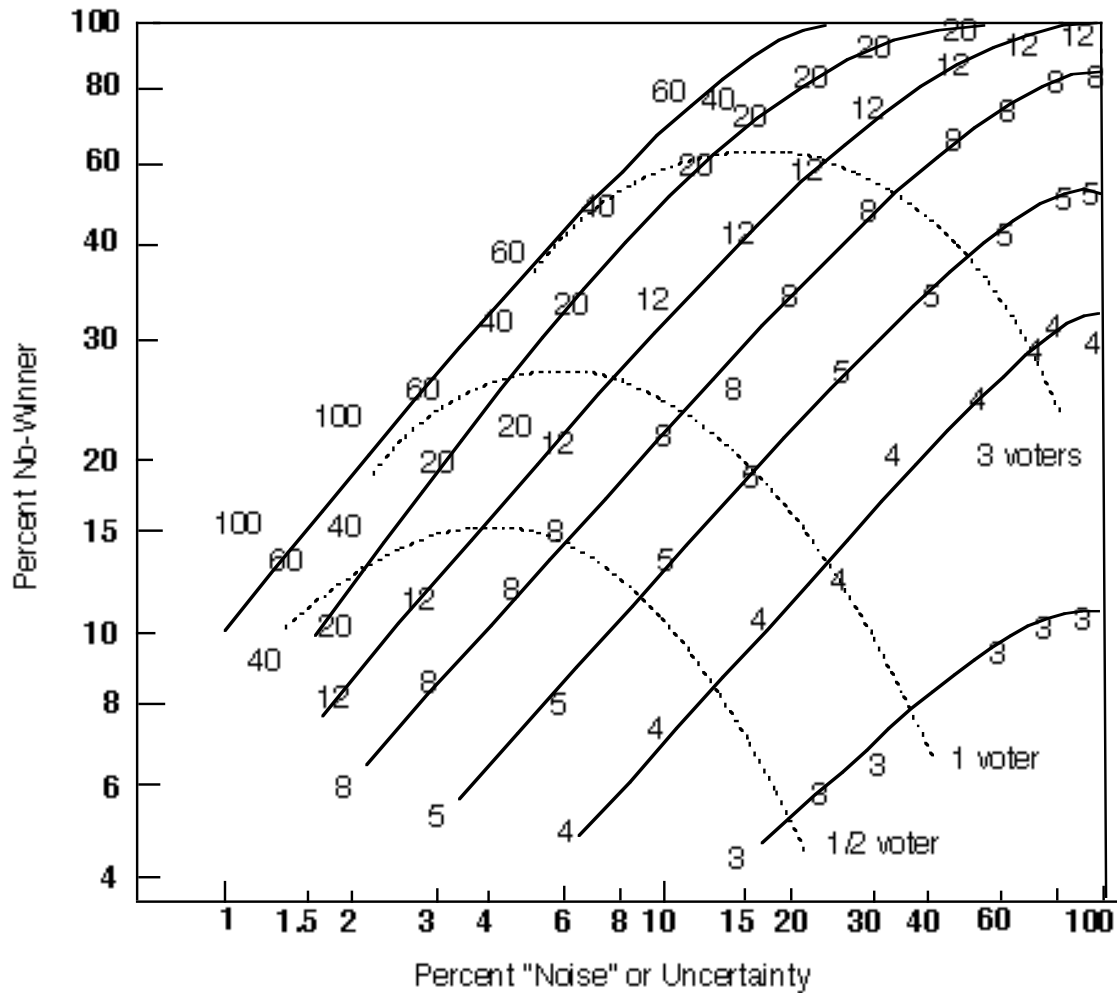


Figure 6: Solid curves: Noise distributed evenly among all agents. Numbers indicate size of random graph

small percent of haphazard votes (e.g. 10%) can have severe consequences on achieving successful outcomes for choice sets larger than twelve alternatives.

