Social Choice, Information Theory, and the Borda Count

by

John Clifton Lawrence

General Algorithm
POB 230351
Encinitas, CA 92023
Phone/fax: 760-633-3778
Web site: http://www.genalg.com
E-mail: jlawrence@genalg.com

September 1, 1999

Abstract

This paper adds to the drumbeat of those who clamor for the Borda Count (BC) by eliminating the stigma of "arbitrariness" previously associated with the BC. It is proven that the BC is rational in the sense that, if a candidate dies, the expected value of the social profile is identical to the original with the dead candidate's name removed. Finally, the voting paradox is resolved by the BC when the most likely individual profiles are used. The fineness or coarseness of the grid on which individuals specify their preference profiles determines the amount of information conveyed. Since this grid is traditionally determined by the number of alternatives, there is no such thing as an irrelevant alternative. The problem of social choice, in general, can be viewed as the transmission of information from multiple sources (the individuals) to one receiver (society). Since there are finite information transmission constraints, there will be some probability of error, P(e), regarding the placement or ranking of alternatives both in individual and social profiles. As individual information is increased, P(e) in the social profile can be made to approach zero as closely as desired.

Keywords

Arrow, Borda, Condorcet, Shannon, social choice, information theory, irrelevant alternatives, binary independence, rational behavior, preference profiles, social welfare function, voting paradox, impossibility theorem, probability of error

Introduction

In 1948, Claude Shannon put forth a Mathematical Theory of Communication which was inherently probabilistic. A communication system model was postulated (see Figure 1). It was shown that there exists a parameter called *channel capacity* which determines whether or not information can be transmitted at an arbitrarily low probability of error. For transmission rates below the channel capacity, arbitrarily good performance can be achieved by proper coding while, for information rates above channel capacity, arbitrarily low error rates cannot be achieved.

Social choice has been considered for the most part to be deterministic although there have been some attempts to take an information theoretic approach. [Heller, Starr and Starrett, 1986], [Hammond, 1982]. Some have considered the use of lotteries on the alternatives to select a winner, [Fishburn, 1984], and some have replaced the individual and social orderings themselves by a set of probabilities. [Intriligator, 1973] A game theoretic approach in which the objective is to maximize the expected utility over the alternatives has been suggested. [von Neumann and Morgenstern, 1944], [Luce and Raiffa, 1958]. The probability of a cycle or social intransitivity has been computed. [Williamson and Sargent, 1967]. For Condorcet-like social welfare functions (SWFs) this is equivalent to computing a probability of error, P(e).

However, no one has considered the individual preference orderings themselves to be probabilistic by virtue of the fact that each individual is allowed to specify only a finite amount of information upon which the SWF must then determine a social preference ordering. The uncertainty with regard to the individual's "true" preferences (those that would be manifested if infinite information were available) leads to possible errors in the social ordering. We show that the P(e) can be made arbitrarily small by increasing the information flow from the individuals. The fact that P(e) will always be non-zero with finite information constraints coincides with Arrow's [1963] Impossibility Theorem. Although a SWF that provides a social ordering that has even a single error, in the sense that the ordering doesn't meet Arrow's conditions and axioms for one pair of alternatives and one domain element (but provides correct solutions in every other case), would traditionally be considered not to exist, we take the approach of allowing errors in the SWF and then trying to minimize them.

A model of a social choice system is given in Figure 2. If there are m alternatives, the j^{th} individual specifies his or her preference ordering as $x_{1j} R^1 x_{2j} R^2 \ldots x_{(m-1)j} R^{m-1} x_{mj}$, where x_{ij} is one of the m alternatives and R^i
$(0 < i < m)$ can be replaced with either P (preference) or I (indifference). The problem is identical to the problem (without indifferences) of placing each of m balls in m slots, one ball to a slot, and (with indifferences) to the placing of m balls in m slots without restrictions as to the number of balls that can be placed in each slot. When there are indifferences, there will be some "open" or "blank" slots i.e. slots with no balls in them. In the traditional model those slots are "closed up" or eliminated. In general, there can be m' ≠ m slots. Each individual, ideally, should have unlimited freedom of expression when specifying his or her preference orderings and not be limited by a number of slots equal to the number of alternatives. However, by imposing the restriction that an individual must express his or her preferences in terms of this number of slots, the rational individual will feel constrained to select his or her profile by projecting each alternative from his or her "true" preferences (expressed in a number of slots limited only by his or her "sensitivity") onto the slots available. These principles are illustrated in Figure 3. Other terms for sensitivity used in the literature are finite perception [Fine, 1995] and sensibility [Ng, 1975], [Svensson, 1985]. When the open slots are eliminated, the information expressed per alternative is not the same when indifferences are involved as when there are only strictly preferences since the position of the open slot (which conveys information) is unknown.

