

# Collective Choice and Mutual Knowledge Structures<sup>†</sup>

**Diana Richards**  
Department of Political Science  
University of Minnesota  
267 19th Avenue South  
Minneapolis, MN 55455, USA

**Whitman A. Richards**  
Media Arts and Sciences and  
Department of Brain and Cognitive  
Science  
Massachusetts Institute of  
Technology  
E10-120  
Cambridge, MA 02139, USA

**Brendan D. McKay**  
Department of Computer  
Science  
Australian National University  
Canberra, ACT, 0200,  
Australia

---

**ABSTRACT.** An important puzzle in the study of complex systems is the conditions under which the aggregation of information from interacting agents results in a stable or an unstable collective outcome. We present a general framework for thinking about the stability and instability of collective outcomes that focuses on the effects of mutual knowledge. We show that if a complex system of aggregated choice respects a mutual knowledge structure, then the prospects of a stable collective outcome are considerably improved. Our domain-independent results apply to collective choice ranging from perception, where an interpretation of sense data is made by a collection of perceptual modules, to social choice, where a group decision is made from a set of preferences held by individuals.

**KEYWORDS:** Complex system, stability, collective choice, knowledge structures, cognition

---

## 1. Introduction

In the study of complex systems, defined as collections of interacting entities or agents, one often seeks to understand the behavior of the whole from the behavior of its constituents. When a complex system involves the aggregation of orderings, we label this as the problem of collective choice, where information about orderings of alternatives held by individual entities is aggregated to a collective outcome. Unfortunately, the aggregation of elements into a collective choice is not always straightforward. For example, Arrow (1963) showed that no voting procedure exists for passing from any set of individual preferences to a social choice that satisfies a set of

minimally desirable conditions. Similarly, Doyle and Wellman (1989) used a version of Arrow's Theorem to show that rational reasoning in distributed intelligent systems is similarly constrained. Aggregation processes are no longer seen as simple sums of the components, but are understood as having the potential for unstable, arbitrary, intransitive, and chaotic behavior (Schofield 1980, Saari 1987, 1991, D. Richards 1994, W. Richards et al., 1993).

Yet in all collective behavior, the ability of constituent elements to yield a relevant stable outcome is an essential aspect of "reasonable" or "intelligent" behavior. Clearly, if collective outcomes have little relation or a chaotic relation to the actions and desires of its elemental constituents, then it violates our intuitive understanding of a rational collective choice. An important assumption in the social choice setting is that a collection of individual preferences can be aggregated to a stable group choice that properly reflects the consensus of the group. A similar assumption is made in the modeling of our everyday perceptions where we desire that the explanations of our sensory observations--i.e., our percepts--are consistent with the biases and beliefs held by the constituent neural information-processing modules (Marr 1982). Here again, in the presence of competing beliefs, we strive for a stable percept, in the sense that competing explanations for the data will not overturn the original percept (as it does when we experience perceptual multistabilities.) Although there are many (Bayesian) techniques for assuring that some "maximum likelihood" conclusion can be reached in a probabilistic framework (e.g., see Clark and Yuille 1990), among these approaches, ours most closely resembles Pearl's (1988), which stresses how causal graphs provide constraints that lead to tractable probabilistic inference under uncertainty.

Thus, an important puzzle in the study of complex systems is to understand conditions under which the collective behavior of communicating or interacting entities can reach a consensus or stable state. This question is so fundamental that it appears throughout the disciplines and at all levels of analysis. To illustrate the universality of this puzzle, consider the following questions from various scientific fields:

- What kinds of bonding relations allow molecules to form stable or unstable configurations, perhaps as in the case of protein folding and its pathologies?
- What kinds of structural interactions among species allow them to form a stable or a transitional biome?
- How do the modules of the brain aggregate data to create a rational collective perception or behavioral response? Under what conditions will sensory input lead to conflicting percepts?
- When do groups of individuals reach a common interpretation of a mutually observed event and when do their differing interpretations fail to be reconciled?

- Under what conditions are collective agreements among individuals, committees, or nation-states achieved and under what conditions does negotiation or voting fail to reach a stable collective compromise?

The new contribution of our approach is to consider the conjunction of complex decision making systems and mutual knowledge structures. By mutual knowledge we mean that all entities hold the same meta-information about the relevant characteristics or relationships among the choice set.<sup>1</sup> In particular, we assume that the empirical structure of the context imposes strong constraints on contextually-feasible preferences held by constituent entities. Therefore, a key assumption of our model is that preference orderings must respect the empirical relationships among choice elements. Note that this is an important change from many existing approaches, which assume that agent-held information (such as preferences) are the primitive building-blocks. In our conceptualization, agent-held information is not primitive but is embedded within a knowledge structure.

