Graphical Models and Collective Choice
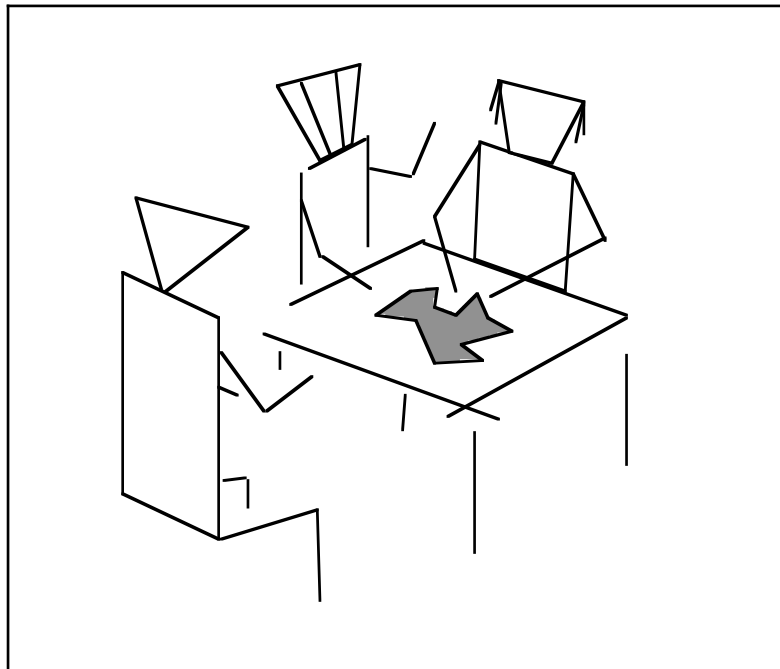
Whitman Richards

Mass. Inst. of Technology   CSAIL 32-364
Computer Science and Artificial Intelligence Lab
Cambridge, MA 02139   617-253-5776   wrichards@mit.edu

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
24-08-2005	Final Performance	01-05-2000; 31-07-2005

4. TITLE AND SUBTITLE
Graphical Models and Collective Choice

5a. CONTRACT NUMBER

5b. GRANT NUMBER
F49620-03-1-0213

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Whitman Richards

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Mass. Institute of Technology
Computer Science & Artificial Intelligence Lab
32 Vassar Street 32-364
Cambridge, MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 North Randolph Street
Suite 325, Room 3112
Arlington, Virginia 22203-1768

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT

When groups of individuals make collective decisions, it is obvious that if all members share similar goals, agreement will be reached more quickly than if members have diverse opinions. This common-sense notion is quantified by using the framework of graphical models to relate choices that group members may favor. It is shown that the similarity relations between members' choices play the dominant role in the ease or difficulty with which the group can reach agreement. If similarity relations among members' choices are sparse, then consensus is likely to be very fragile and easy to disrupt. Five specific findings of this nature are reported. In addition, there are three other spin-off results. Two are related to biological system design; the third relates measures for predictability and Shannon information.

15. SUBJECT TERMS
Graphical Models; Social Networks; Collective Choice; Decision-Making; Uncertainty

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON W.A. Richards
a. REPORT	b. ABSTRACT	c. THIS PAGE		19	19b. TELEPHONE NUMBER (include area code) 617-253-5776

AFOSR   22 Aug 05
Graphical Models and Collective Choice
Whitman Richards

MIT 32-364   Cambridge, Ma 02139   **whit@csail.mit.edu**
617-253-5776   Fax: 617-253-8335

Project Summary

Graphical models have led to important advances in probabilistic reasoning (Pearl, 1988.) We have applied similar constraints on decision-making by groups of individuals. A graphical model makes explicit relations among alternatives in the choice domain, and is a way of representing the mental models of voters. The form of the model, and the degree to which members of the group respect the model, play key roles in achieving consensus. The results have implications for social choice, decision-making, belief dynamics, man-machine interactions that entail interface agents, command and control, and the ease with which small groups of constituents are able to alter or block consensus of a much larger majority.

Project Participants: John Kraemer (Dept. Brain & Cognitive Science)
Michael Coen (Elec. Engr. & Computer Science)
Galen Pickard (Elec. Engr. & Computer Science)

Introduction

Collective choice occurs when groups of individuals, nations, neural assemblies, or more generally "agents" aggregate information or participate in group decision making. The goal is to select one alternative among many. If each agent is entitled to one vote, an obvious and common method for finding a winning alternative is to choose the one receiving the most votes. This procedure, called the Plurality procedure, is the one typically implemented in "winner-take-all" networks. Unfortunately, in noisy or controversial choices, this winner may represent the opinions of only a very small percent of the population. If such a winner were challenged head-to-head by another alternative in a pair-wise contest, the outcome frequently will be different (Saari, 1991.)

A more realistic scenario assumes voters (or agents) have some minimal information about the set of choices (Runkel, 1956), and use this information to decide for whom to vote (Saari, 1994.) Some of this information may be in the form of institutional constraints (Schelling, 1971; Young, 1998.) In these cases, voters will have a model of at least part of the choice set. Unlike the Plurality