The dashed lines summarize simulation results when a small group of voters are uncertain, and vote haphazardly 100% of the time. (Recall that the voting power for any type of voter is chosen from a uniform distribution of weights.) For a single type of rogue voter among a group of four types (alternatives), the effect on the outcome will be equivalent to distributing the

noise over 25% of the total votes cast. Hence the dashed curve labeled “1 voter” crosses the 4- alternative solid curve at a point directly above 25% noise on the abscissa, corresponding to about 12% no-winners in each case. Similarly, if there are eight different voter types (i.e. a random graph relating eight alternatives), then the same dashed line labeled “1 voter” will cross the 8- alternative solid curve directly above  $1/8 = 12\%$  noise, corresponding to about 22% no winners whether or not the noise is concentrated in one type of voter, or distributed across all voters. For three voters, the calculation is similar, simply finding the noise equivalent if all rogue voter’s votes were distributed across all voters. The lowest dashed curve labeled 1/2-voter corresponds to one voter who votes haphazardly 50% of the time.

### 7.3 Haphazard votes for third or less desired alternative

One might expect in practice that uncertainty will increase for less preferred alternatives. In other words, given two alternatives being compared, if these alternatives are third or fourth ranked in an voter’s preference ordering, uncertainty over which to favor should be much higher than for the first and second choices. Consider then voters who introduce noise only if both of the two alternatives being contested are third or higher choices. Thus in the shared domain model, the voter’s first choice or ideal point is not adjacent to the two contested alternatives. Fig. 7 shows the results are dramatically different from the previous case.

First, although only results for 40 vertex random graphs are shown, the size of the graph ( $n > 10$ ) makes little difference in the main effect. Rather, unlike the earlier results, here the edge probability of  $M_n$  (or  $G_n$ ) drastically changes the relations between voter uncertainty and the probability of no winner. For highly connected random graphs [  $p(e) \rightarrow 1$  ], noise is ineffective – as expected as the covering becomes complete– whereas for sparse graphs such as chains, an almost trivial amount of noise or uncertainty can create a high probability of no-winners.