Such knowledge structures and relationships are present in all choice contexts, whether physical, social, or cognitive. Their origin lies in the laws and regularities that bring order into our world (Thompson 1968, McMahon 1976, Jepson et al 1995). To illustrate, consider the frequency of vocal sounds made by animals: large animals make low pitch sounds because they have large vocal tracts, whereas small animals with smaller vocal cavities will make higher pitched sounds. If we hear a sound made by an unseen animal in the forest, we can guess the size, and hence the category of the animal. We all share and use this kind of intrinsic knowledge in order to make rational perceptual inferences from sense data. Second, at a more cognitive level, stories, like other forms of linguistic communication, must have a known intrinsic structure in order for the meaning to be understood by the audience (Campbell 1949). To be more explicit, consider types of stories as reflected in categories of films. Film categories have a natural ordering from "light cognitive" such as romantic comedies to "heavy physical" such as martial arts. People who prefer light cognitive films will typically avoid violent films, and vice versa. However, both groups of people may accept documentaries. Video stores typically use these dimensions to organize store layouts, placing the more neutral category of documentaries between the extreme categories.<sup>2</sup> Finally, our daily social interactions are also very regularized and lawful, following certain traditions and conventions. In the U.S., we drive on the right side of the road, with the steering wheel on the left. In Japan and Britain, it is the opposite. These conventions dictate the placement of traffic signals, signs, and which way we look first before crossing the street. Such strong correlations at all levels-- perceptual, cognitive, and social--impose an enormous amount of structure on our thoughts and behaviors, and affect our ways of holding and sharing knowledge.

The main result of this article is to capitalize on shared knowledge structures to account for the empirical observation that many collective choice contexts do in fact reach a stable collective outcome. Consider the two areas on which we focus: cognitive science and political science. The majority of everyday perceptions are correct and stable: if humans were continually operating in a world of unstable perceptual illusions they would be incredibly maladapted at surviving in their environment. Yet it is still a puzzle how the various "agents in our minds"--each embodied in one of the many different modules of our brain--aggregate their collective expertise to reach conclusions that are consistent with world behaviors (Fodor 1983, Minsky 1986, Van Essen 1992). Here we show that one answer lies in the agents sharing mutual knowledge. More strongly, we would claim that without a modicum of shared knowledge, collections of neural agents will suffer instabilities. Similarly, social decisions, whether in a family or a governmental entity, are typically achieved in practice. A salient puzzle in the field of social choice is to delimit conditions under which collective agreements are swift and stable, and conditions under which human societies fail to coordinate to a stable collective choice. Again, we propose that mutual knowledge serves as a stabilizing force.

Contrary to previous negative theoretical results about the aggregation of information in complex systems, our results are positive in that a collective equilibrium is guaranteed in many cases for all sets of feasible agent preferences. Yet our results do not suggest stability is always forthcoming. Rather, specific conditions are outlined where instability is predicted.

The goal of this article is to conceptualize the problem of stability and instability in complex choice systems using a very general framework (based on graphs and partial orders) that can apply across fields. To show the generality of our approach, we consider two quite different types of collective choice from the fields of cognitive science and political science respectively: perceptual choice: the choice of an interpretation for our sense data made by a collection of perceptual modules, and social choice: the choice of a group decision from a set of preferences held by individuals. Because it is the organizational form or "structure" of knowledge--not its specific content--that is sufficient to induce an equilibrium in collective outcomes, our approach applies to many domains where the collective choice process incorporates empirical relationships. We encourage others to think about how our results and framework help to reconceptualize the question of stability and instability in broader applications of complex system theory.

## 2. Complex Systems and World Structure

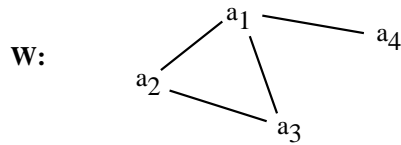
Collective choice, broadly defined, is any situation where a collection of agents must aggregate information in order to reach a single collective decision. As we have already pointed out, the collection of agents can be as diverse as voters that must reach a group decision to neural

modules within a biological system that must reach a collective consensus as to the next behavioral act. We begin with a set of agents who choose over a finite set of alternatives, denoted  $A = \{a_1, \dots, a_n\}$ , where  $A$  contains at least two elements. Agents and alternatives exist in a perceived or empirical "world" governed by a shared knowledge structure, represented by a labeled graph  $W(A, e)$  with a set of vertices  $A$  and a set of edges  $e$ . A vertex may have one or more edges, but it is assumed that  $W$  is connected. Each edge  $e = \{a_i, a_j\}$  of  $W$  linking  $a_i$  and  $a_j$  corresponds to a change of one parameter between alternatives  $a_i$  and  $a_j$ . For example, Figure 1 shows a graph  $W$  representing parameter changes between four alternatives,  $a_1, \dots, a_4$ . Each edge in  $W$  connecting two alternatives indicates that those alternatives differ in a single attribute. For example, if the alternatives are choices among soft drinks,  $a_1$  and  $a_4$  may be two brands that have similar taste but differ in that one is caffeinated. Alternatives  $a_1$  and  $a_2$  may both have no caffeine but differ in their citrus blend, such as comparing Seven-Up to Sprite.