procedure, alternative choices now play an important role in the vote, for example when first choices are thwarted or not viable. When such information is incorporated into a tally, two maximum likelihood methods for aggregating votes are the Borda Count (Borda, 1786) and the Condorcet tally (Condorcet, 1785), as shown by Young (1986.) Both procedures utilize information provided by a voter's preference rankings of alternatives. The Borda Count uses this information by weighting a voter's preferences inversely to rank; the Condorcet tally proceeds by conducting pairwise comparisons between all alternatives. Outcomes from both procedures are highly correlated (Richards, 2005.) Here we have favored the Condorcet procedure because instabilities in outcomes are made explicit when no alternative can be found that will beat all others. Indeed, as shown in Fig. 1, if individuals cast votes haphazardly without constraint on their preference rankings of alternatives, then typically *NO* Condorcet winner will be found. This finding is inconsistent with real life scenarios, implying that voters indeed have shared models about how alternatives in the choice set are related. The primary focus of our research has been to explore how perturbations in any shared model of the domain (or equivalently, the individuals' belief structures) will disrupt consensus.
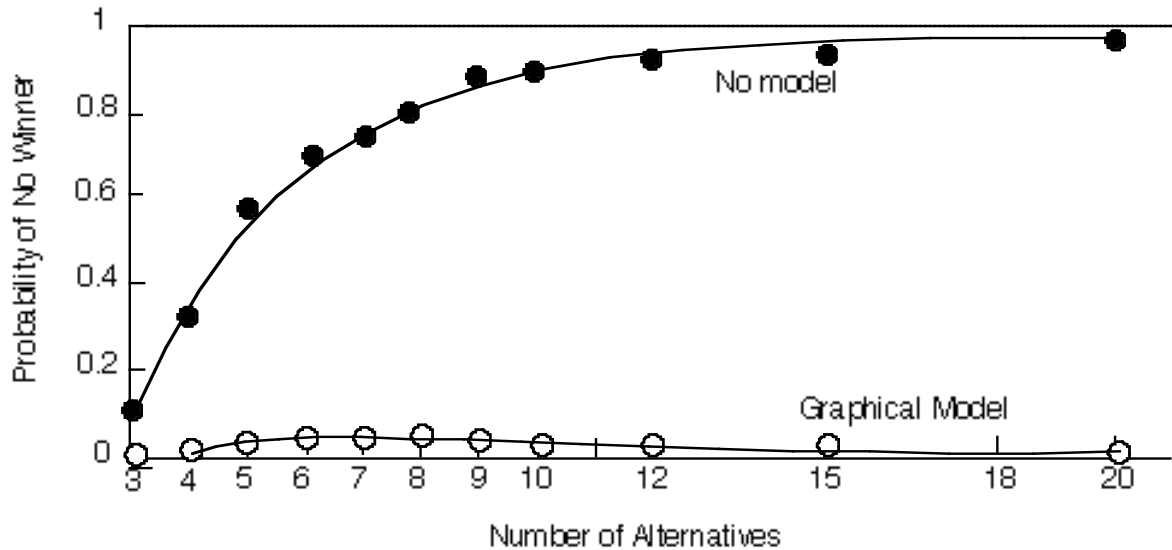

Figure 1: Top curve: random preference orders. Bottom curve: preference orders of voters respect a shared domain model relating alternatives.

2.0 The Shared Model Constraint

2.1 Graphical Model M_n:

To clarify how a shared model for the domain constrains a voter's or agent's preference rankings, consider the graphical model M_n in Fig. 2. There are

four alternatives, a_1, a_2, a_3, a_4. An edge ij connecting two alternatives indicates a non-metric similarity relationship between a_i and a_j (Shepard 1980, Borg & Lingoes 1987.) The main assumption is that if an agent most prefers alternative a_i, then that agent's second choices will be those alternatives a_j that are most similar
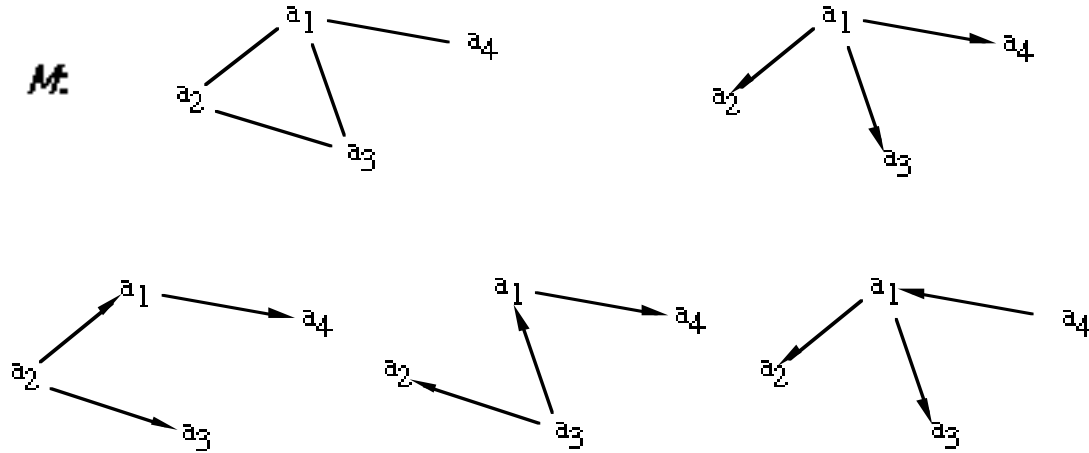

Figure 2: Set of Partial Orders Induced from **M**.

to his ideal point a_i. For example, given this particular model M_n, if an agent's first choice is a_3, then equally preferred second choices will be a_1, a_2 and a_4 will be the least desirable choice. Thus each agent has a (weak) preference ordering over the alternatives in the choice set, induced from the shared global model, M_n. (Details are discussed elsewhere: Richards et al 1998, 2002; Richards 2005.)

2.2 Definitions and Notation

Let $w = (w_1 ... w_n)$ 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 $| a_j > a_k |$ denote the number of voters for whom a_j is preferred to a_k. Then an alternative $a_j \in A$ is the alternative most preferred by the group if for all $a_k \in A$, $a_k =/= a_j$, $| a_j > a_k | > | a_k > a_j |$. 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.