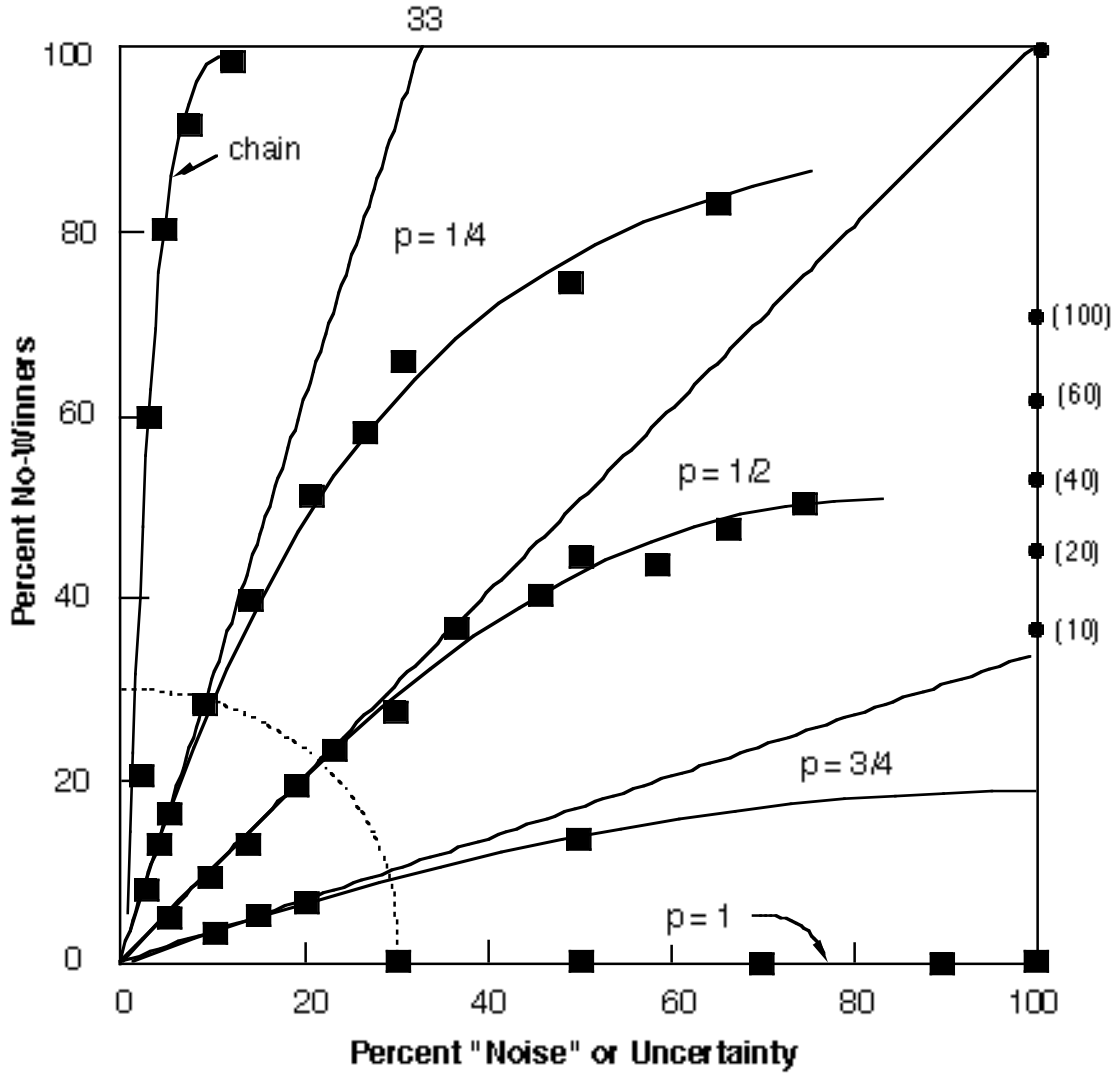


Fig. 7: Haphazard votes for alternatives not immediately similar to ideal point (i.e. non-adjacent vertices in  $G_n$ .) Data are for 40 vtx random graphs with edge probability,  $p$ , as indicated.

We also see a rather pleasing correlation between the edge probability of  $M_n$  (i.e.  $G_n$ ) and the asymptotic slope of the relation between no-winners and uncertainty or noise. As the noise approaches zero, the slopes of the curves are  $(1-p)/p$  for edge probability  $p$ . The cases for  $p = 1/2$  and  $p = 1$  illustrate. When  $p = 1$ , the slope is zero; whereas for  $p = 1/2$  the asymptotic slope is one.