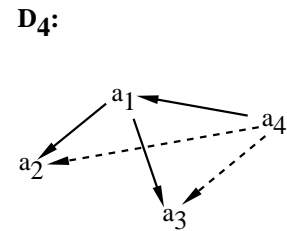
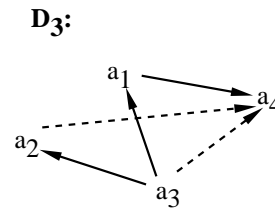
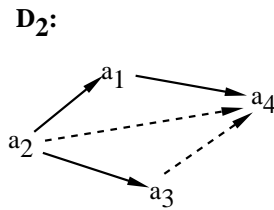
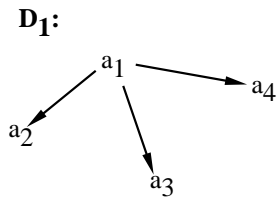
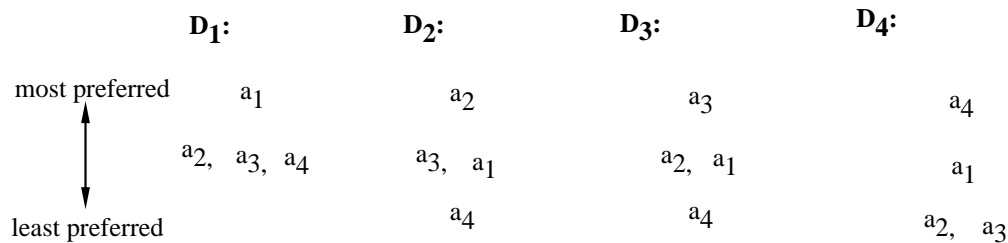
A preference ordering is modeled as a set of transitive paths through a choice space, where a path cannot "jump" from any alternative to any other, but must respect relationships among alternatives consistent with "world structure"  $W$ . An ordering which jumps around, violating the relationships among alternatives in  $W$ , is not allowed in our model.<sup>3</sup> Hence, preferences are defined over  $W(A, e)$  and incorporate not only the set of alternatives  $A$  but also the relationships among the alternatives. We assume that each agent has a unique most-preferred alternative, called an ideal point. Thus, for every  $a_j \in A$ , there may be some agents who have  $a_j$  as their ideal point. Let  $D_j$  be the partial order induced from  $W$  beginning at ideal point  $a_j$ . Thus,  $D_j = (A, P)$  is an asymmetric and transitive directed graph where  $(a_j, a_k) \in P$  iff the number of edges on the shortest path from  $a_j$  to  $a_j$  is less than the number of edges on the shortest path from  $a_j$  to  $a_k$ . If  $(a_j, a_k) \in P$  then we say that  $a_j$  is preferred to  $a_k$ , denoted  $a_j \preceq a_k$ . If the number of edges on the shortest path from  $a_j$  to  $a_j$  is equal to the number of edges on the shortest path from  $a_j$  to  $a_k$  then we say that  $a_j$  and  $a_k$  are noncomparable for those agents with ideal point  $a_j$ . Figure 1 shows the set of four directed graphs induced from the graph  $W$  and then the corresponding sets of feasible preferences over the set of alternatives.<sup>4</sup>

It is a central assumption of our framework that agents' preferences are consistent with their representation of the relationships among alternatives. Given a most-preferred category, the ordering over the remaining alternatives in the choice set must be consistent with the extent of differences among alternatives, which is captured by the graph  $W$ . Therefore, preferences follow directly from the graph  $W$ . Note that these relationships need not be simple one- or two-dimensional constraints, but can involve complex graph structures.<sup>5</sup> Since the number of parameters can be as large as the number of edges in  $W$ , the form of  $W$  is quite flexible, including "rings", low or high dimension graphs, or highly connected or sparsely connected graphs. The case of linear orders (e.g., Black 1958) is a special case of the set of all configurations.

FIGURE 1.

**D<sub>i</sub>: Set of partial orders induced from W:**

(no arc means alternatives are indifferent, dashed lines are implied by transitivity over preferred or indifferent alternatives)

**Feasible preference orders from W:**

Let  $\mathbf{w} = \{w_1, \dots, w_n\}$  be the normalized weights over the  $n$  preference types, i.e.,  $w_i$  is the proportion of agents with ideal point  $a_i$  and thus the proportion of agents with the partial order  $D_i$  over the set of alternatives  $A$ . Our approach is to check for the possibility of a stable collective outcome using pairwise comparisons between the alternatives, which corresponds to pairwise plurality rule (Saari 1994). In the perceptual case, this is roughly equivalent to the modules forming cliques with shared preferred interpretations of the observations, and then "voting" to determine if the interpretation favored by one clique will beat all others. Although different aggregation rules could be chosen, a pairwise comparison is one of the most stringent. In addition, we assume that the preference orderings of the agents or perceptual modules are aggregated "sincerely", specifically that the weights are based solely on the partial orders of an agent's preferences. <sup>6</sup>

A top cycle exists if there is some set of alternatives such that  $a_i$  beats  $a_j$ ,  $a_j$  beats  $a_k$ , and  $a_k$  beats  $a_i$  and there exists no other alternative that beats all the alternatives in the cycle. If a top cycle is impossible, then for any set of weights  $\mathbf{w}$  there is a top-ranked alternative. (Note that there may be additional cycles that are not top cycles; these do not concern us since we are interested in the existence of a top-ranked alternative.) For example, in the stable aggregation of information from several perceptual modules,  $I_1$  would be an unstable majority consensus if there was another interpretation,  $I_2$  that was favored over  $I_1$  by another majority clique, and  $I_2$  was beaten by a third clique favoring  $I_3$ , which in turn would be beaten by the original  $I_1$  interpretation. Such top cycles can be easily created in laboratory situations where perceptual information is controlled to be ambiguous, breaking the natural regularities and ordering relations normally encountered in the world (Gregory 1970, Marroquin cited in Marr (1982), Rock 1983).