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.

2.3 Methods: Our results are largely based on Monte Carlo simulations. The procedure is to construct a connected graph 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 (i.e. the model M_n) determines the set of n feasible preference orders, with each preference order assigned a weight w_i, i = 1,....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 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, (>90%), 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.

3.0 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 (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. Hence variations in voting strengths for one alternative become diluted with much less impact. Elsewhere we have documented the robustness of the Borda and Condorcet tallies over the more common winner-take-all Plurality methods. (Richards & Seung 2004, Richards, 2005a.)
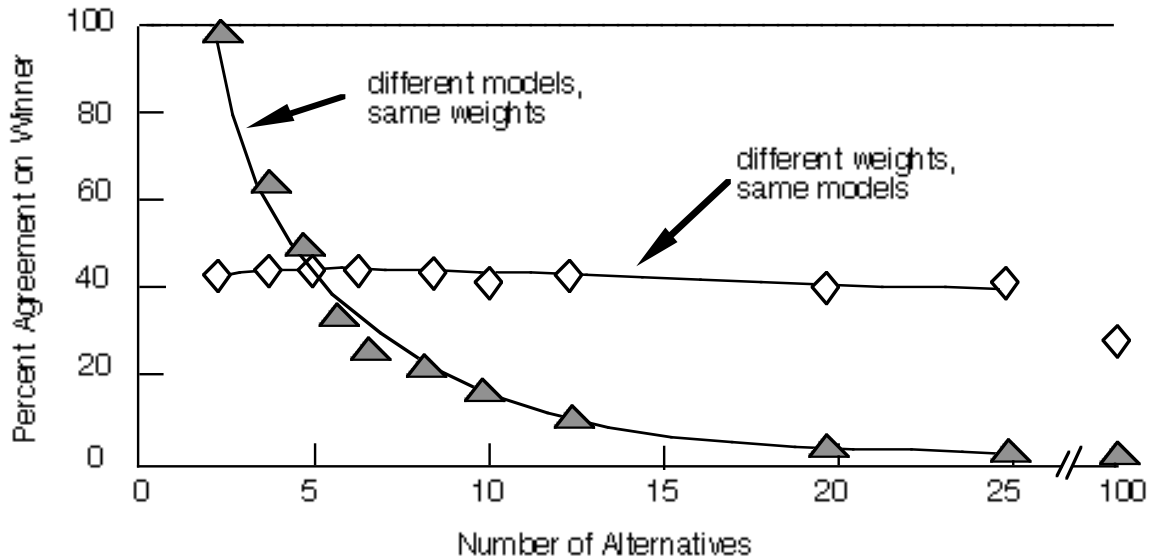

Figure 3: Robustness of winners to perturbations in either weights on nodes (open diamonds) or to the structure of model (fgray triangles.) Models are random graphs; weights are taken from a uniform distribution.

To further reinforce the importance of model M_n in a choice domain, rather than weights on alternatives, consider Figure 3. 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. (This percent varies with level of "noise on weights" introduced, but still remains essentially flat over the indicated range.) 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.)

Finding #1: The shared model M_n plays a dominant role in robustness of outcomes.

Elsewhere we have shown that for n>10 the expected agreement in Condorcet outcomes declines linearly with the number of vertices in M_n that are revised. Specifically, the relation is (n-k)/n, where k is the number of vertices in M_n whose edge sets have been altered (Richards, 2005b.) This finding has led to an insight regarding the relation between measures of prediction of outcomes (d') and the information content of a graphical model, to be described briefly in section 6.0.

4.0 Perturbing the Graphical Model

4.1 Directional edges for M_n (Digraphs)
The lowest curve in Fig.1 (open circles) shows the power of the shared model M_n in helping to achieve consensus: 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. We now relax the constraints imposed on M_n .
Let us continue to require that each voter's preference ordering on alternatives be fixed. However, the shared model M_n with bi-directional edges will be replaced by a new form of M_n with directed edges. (The D_i's as before will be limited to three levels as in Fig. 2.) The perturbation is equivalent to choosing edges at random from a uniform distribution of all $_nC_2$ *2 edges.
The top curve in Fig. 4 (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.
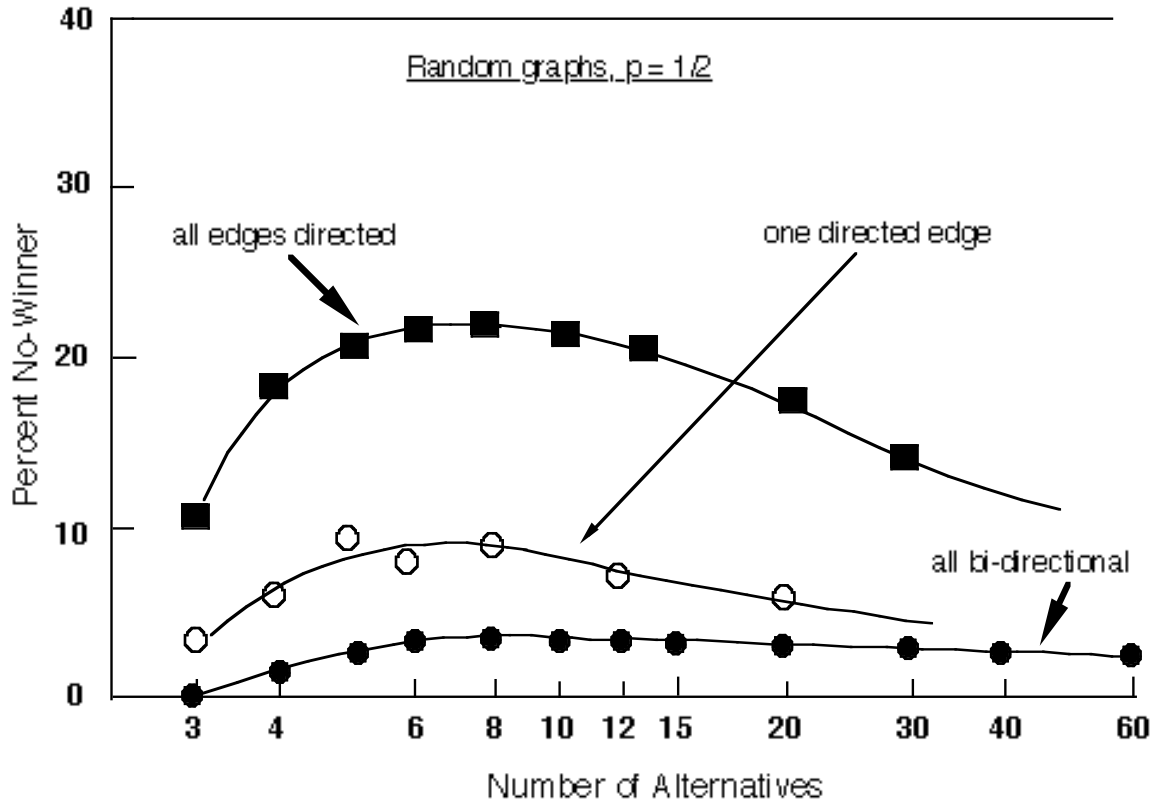