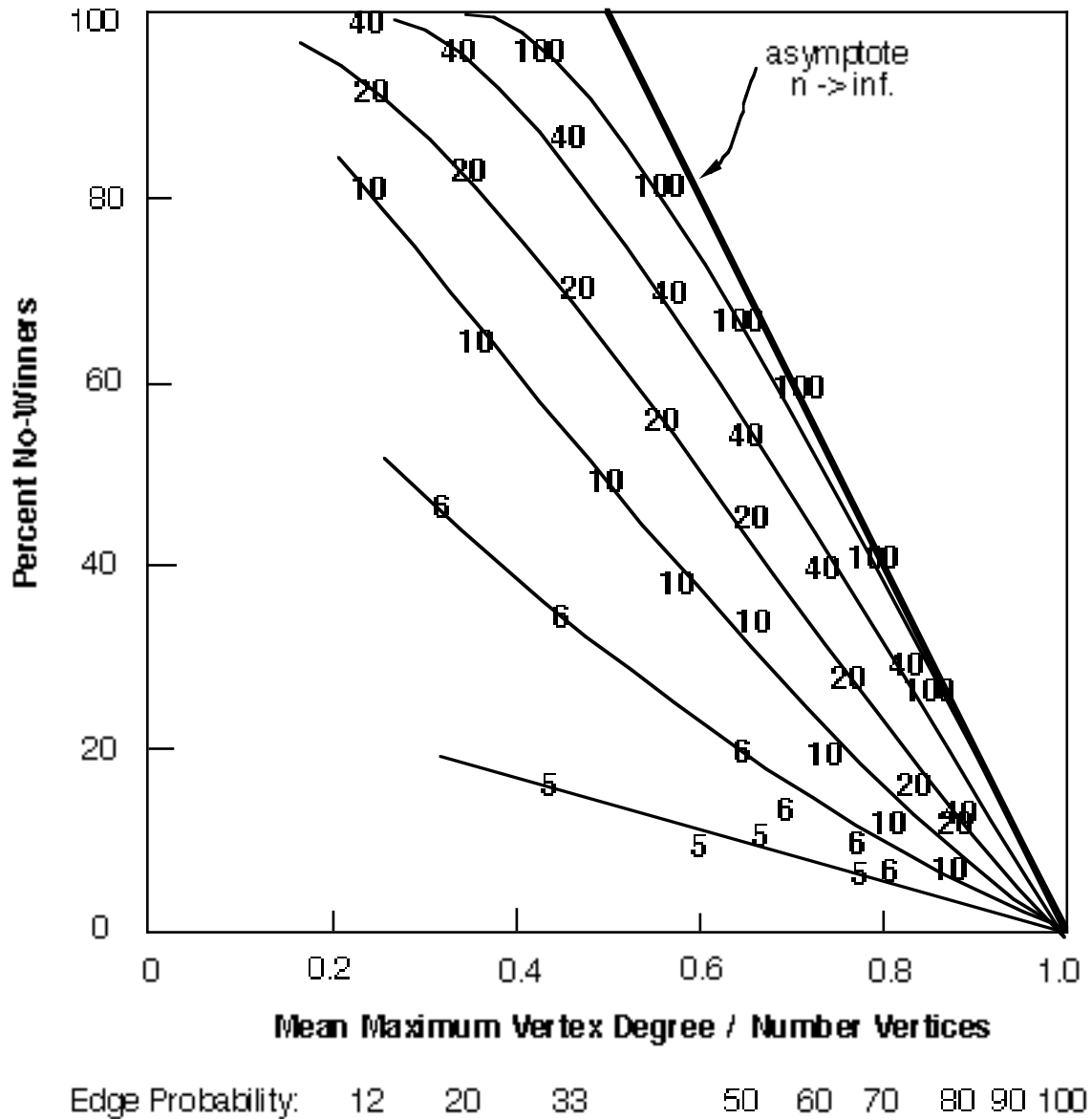


Fig. 8. Same data as in fig. 7, replotted to show asymptotic behavior for large  $n$ .

Figure 8 provides another plot of a portion of the same data, revealing a second asymptotic property of uncertainty or noise limited to third or lower-ranked preferences. Here, the abscissa is the mean maximum degree of  $M_n$  ( $G_n$ ) normalized by the number of vertices  $n$ . For all points shown, the noise is fixed at the maximum of 100%, provided that the two contested alternatives are not adjacent to the voter's first choice. There appears to be an asymptotic bound on the percent no-winners versus the normalized mean maximum degree of  $M_n$ . The theoretical explanation is under study.

## 8. Summary

There are three main points: First, outcomes are very sensitive to variations of preference orderings, especially those that are inconsistent with the model  $M_n$  for the domain held by the majority. Second, when voting uncertainty is limited to deciding between choices that are third-ranked or lower (i.e. choices not adjacent to the voter's ideal point), then the type of graph (i.e. structure of  $M_n$ ) is the main factor in blocking unique winners, with the deleterious effects of uncertainty increasing as the similarity relationships in the domain model  $M_n$  becomes more sparse [Fig.7]. Equivalently, for random graphs, as the mean maximum vertex degree decreases [Fig.8], the percent of no-winners increases roughly monotonically when uncertainty is limited to third or higher-ranked alternatives. The size of  $M_n$  makes little difference for  $n > 40$  alternatives. Thirdly, and in contrast to uncertainty in less desirable preferences, when uncertainty or noise pervades *all* choices except the voter's first choice, then the odds for outcomes with no-winners increases with the size of the domain model  $M_n$ , and the structure of  $M_n$  is much less important [Fig.6].