We define an equilibrium as an alternative that cannot be overturned in a pairwise vote by any other alternative in the choice set. Let  $|a_i \preceq a_j|$  denote the number of agents for whom  $a_i$  is preferred to  $a_j$ . Then the following defines an equilibrium concept in our framework:

Definition. For a structure  $W$  with preferences  $D_i$ , an alternative  $a_i \in A$  is a knowledge-induced equilibrium if and only if for all  $a_j \in A$ ,  $a_j \neq a_i$ ,  $|a_i \preceq a_j| > |a_j \preceq a_i|$ .

We say that a configuration of alternatives is stable if there exists a knowledge-induced equilibrium for all possible weights of preferences. An alternative  $a_j$  can be reached from alternative  $a_i$  if it is possible to move from  $a_i$  to  $a_j$  by a sequence of pairwise decisions. If an alternative  $a_i$  can be reached from itself, then a cycle exists. Clearly if there are no cycles then an equilibrium exists.<sup>7</sup> A collective outcome is robust (also called structurally stable in the social choice literature) if it remains as an equilibrium after arbitrarily small changes in individual-level data. In this paper we focus exclusively on enumerating stable configurations and we leave the question of robustness to future research.<sup>8</sup>

### 3. Results

In its simplest form, our framework is a mapping from a "world structure" to a collective ordering over the set of alternatives in that "world." Thus, we are determining the set of world structures, represented by the number of distinct graphs  $W$ , for which an equilibrium exists for all possible distributions over orderings.

However, the combinatorial complexity of the problem increases very rapidly in the number of alternatives. Therefore, the analysis must rely on a combination of formal proofs and

numerical techniques to examine the extent of stability. The combination of the two approaches also serves to corroborate our results.

We use four lemmas for the results that follow. The first lemma uses the fact that the graph representation of the collective ordering satisfies the definition of a tournament. More specifically, recall that the combination of the set of feasible preferences orderings,  $D_i$ ,  $i = 1 \dots n$ , (as in Figure 1) and a distribution over these preference orderings  $\mathbf{w}$  generates a pairwise ordering between each pair of alternatives in  $W$ . Let  $\preceq_s$  denote a collective preference relation where  $a_i \preceq_s a_j$  iff the proportion of agents for which  $a_i \preceq a_j$  is greater than the proportion of agents for  $a_j \preceq a_i$ . (We can ignore the case of ties, where  $|a_i \preceq a_j| = |a_j \preceq a_i|$ .) The pairwise comparisons result in a digraph  $S$  where all alternatives in  $S$  are compared and either  $a_i$  beats  $a_j$  or  $a_j$  beats  $a_i$ . Hence  $S$  is a tournament where an arc from  $a_i$  to  $a_j$  implies  $a_i \preceq_s a_j$ . The following lemma establishes that when considering the possibility of cycles among alternatives in a tournament, it is sufficient to examine three-cycles rather than cycles of all lengths. This lemma is important to reduce the numerical task of verifying stability.

Lemma 1: If the tournament  $S$  has a top cycle of any length, then  $S$  has a top three-cycle.

Proof. Restrict the tournament  $S$  to the subgraph of the tournament induced by the vertices that are in the top cycle. Then the tournament subgraph must have a top three-cycle (Moon, 1968, p. 11).

The following two lemmas are used to generate the inductive results examining top-cycles in graphs up to ten vertices (subsequent Fig. 3). Suppose  $W$  is a graph with vertices  $a_1, a_2, \dots, a_n$ , carrying weights  $w_1, w_2, \dots, w_n$  respectively. For each  $a_i$ , define  $\Sigma(a_i)$  to be  $w_i$  plus the total weight of all the neighbors of  $a_i$  and  $\Sigma^+(a_i)$  to be  $\Sigma(a_i) + w_i$ . An induced subgraph  $H$  is a subgraph that is induced by some subset of the vertices of  $W$ . We say that a subgraph  $H$  is covered if there is a vertex  $v$  in  $W$  that is not in  $H$  such that every vertex of the subgraph is adjacent to  $v$ .

Lemma 2: Suppose  $W$  has diameter 2. If  $a_i$  is adjacent to  $a_j$  then  $a_i \preceq_s a_j$  iff  $\Sigma^+(a_i) > \Sigma^+(a_j)$ ; if  $a_i$  is not adjacent to  $a_j$  then  $a_i \preceq_s a_j$  iff  $\Sigma(a_i) > \Sigma(a_j)$ .

Proof. Follows from the definition of  $\preceq_s$ .