Figure 4: Probability of top-cycles (i.e. no winner) when the shared domain model is perturbed (top and middle) versus the ideal case where all preference orders respect the domain model (solid dots.)

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.

Finding #2: A random assignment of directional edges in the shared domain model M_n raises the probability of no winner by about 5-fold, as compared with

9

a random assignment of bi-directional edges. Hence consensus favors groups where voters see relationships in the domain in a reciprocal fashion.

4.2 Subgroups of Voters who violate M_n

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. 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. Obviously, as the number of haphazard votes increases, the probability of no-winner will also increase (see Figs. 1 & 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.

The solid curves in Fig. 5 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.

Finding #3: Even a small percent of haphazard votes (e.g.10%) can have severe consequences on achieving successful outcomes for choice sets larger than twelve alternatives.

We turn next to the dashed lines in Fig 5. These summarize 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
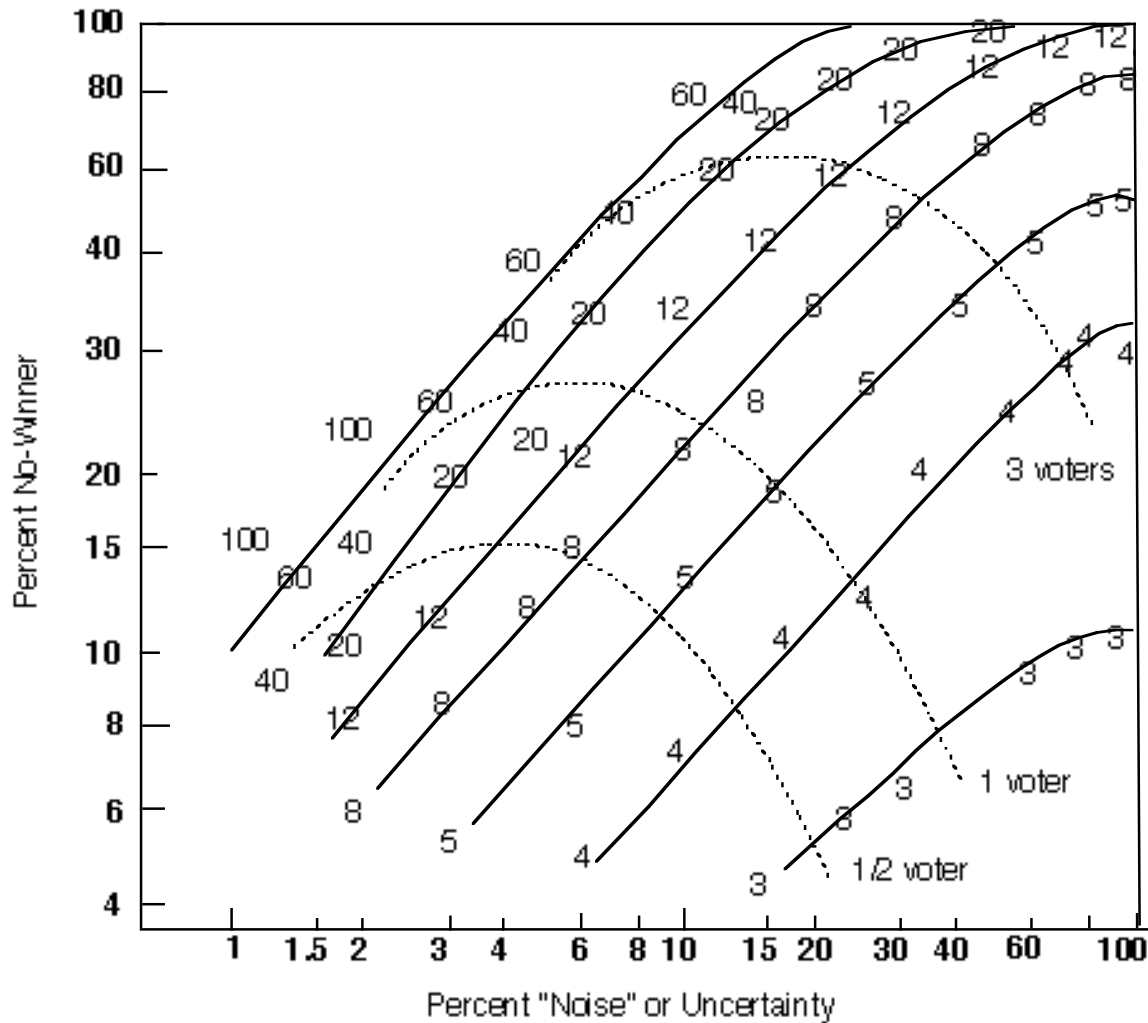

Figure 5: Solid curves: Noise distributed evenly among all agents.
Numbers indicate size of random graph

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.