**Acknowledgements:** Supported under AFOSR contract #6894705.

## 9.0 References

- Arrow, K.J. (1963) *Social choice and Individual Values*. Wiley, NY.
- Borda, J-C. (1786) Memoire sur les elections au Scrutin. Histoire de l'Academie Royal des Sciences.
- Borg, I & Lingoes, J. (1987) *Multidimensional Similarity structure Analysis*. Springer-Verlag, NY.
- Campbell, C. & Tullock, G. (1965) A measure of the importance of cyclical majorities. *Econ. Journal*, 75, 853-857.
- Condorcet, Marquis de (1785) Essai sur l'application de l'analyse a la probabillite des decisions rendue a la pluralite des voix, Paris (See Arrow, 1963).
- Jones, B; Radcliff, B; Taber, C. Timpone, R. (1995) Condorcet Winners and the Paradox of Voting. *Amer. Pol. Sci. Rev.* 89, 137-144.
- Kelly, J.S. (1986) Condorcet winner proportions. *Social Choice and Welfare*, 3, 311-314.

- Richards, W. (2005) Anigrafs (book in progress).  
[www.csail.mit.edu/people/whit](http://www.csail.mit.edu/people/whit)
- Richards, D., McKay, B. and Richards, W. (1998) Collective Choice and mutual knowledge structures. *Adv. in Complex Systems*. 1, 221-36.
- Richards, W., McKay, B. and Richards, D. (2002) Probability of Collective Choice with shared knowledge structures. *Jnl Math. Psychol.* 46, 338-351.
- Richards, W. & Seung, S. (2004) Neural Voting Machines. AIMemo 2005-29, MIT-CSAIL technical series. (early version Cog. Sci. Soc. 2003.)
- Runkel, P.J. (1956) Cognitive similarity in facilitating communication. *Sociometry*, 19, 178-91.
- Saari, D. G.: 1994, *Geometry of Voting*, Springer-Verlag, Berlin.
- Shepard, R (1980) Multidimensional Scaling, Tree-Fitting and Clustering. *Science* 210, 390-398.
- Young, H.P. (1986) Optimal ranking and choices from pairwise comparisons. In: B. Grofman & G. Owen (Eds.) *Information Pooling and Group Decision Making*. JAI Press, Greenwich, CT.

## 10. Appendix: Robustness of Three Tally Procedures

### 10.1 Definitions of the tallies

**Plurality P** (winner-take-all): Let there be  $n$  alternative choices  $a_i$  with  $v_i$  of the voters preferring alternative  $a_i$ . The winner is

$$\text{Plurality\_winner} = \arg\text{Max}(i) \{v_i\} \quad [1]$$

Note that no information about any similarity relationships among alternatives is captured in [1]. In other words, the Plurality voting method does not consider second ranked preferences of the voters.

**Borda B:** Assume the alternative choices are related by a model  $M_n$  that is held in common by all voters. In the unperturbed case, each voter's ranking of alternatives reflects information about choice relationships. (Note that the effective role of is to place conditional priors on the choice domain.) Although the shared model  $M_n$  has typically been represented as a graph,  $G_n$ , it is more convenient to use the matrix  $M_n$  where the entry "1" indicates the presence of the edge  $ij$  in  $G_n$  and 0 otherwise. For simplicity, we assume that the edges of  $M_n$  are undirected, meaning that if alternative  $a_1$  is similar to alternative  $a_2$ ,

then  $a_2$  is equally similar to  $a_1$ . However, directed edges require only a trivial modification to the scheme.

With  $\mathbf{M}_n$  expressed as the matrix  $\mathbf{M}_n$  we can include second choice opinions in a tally by defining a new voting weight  $v_i^*$  as

$$v_i^* = \{ 2 v_i + \sum_j \mathbf{M}_{nj} v_j \} \quad [2]$$

where now first-choice preferences are given twice the weight of the second-ranked choices, and third or higher ranked options have zero weight. (This is a simplified, revised Borda procedure.) The winner is then

$$\text{winner\_Borda} = \text{argMax}(i) \{v_i^*\} \quad [3]$$

### **Condorcet C:**

Definition: let  $d_{ij}$  be the minimum number of edge steps between vertices  $i$  and  $j$  in  $\mathbf{M}_n$ , where each vertex corresponds to the alternatives  $a_i$  and  $a_j$  respectively.