Lemma 3: Let  $G_1$  and  $G_2$  be two graphs of diameter 2 in which  $H$  is an induced subgraph. Suppose that for each vertex  $v_2$  in  $G_2 - H$ , there is a vertex  $v_1$  in  $G_1 - H$  such that the

neighborhood of  $v_2$  in  $H$  is a subset of the neighborhood of  $v_1$  in  $H$ . Then, if  $G_1$  has a top-cycle with zero weight outside  $H$ , so does  $G_2$ .

Proof. Give  $G_2$  the same weights as  $G_1$  in  $H$  and give zero weights to vertices outside of  $H$ .

Since  $G_1$  and  $G_2$  have the same weights inside  $H$  and zero weights outside  $H$ ,  $\Sigma(v)$  and  $\Sigma(v)$  are the same in both  $G_1$  and  $G_2$ , and similarly  $\Sigma^+(v)$  and  $\Sigma^+(v)$  are the same in both  $G_1$  and  $G_2$ , for all vertices in  $H$ . So the top-cycle in  $G_1$  is also a cycle in  $G_2$ . If the cycle in  $G_2$  was not a top-cycle, there would be a vertex  $v_2$  in  $G_2 - H$  which dominates the cycle. But then the  $v_1$  given by the lemma would dominate the cycle in  $G_1$  which is impossible since it is a top-cycle.

As will be seen below, two important subgraphs with top-cycles are the five-ring (a "pentagon") and the five-ring with one chord (a "house"). Lemma 3 implies that a graph with an induced uncovered house has a top-cycle provided it has diameter 2. Furthermore, a graph with an induced uncovered pentagon has a top-cycle if it has diameter 2 and there is some vertex  $v$  on  $P$  such that  $P - v$  is not covered. Another important graph is the special case of rings with  $n \geq 5$ . The following lemma shows that rings with five or more alternatives cannot guarantee an equilibrium.

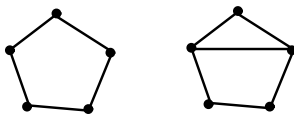
Lemma 4. All rings,  $n \geq 5$ , have a top cycle.

Proof. Label the graph  $R_n$  counterclockwise with weights  $w_1, w_2, \dots, w_n$ . For even values of  $n \geq 6$ , the weights  $\mathbf{w} = (2, 1, 5, 0, \dots, 0, 5, 0, \dots, 0)$  with  $w_{\frac{n}{2}+2} = 5$  result in the top-cycle  $a_3 \text{ } \text{f}_s \text{ } a_2 \text{ } \text{f}_s \text{ } a_n \text{ } \text{f}_s \text{ } a_3$  with  $a_3 \text{ } \text{f}_s \text{ } a_1$  and  $a_3 \text{ } \text{f}_s \text{ } a_4, \dots, a_{n-1}$ . For odd values of  $n \geq 5$ , the weights  $\mathbf{w} = (5, 2, 1, 0, \dots, 0, 5, 0, \dots, 0)$  with  $w_{\frac{n+1}{2}+1} = 5$  result in the top-cycle  $a_1 \text{ } \text{f}_s \text{ } a_2 \text{ } \text{f}_s \text{ } a_3 \text{ } \text{f}_s \text{ } a_1$  with  $a_1 \text{ } \text{f}_s \text{ } a_4, \dots, a_n$ .

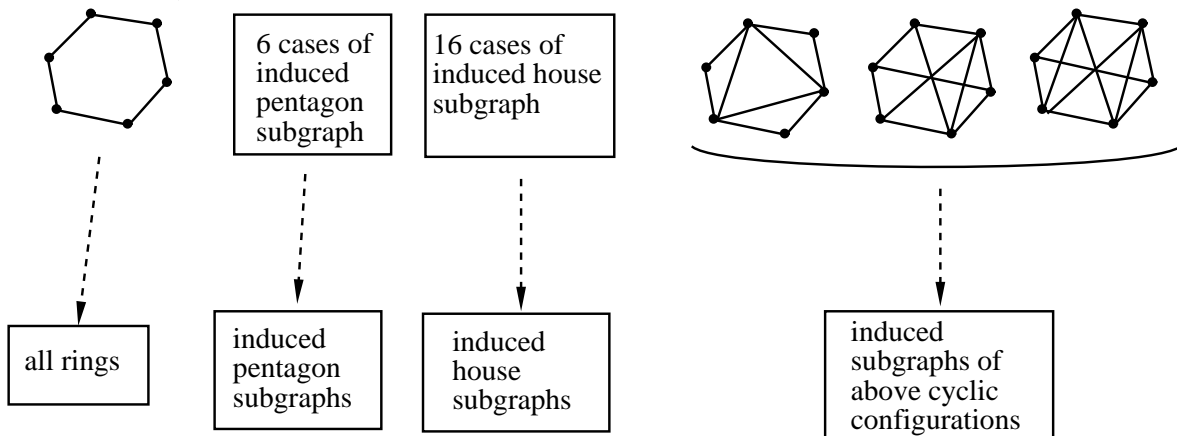
**FIGURE 2. Set of primitive unstable structures**