Finding # 4: Regardless whether a fixed number of uncertain votes are cast as a block for one or more alternatives, or distributed over all voters, the disruption of consensus will be the same.

4.3 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
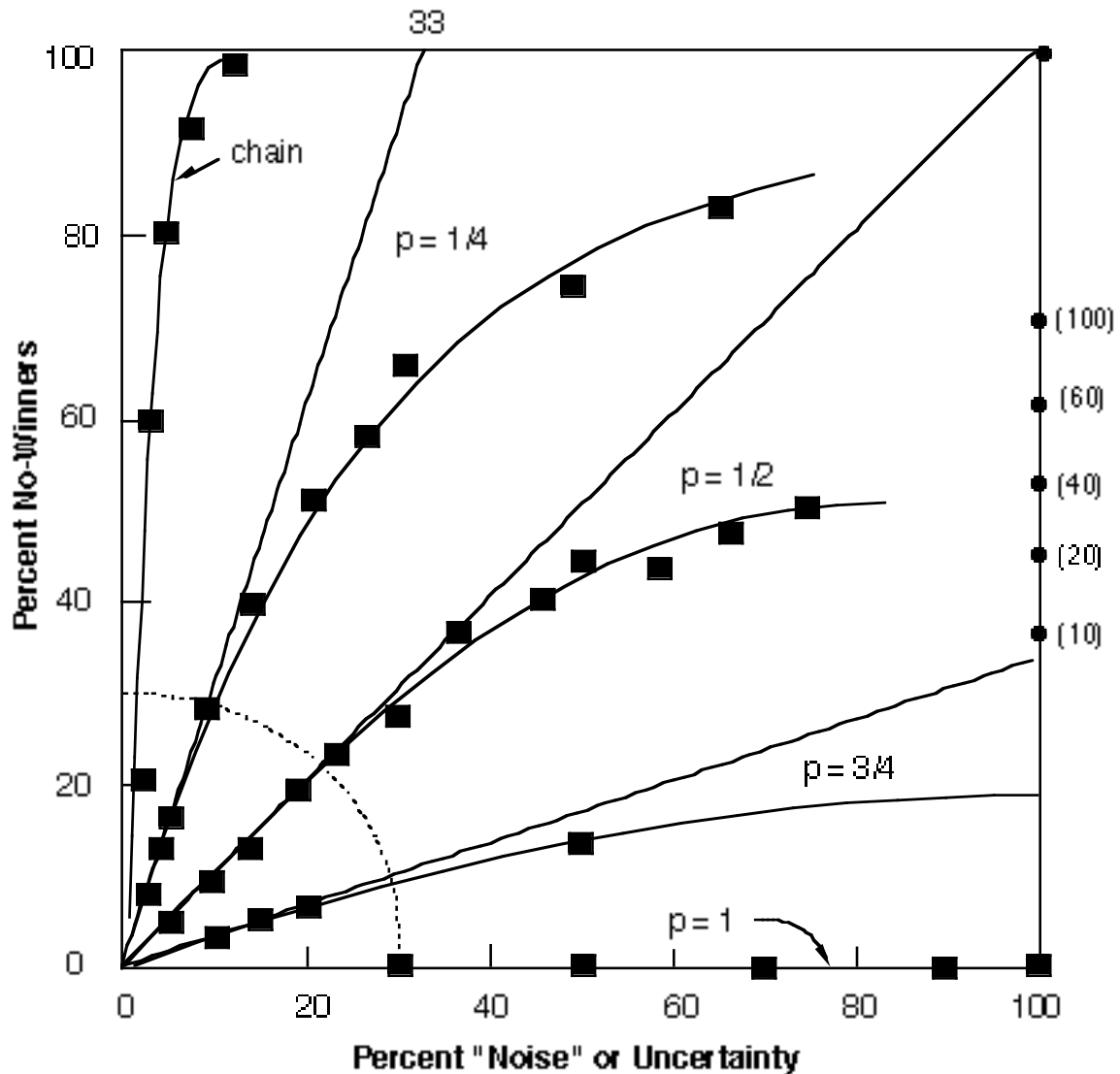

Fig. 6: Haphazard votes for alternatives not immediately similar to ideal point (i.e. non-adjacent vertices in Gn.) Data are for 40 vtx random graphs with edge probability, p, as indicated.

12

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. 6 shows the results are dramatically different from the previous case presented in section 4.2 and Fig. 5.

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) -> 1], noise is ineffective – as expected as the graphical 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.

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.

Finding # 5: The sparseness of M_n (i.e the edge probability in Gn) has very significant effects on consensus when uncertainty in voting occurs only for third or less desirable options.

Note: further related results appear in Richards, 2005a.