Then a pairwise Condorcet score  $S_{ij}$  between alternatives  $a_i$  and  $a_j$  is given by

$$S_{ij} = \sum_k v_k \text{sgn}[d_{jk} - d_{ik}] \quad [4]$$

with the sign positive for the alternative  $a_i$  or  $a_j$  closer to  $a_k$ . Note that if  $a_i$  or  $a_j$  are equidistant from  $a_k$ , then  $\text{sgn}=0$  and the voting weight  $v_k$  does not contribute to  $S_{ij}$ . Furthermore, as before in the Borda method, we again impose a maximum on the value of  $d_{ij}$  of 2, which means that third or higher ranked alternatives do not enter into the tally.

A Condorcet winner is then

$$\text{winner\_Condorcet} = \text{ForAll}_{i \neq j} S_{ij} > 0 . \quad [6]$$

## 10.2 Results

Figure 9 provides some additional data on the robustness of the three aggregation procedures. Here, connected random graphs of edge probability 1/4 were generated, with weights on nodes chosen from a uniform distribution. Winners were then found using Condorcet (C), Borda (B) and simple Plurality (P) tallies, the latter being simply that node with maximum input weight. Then the input weights were diddled using a uniform sampling from +/- 50% of the initial node weight (hence an average of 25% variation.) Using the same

graphical model  $M_n$ , a second set of winners were calculated. Figure 9 shows the probability that the winners were the same.

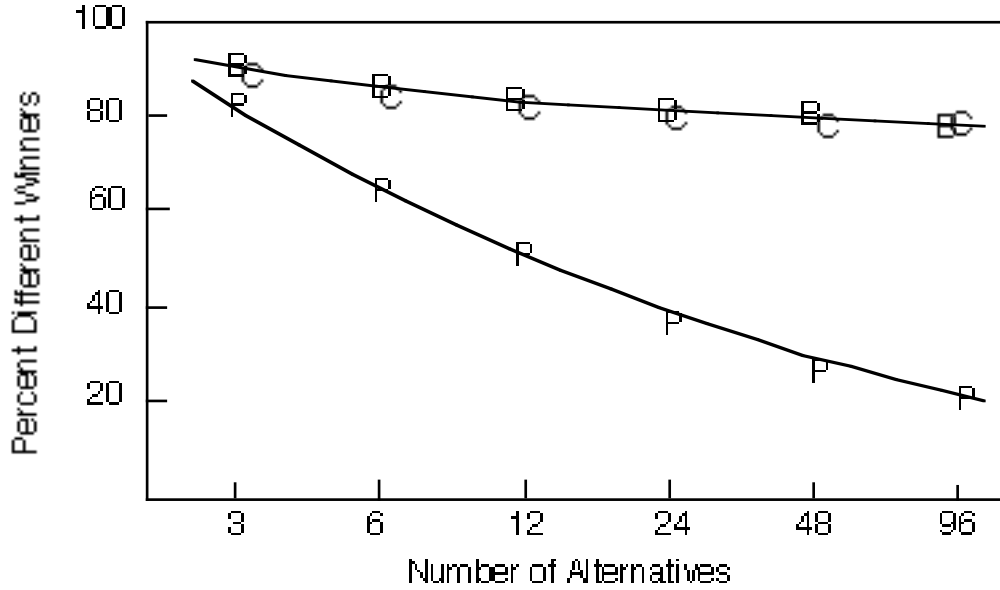


Fig. 9: Percent different Winners with mean weight variation of 50%;  
B = Borda; C = Condorcet; P = simple Plurality (winner is vertex  
having maximum weight.)

For the Borda and Condorcet procedures, over 80% of the winners remained the same (curve labeled BC). In contrast, the simple Plurality method (P) was not robust to noisy input weights. Furthermore, as shown earlier in Fig. 3, the Plurality winner seldom agreed with the Borda or Condorcet winners constrained by graphical models describing similarity relations among 10 or more alternatives. This difference shows the large effect prior knowledge about the domain can have on the determining optimal (maximum likelihood) choices.