**n = 3 or 4: all structures stable**

**n = 5:  
2 unstable out of 21  
structures:**



**n = 6:  
26 unstable out of  
112 structures:**

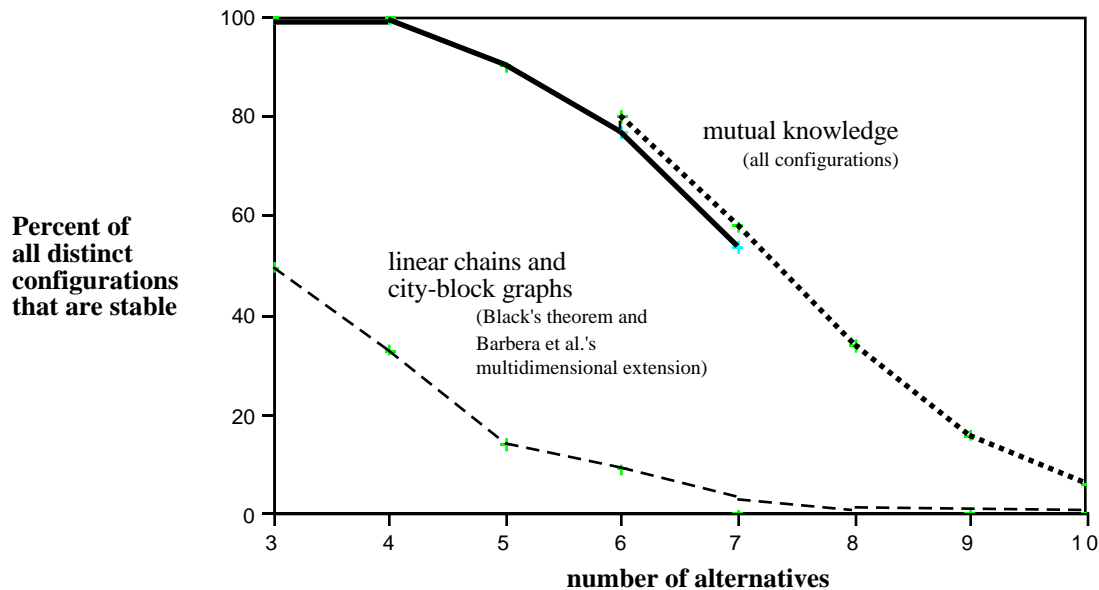


The following observations summarize the effect of mutual knowledge structures. These results were obtained by a combination of computation (McKay 1990) and induction from the lemmas. (In addition, results for  $n = 1$  to 7 were verified by a Mathematica program that checked all distinct graphs for the possibility of top cycles (solid line in Fig. 3).):

- An equilibrium exists for all preference distributions over structures with four or fewer alternatives.
- In the case of five alternatives, only two structures (of a total of 21) fail to guarantee an equilibrium for all sets of preference distributions: the five-ring and the special case of a five-ring with one chord (a "house").<sup>9</sup>
- In the case of six alternatives, 26 out of 112 structures have the possibility of cycles, or equivalently, 77% of all knowledge structures over six alternatives guarantee a stable collective outcome for all sets of agent weightings. Twenty-two of these cases are graphs that contain either an induced pentagon subgraph or an induced "house" subgraph. The remaining four are the six-ring and three special symmetries (see Fig. 2).
- In the case of seven alternatives, 397 out of 853 structures had the possibility of cycles; hence just over 50% of the seven-vertex structures are stable for all agent weights. All of the structures with top-cycles had induced six-vertex subgraphs with top-cycles.
- For more than seven alternatives, using Lemma 3 we can estimate the number of connected graphs that will not have top cycles (dashed line in Fig. 3).

Figure 3 summarizes the dramatic effect of mutual knowledge structures. When no knowledge structure constraint is present, the baseline is at 0% stable for three or more alternatives because there is always at least one set of ordering weights for which cycles exist (Arrow 1963). However, Black (1958) showed that cycles are precluded in the case of single-peaked preferences over a one-dimensional set of alternatives and Barberà et al (1993) extended this result to show that cycles are also precluded using a median voting rule if preferences are multidimensional single-peaked over alternatives ordered in a city-block. Figure 3 shows the improvement in stability over previous results. The effect of mutual knowledge is a dramatic improvement in the stability of collective choice, particularly for smaller sets of alternatives.

**FIGURE 3. Percent of configurations that are stable under mutual knowledge (compared with no constraints\* and linear or city-block constraints)**



\*With no constraints, the number of distinct configurations that are stable for all preference weights is zero (Arrow's theorem).

#### 4. Discussion and extensions

This article presented a general framework for thinking about the effect of mutual knowledge structures on collective choice aggregation in complex systems. We have shown that if there is a mutual knowledge structure among alternatives and if elemental orderings respect this structure, then collective stability is much more common than predicted based on previous frameworks. Our general framework dovetails nicely with existing convergence results. For example, well-known restrictions of preferences that yield convergence, such as Black's single-peakedness condition (1958) or Barberà et al.'s "city-block" condition (1993), are special cases of a large set of graphs that can result in stable collective outcomes for reasonable aggregation rules. Hence these two earlier stability results are the tip of the iceberg of the set of all preference restrictions that induce stability. Our general framework allows for a complete cataloging of all sets of structures that guarantee stability under all possible weights.