5.0 Information and d'
In the course of exploring similarity measures between two different graphical models, a relation between Shannon Information content (in bits) and Signal Detection measures (d') was discovered. Although such a relation has been sought since the 60's, there is no convincing proposal (Luce, 2003), excepting that of Birdsall (1955) where the criterion Beta was specified for several different, quite specific situations. Our new insight comes from the study of how one individual, with his particular graphical model, might predict the choice of winner of another individual with a different graphical model. It can be shown that the false positives and misses of one's ability to predict the other's responses are then equal, on the average. Thus, the hit rate can be translated into an information measure, without worry about the false positive rate, which is known. The theoretical framework we choose to prove this assertion is adapted from Ihler, Fisher & Willsky (2004). The key is to use the Kullbach-Leibler divergence as a dissimilarity distance between

the two graphical models. Unfortunately, the write-up with John Fisher has been delayed. The target for an MIT AI Memo is November.

Finding #6: Our paper will answer a 40 yr old question of how d' and bits can be rigorously related. This result opens the door to assigning information measures (bits) to predictions about categorical assertions, and, indirectly, may provide a measure (in bits) for the significance of cognitive events.
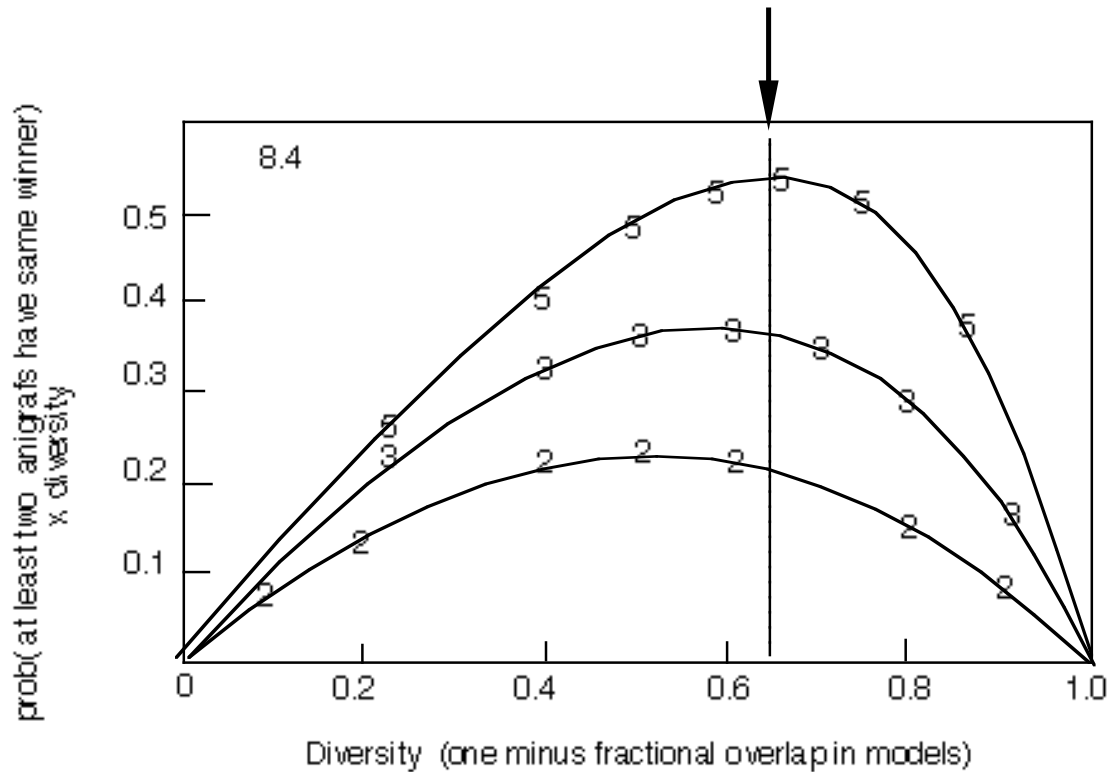

Fig. 7: Similarity between two graphical models is defined as the percent of vertices with edges in common. Diversity is the opposite: one minus similarity. The graphs show how to maximize diversity in individual cognitive structures (represented as graphical models) while still maintaining good communication between members of a group of sizes 2, 3, and 5. The optimal graphical similarity between the members' models should be one-third (i.e. diversity equals two-thirds.)

Another twist on this problem is when two parties wish to collaborate. Each party should bring to the partnership different perspectives (i.e. models), but at the same time they must interact without misunderstandings. Hence we have a trade-off between the similarity of the two models needed for understanding (d' high) and the dissimilarity of the models which provides different perspectives

(information content high.) Fig. 7 illustrates that an optimal solution is when the similarity between the graphical models is one-third, as measured by the fraction of vertices with different edge sets. (See Richards, 2005b for further details and elaborations.)

6.0 Neural Voting Machines

The Borda and Condorcet methods are maximum likelihood aggregation procedures for social decision-making under uncertainty (Young, 1986.) The Borda method is easy to implement, requiring O[n] calculations, and always will yield a consensus, excepting ties. The Condorcet method is a tournament where each alternative is contested pairwise with all others. This method has the advantage of NOT yielding a winner if there is no clear majority winner, and, furthermore, places no arbitrary weights on lower ranked preferences. The clear disadvantage is that the pairwise comparisons require $O[n^2]$ calculations.
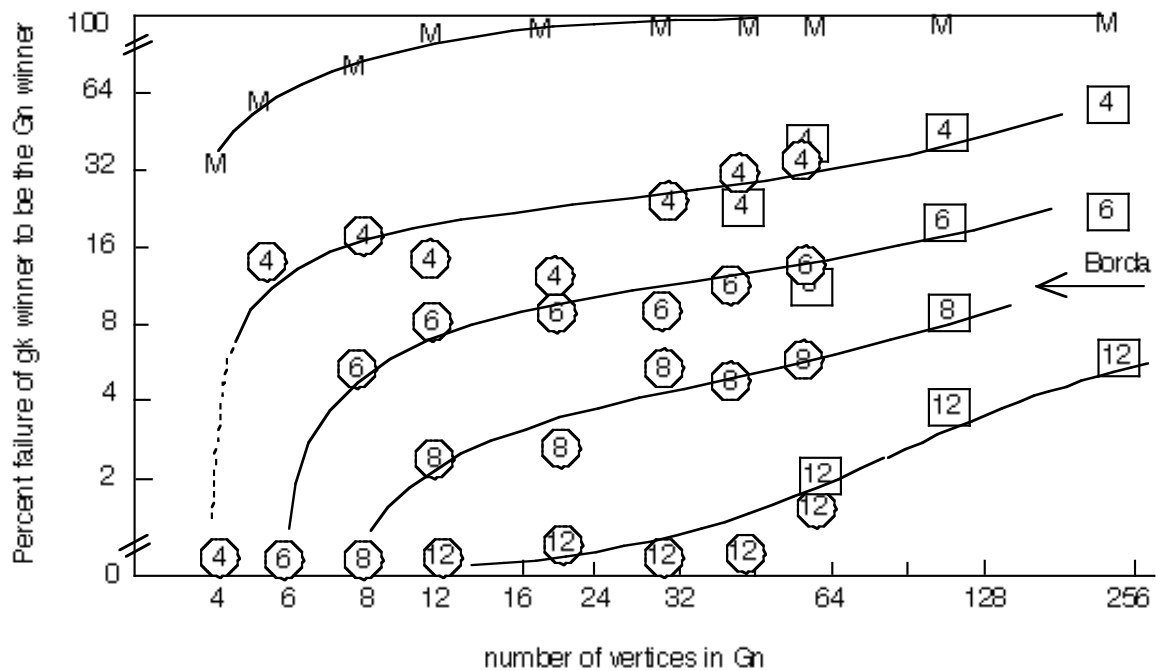

Figure 8: Winners for gk compared with winners for Gn, with the numbers along the curves indicating values for k. Over the range of n, the Borda winner matches 90% of the Condorcet winners (arrow.) The maximum weight node in Gn is rarely the winner for n > 12, as shown by the curve labelled "M".

We have designed an approximate Condorcet algorithm that requires only O[n,k] calculations, where k can be as small as 12 for 60 alternatives, or 25 for 200. The algorithm is 97% accurate, and fails by misses, not false positives. In other words, almost invariably for random graphs, no winner is delivered when the algorithm fails to produce the correct winner.

15

The trick is to find the largest Borda scores in the landscape of Borda winners for the original graph Gn (O[n]). Then take the top k Borda winners, and form the subgraph gk of Gn. These calculations are O[k | 2], which is insignificant compared with O[n | 2]. Figure 8 shows the success rate of the algorithm for k = 4, 6, 8, and 12. Also shown is the percent of Borda winners that agree with the true Condorcet winner for Gn (roughly 90%.) Note that the approximation is quite good, and very efficient.

We have also designed a simple neural network that can carry out this calculation. It has only one more layer than the Borda network. The graphical model does not appear explicitly as we would visualize a graph. Rather, the edges of the graphical model are made explicit. This is necessary in order to capture a theoretical adjacency assertion needed to compute pair-wise winners. The edge assertions can also be cast as correlations between alternatives, thus opening possibilities for weighted edges in the graphical model, or for learning new relationships (Richards & Seung, 2004.)

Finding #7: The major impact is probably in Theoretical Neuroscience. It is now clear that biological systems can indeed carry out information aggregation procedures that are much more optimal than the popular winner-take-all method.

7.0 Multimodal Dynamics: cross-modal clustering (M.H.Coen)
Graphical models that represent similarity relations among alternatives lie at the heart of our research. How are these models learned?

A new answer to this question has been the result of Michael Coen's research. Rather than attempting to categorize (or cluster) data obtained from one type of source (i.e. one sensory modality), he cross-correlates information from two different modalities, with one modality "training" the other. Thus, the method is self-supervised. A simple example is the correlation in a video stream between the sounds the speaker utters and her lip movements. The data from both sources is unlabeled; in other words the input is simply two scatter plots of points (observations.) From these data, we wish to recover meaningful phonetic categories, or in this case, the vowels.