Our approach entails many implications of interest to a variety of settings. First, important theoretical ramifications emerge about the behavior of complex systems simply by bringing to the

forefront the role of how knowledge or information is shared among agents. In our conception of the problem, alternatives in any collective choice setting do not "float freely" in some generic choice space but are highly structured vis-à-vis each other through knowledge systems. The presence of a knowledge system, we assert, is a basic component of all social and biological choice, in that agents, whether individuals or neural modules, are embedded in a context that includes physical laws, social conventions, or cognitive groupings and classifications.

The implication of this construction is that "elemental orderings"--or more generally, how agents treat the information they receive from the environment--are not the primitive concept but emerge in conjunction with a knowledge system over the set of alternatives. Our framework places preferences and knowledge structures as coupled concepts, where preferences only make sense as following in a consistent manner from the categories and relationships used to structure a choice environment. However, agents still may have conflicting interests within their knowledge structure. Furthermore, by bringing knowledge structures to the forefront, our construction differs from traditional connectionist approaches that are based on local and statistical information. Instead, the focus shifts to the form of the meta-information and the belief structure that is shared among agents.

Hence, in our framework, it is the form of the mutual relationships, not the specific content, that plays the key role in achieving stability of outcomes. This implies that the highly abstract framework we present here can provide insight into the stability of collective aggregations in many contexts, ranging from how individuals reach collective decisions, to how neural conclusions are reached in perception, to the stability of ecological biomes, and even to the stability of biochemical aggregations such as proteins. Since mutual knowledge consists simply of informational components, it can be conceptualized as "hardwired" (as in cells or neural modules), or as energy bonds, or as subjective schemas or models (as in economic theories, scientific consensus, or epistemic communities).

However, not all complex systems will result in a stable collective outcome. Our results point to specific knowledge configurations that fail to guarantee an equilibrium. Ultimately, these predictions can be subject to empirical or experimental verification. For example, Hutchins (1995, 241) details an account of a perceptual error in navigation that occurred on the Chesapeake Bay where an oncoming vessel was falsely interpreted as a boat traveling in the same direction which was being overtaken. Surprisingly, the network configuration of hypotheses that Hutchins presents for this actual example of perceptual instability corresponds to the eleven-edge configuration for  $n = 6$  shown in Figure 2! Such instabilities obviously can go beyond the actual network, however. For example, if mutual knowledge fails, i.e., if there is a problem in the transfer of information across entities or if entities are structured with different informational meta-schemes, then one should expect problems in the stability of the collective aggregation. Typically

mental models of complicated processes (such as nuclear power stations) may fail under stress because such links are broken (Moray 1990). Third, the characteristics of stable and potentially unstable information structures help account for why seemingly small changes can result in very large effects on aggregate outcomes. The simple addition or subtraction of an "edge connection" in our framework can be the difference between guaranteed stability and the potential for instability. The exploration of such "keystone" elements and links, applicable both to biology and to social decision making, is an exciting prospect.

The general framework we present can and should be extended in several ways. First, relationships among alternatives can be represented in multiple formats, such as through functional relationships, set partitions, and directional cause-and-effect paths. The graphical approach used here easily extends to these different forms of representing knowledge systems. Second, relationships among alternatives may be hierarchical; this is also a straightforward theoretical extension (although not computationally easy!). Third, knowledge is often held by specialized "experts" and distributed differently among different elements. The extent and accuracy with which this knowledge is shared is an enticing issue, because presumably it will affect the stability of the underlying structure. Fourth, the question of the degree to which a mutual knowledge structure emerges through the dynamics of interaction among agents remains a large unexplored topic, suggesting further work examining knowledge structures emerging as dynamic equilibria. Fifth, although we have demonstrated how relationships among categories of alternatives can induce stable outcomes, we have not demonstrated the empirical frequency of stability-inducing graph structures. However, the large number of graph structures that do result in stable outcomes suggests great promise and provides better insight as to why in practice most collective decision making is stable and "rational"--contrary to past theoretical expectations.

## 5. References

- Arrow, K. J. 1963. Social Choice and Individual Values. New Haven: Yale University Press.
- Barberà, S., Stacchetti, E. and Gul, F. 1993. Generalized Median Voter Schemes and Committees, Journal of Economic Theory 61, 262-89.
- Black, D. 1958. The Theory of Committees and Elections. London: Cambridge University Press.
- Campbell, J. 1971. The Hero with a Thousand Faces. Princeton: Princeton University Press.
- Clark, J. J., and Yuille, A. L. 1990. Data Fusion for Sensory Information Processing Systems. New York: Kluwer.
- Doyle, J. and Wellman, M.P. 1989. Impediments to Universal Preference-Based Default Theories, Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning, San Francisco: Morgan Kaufmann, pp. 94-102.
- Fodor, J.A. 1983. The Modularity of Mind. Cambridge: M.I.T. Press.
- Gregory, R. 1970. The Intelligent Eye. New York: McGraw Hill.
- Harary, F. H. 1968. Graph Theory. Reading, Mass.: Addison-Wesley.
- Hutchins, E. 1995. Cognition in the Wild. Cambridge: M.I.T. Press.
- Jepson, A., Richards, W. and Knill, D. 1995. Modal Structure and Reliable Inference, In D. Knill and W. Richards (eds.), Perception as Bayesian Inference, New York: Cambridge University Press.
- Marr, D. C. 1982. Vision. San Francisco: Freeman.
- McKay, B. D. 1990. Nauty User's Guide (version 1.5) Tech. Report TR-CS-90-92, Dept. Computer Science, Australian National University. For most recent version, see: <http://cs.anu.edu.au/people/bdm/nauty/>
- McKelvey, R.D. 1979. General Conditions for Global Intransitivities in Formal Voting Models, Econometrica 47, 1085-112.
- McMahon, T.A. 1976. Allometry, In Yearbook of Science and Technology. New York: McGraw Hill, pp. 48-57.
- Minsky, M. 1986. Society of Mind. New York: Simon and Schuster.
- Moon, J. W. 1968. Topics on Tournaments. New York: Holt, Rinehart and Winston.
- Moray, N. 1990. A Lattice Theory Approach to the Structure of Mental Models, Phil. Trans. R. Soc. Lond. B. 327, 577-83.
- Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems. San Mateo, Calif.: Morgan Kaufman.

- Richards, D. 1994. Intransitivities in Multidimensional Spatial Voting: Period Three Implies Chaos, Social Choice and Welfare 11, 109-119.
- Richards, D. 1996. Knowledge Structures and Equilibrium in Multidimensional Social Choice. Presented at American Political Science Association Meeting.
- Richards, D. and Richards, W. 1995. Knowledge-Induced Stability in Collective Choice, M.I.T. Center for Cognitive Science Occasional Paper #49.
- Richards, W., Wilson, H. and Sommer, M.. 1993. Chaos in Percepts? Biological Cybernetics 70, 345-349.
- Rock, I. 1983. The Logic of Perception. Cambridge: M.I.T. Press.
- Saari, D. G. 1987. Chaos and the Theory of Elections, In A. Kurzhanski and K. Sigmund (eds.) Dynamical Systems, Berlin: Springer-Verlag.
- 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 Berlin: Springer-Verlag.
- Schofield, N. 1980. Formal Political Theory. Quality and Quantity 14, 249-75.
- Schofield, N. 1983. Generic Instability of Majority Rule. Review of Economic Studies 50, 696-705.
- Shepsle, K.A. 1979. Institutional Arrangements and Equilibrium in Multidimensional Voting Models, American Journal of Political Science 23, 27-60.
- Thompson, D. W. 1968. On Growth and Form New York: Cambridge University Press.
- Van Essen, D. C., Anderson, C. and Felleman, D. J. 1992. Information Processing in the Primate Visual System: An Integrated Systems Perspective. Science 255, 419-423.

### Endnotes

---

† richards@polisci.umn.edu, wrichards@mit.edu, bdm@cs.anu.edu.au. The first author benefited from support from the National Fellows Program of Hoover Institution at Stanford University, and a visit to the Santa Fe Institute in February and March of 1997. The second author was supported in part by the Mitsubishi Electric Research Lab. The authors would like to thank Josh Tenenbaum and Paul Edelman for helpful comments.

<sup>1</sup> We use the game-theoretical terms for shared information, where mutual knowledge refers to information held by all agents, which is a weaker informational condition than common knowledge, which requires that all agents hold the same information, and know that all others hold this information ad infinitum.

---

<sup>2</sup> This example is supported by experimental results examining feasible orderings and structure among film categories using multidimensional scaling methods.

<sup>3</sup> For example, if  $W$  places an ordering on film categories, an individual's preferences over films cannot jump from "romantic comedy" to "superaction," but must move consistently through the set of categories, such as via comedy, drama, and adventure.

<sup>4</sup> Note that the dimension of a partial orders induced from  $W$  may be greater than one (as is the case in Figure 1). In other words, the effect of  $W$  is not simply to create a set of one-dimensional partial orders

<sup>5</sup> We choose a graph-theoretic approach to represent and analyze the collective choice problem because we found it more concise and powerful than other approaches, although one can also represent our knowledge structure approach using partitions in a choice space (e.g., Richards and Richards 1995) or functional relationships (e.g., Jepson et al. 1995).

<sup>6</sup> In the case of social choice, the agents are individuals and thus strategic decision making (and extensions to mutual and common knowledge) becomes more important. The issue of strategy-proofness is in progress; a preliminary treatment is in Richards (1996).

<sup>7</sup> Note that in the case of ties the equilibrium need not be unique: we are not striving to reach a unique global maximum likelihood solution.

<sup>8</sup> To give some preliminary insight into the issue of robustness, this can be evaluated by determining the rank of an  $m \times n$  matrix, where the columns are the  $n$  alternatives and the rows are the  $m$  pairwise comparisons and the entries are the signed weights between the two alternatives. When this rank is equal to or greater than  $n - 1$ , then the system will be robust (Saari 1994, Richards and Richards 1995).

<sup>9</sup> For example, labeling the vertices of a five-ring clockwise as  $a, b, c, d,$  and  $e$ , the weights  $(5/13, 2/13, 1/13, 5/13, 0/13)$  result in the top cycle  $a \succ b \succ c \succ a$ , with  $a \succ d$  and  $a \succ e$ .