Fig. 9 shows the result for vowels. Here we have two "slices" through a high dimensional feature space of a sound stream. Clusters emerge in each slice through an iterated process of one slice supervising (training) the aggregation
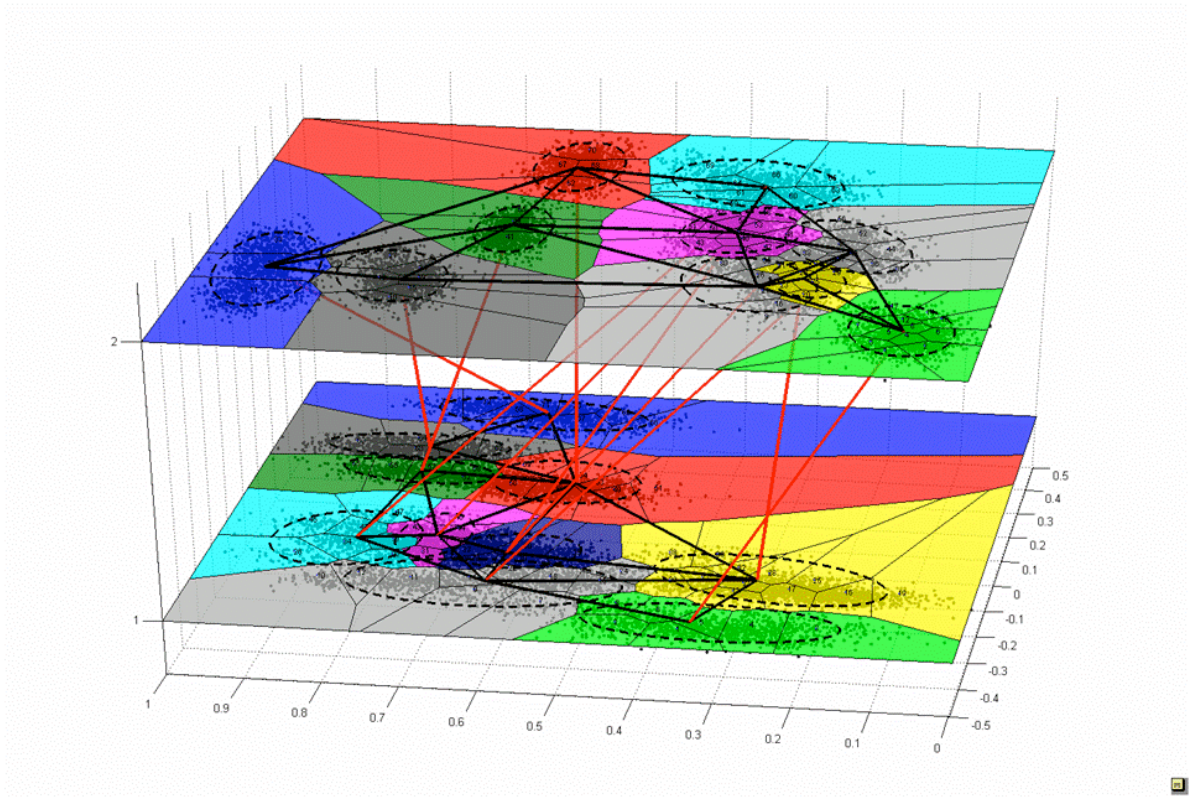
Fig. 9: The edges between the planar slices link clusters in the two sensory channels; the edges within each slice show the graphical relations within each slice. (from M. H. Coen, 2005.)

of information in the other, consistent with the evidence provided by the correlations.

When viewed as belief systems, we see in Fig. 9 that context, here represented as the different slices, may change similarity relations on the one hand, but at the same time, data from different contexts can be used to learn the basic categories of the domain.

Finding #8: First self-supervised learning algorithm using multi-modal information.

Note: Thesis available October 05. For preliminary version see Coen 2005.

8.0 Implications of Research

Our findings show that successful outcomes in collective decision-making are strongly dependent on the integrity and form of the model of the domain shared by the members of the group. Although we have modeled the belief structures of the domain as a graph, and have explored stability using Condorcet aggregation, we expect the results to generalize to other representational forms, as well as to other tally procedures including non-democratic decision-making. Regarding the latter, one might regard the beliefs of group members subject to dictatorial rule as having little impact on the stability of the group. Our results, however, suggest otherwise, especially if the coherence of individual beliefs plays a role in group stability. If so, then even a small group of members can create an environment that is potentially unstable, easily reaching a tipping-point (e.g. Fig. 7.) Our findings thus impact not only the understanding of group decision-making, but also the stability of social networks, negotiations seen as collaborations, as well as collaborations between parties with differing belief structures.

9.0 References
9.1 Supported under Project:

Coen, M. H. (2005) Cross-Modal clustering. Proc. Amer. Assoc. Art. Intell.
        (available at http://www.csail.mit.edu/~mhcoen/aaai05.pdf)
Pickard, G. & W. Richards (2006?) Optimal Borda rank vectors. [in preparation.]
Richards, W. (2005a) Collective choice with uncertain domain models.
        AI Memo 2005-024. (available from: http://www.ai.mit.edu/research/
        publications/)
Richards, W. & S. Seung (2004) Neural Voting Machines. AI Memo 2004-029.
        (available from: http://www.ai.mit.edu/research/publications/)
Richards, W. & S. Seung (2003) Neural Voting Machines, Proc. Cog. Sci. Soc.
         Boston, MA. (an early version.)
Richards, W. & J. Fisher (2005?) Information gain associated with different
         observer models [in preparation.]

9.2 Other citations:
Arrow, K.J. (1963) Social choice and Individual values. Wiley, NY.
Birdsall, T. (1955) The theory of signal detectability. Pp 391-401 In:
        Information Theory in Psychology, Henry Quastler, ed.; Free Press,
        Glencoe, Ill.

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
Condorcet, Marquis de (1785) Essai sur l'application de l'analyse a la
       probabilite des decisions rendue a la pluralite des voix, Paris (See
       Arrow, 1963).
Ihler, A; J. W. Fisher; A. S. Willsky (2004) Non-parametric hypothesis tests
       for statistical dependency. IEEE Trans. Signal Proc., Special Issue on
       Machine Learning Methods.
Luce, D. (2003) What ever happened to information theory in psychology?
       Rev. General Psych. 7, 183-188.
Pearl, J. (1988) Probabilistic Reasoning in Intelligent Systems. Morgan
       Kaufman.
Richards, W. (2005b) Anigrafs (book in progress) www.ai.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. Jrl Math. Psychol. 46, 338-
       351.
Runkel, P.J. (1956) Cognitive similarity in facilitating communication.
       Sociometry, 19, 178-91.
Saari, D. G.: 1991, 'Erratic Behavior in Economic Models', Journal of
       Economic Behavior and Organization 16, 3-35.
Saari, D. G.: 1994, Geometry of Voting, Springer-Verlag, Berlin.
Schelling, T (1971) Dynamic Models of Segregation. Jrl Math. Sociol. 1, 143-
       186.
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
Young, H.P. (1998) Individual Strategy and Social Structure, Princeton.