We can calculate the amount of information expressed by an individual. As defined by Shannon, the amount of information is equal to the logarithm to the base 2 of the number of states involved. Without indifferences, the first ball (or alternative) can be placed in m ways, the second ball in m-1 ways, etc., so that there are m! possible orderings and $\log_2 m!$ bits of information expressed by each individual. With indifferences, each ball can be placed in m ways so that there are m^m possible orderings or $m\log_2 m$ bits of information expressed by each individual. Therefore, the amount of information expressed is a function of the *number of slots* which, in the traditional model is equal to the number of alternatives. If there are a large number of alternatives or slots, each individual can express his or her preferences more precisely especially if indifferences are allowed. Figure 4 shows, in general, how a rational individual will project his or her “true” preferences by expressing then in various situations in which a differing number of slots are available. It can be seen that, if one or more alternatives are removed from the original ordering (due to the death of one or more candidates, for instance), the rational individual will project his or her “true” preferences onto the number of slots appropriate to the number of remaining candidates in order to come up with his or her new preference ordering.

Figure 4 exhibits some interesting phenomena. In **A** the individual’s true preferences are expressed in a number of slots down to his or her “sensitivity” level. For purposes of simplification, we will assume that the individual’s true preference is concentrated in a spike in the middle of the “sensitivity” slot rather than being distributed over the slot. An individual’s sensitivity level results in the placing of alternatives on the finest grid possible for that individual. In Figure 4 **A** the individual is able to discriminate among all 9 alternatives so that the preference profile contains no indifferences. In **B** since there are 9 alternatives, they are expressed in 9 slots in order to be in accordance with traditional social choice theory. Notice that here alternatives d, f and i are indifferent among each other as are a and b. In **C** 3 alternatives are removed. Now c and g are indifferent and a is prefered to b. In **D** a is again indifferent to b, and in **E** a is prefered to b. Thus whether or not an alternative is prefered or indifferent to another alternative depends on the structure of the grid. A finer grid may result in indifferences being discriminated into preferences, but a coarser grid may cause an indifference to become a preference also as shown by the relationship of a and b in Figure 4.

From these examples, it can be seen that what have been considered, heretofore, irrelevant alternatives, aren't really irrelevant at all since they determine the grid over which individual decisions can be made. It is also concluded that individual data specified for one set of alternatives cannot be used with another (larger or smaller) set of alternatives with accuracy or certainty. Instead, individuals must be repolled as the number of alternatives changes, or *the individual profiles must be viewed probabilistically.* As the set becomes larger, indifferences might become preferences and, as the set becomes smaller, preferences might become indifferences and vice versa. However, when deleting alternatives from an alternative set, there is no reason why the larger grid and informational base of the original set of alternatives could not be used with the reduced set of alternatives since that information is already available.

In Arrow's model, the information expressed by individuals has been arbitrarily constrained by requiring that each individual express his preferences only in terms of binary comparisons as required by the condition of irrelevant alternatives, IIA. This has been questioned by other writers. [Fishburn, 1971]. It leads to a distortion of the individual data since what the individual might have preferred to express as an indifference may now be constrained to be a preference. The worst case scenario in which the individual might have prefered to express indifference among all alternatives will result in fully discriminated preferences as shown in Figure 5. Without assuming transitivity, binary comparisons require one bit of information for every possible pair of alternatives from each individual or

bits for preferences only and bits for preferences and indifferences. If we assume transitivity, we need less information since now we need only compare the alternatives in a chain: a with b, b with c etc. Therefore, there are m-1 bits for preferences only and bits for preferences and indifferences. It has been pointed out [Saari, 1995] that this represents a dearth of information which leads to Arrow's Impossibility Theorem. Saari states: "Adopting an informational perspective, then, [Arrow's Theorem and others like it] *just state that procedures for three or more candidates require more information than just the relative rankings of pairs."* [emphasis in the original] Clearly, the binary comparisons lead to an ordering and even a transitive ordering if we assume a rational individual, but that ordering is missing the information required to nail down which alternative is in which slot and is, therefore, arbitrary. The amount of missing information is precisely

$$m\log_2 m - \qquad\qquad \text{bits}$$

If the j^{th} individual specifies his or her preferences in the form $x_{1j}\, R^1 x_{2j}\, R^2 ... x_{(m-1)j}\, R^{m-1} x_{mj}$, then breaking this down into binary comparisons is, technically, not only throwing away information but a misrepresentation of the individual's true preferences in the first place. While the expression of the individual's preferences in m slots by projection represents a compromise with his or her true preferences, it is a much better compromise than that represented by binary comparisons! In fact it is the most congruous method of representing true preferences if congruity is defined as the sum of the differences between the expressed preferences and the true preferences. There also will be a difference between the information gathered if the individual is asked to make the binary comparisons or if the binary comparisons are inferred from an m-ary comparison supplied by the individual since, in the first case, the individual will be supplying information down to his or her sensitivity level and, in the second, the information will be derived from the individual's projection onto an m-ary grid. The amount of information the individual *could* supply is $m'\log_2 m'$ where m' is related to the individual's sensitivity level whereas the information the individual is *asked* to supply is $m\log_2 m$.

A related point is that, in the traditional model, with no "open" slots allowed, the amount of information expressed when there is one indifference is $(m-1)\log_2(m-1)$ bits since there are now m-1 slots. Consequently, the amount of information per alternative is diminished when an indifference is expressed. This gives rise to the possibility that different amounts of information can be gathered from different individuals, and different amounts of information can be gathered regarding different alternatives. In order to avoid this, the following are stated as categorical principles:

Principle 1: The amount of information collected from each individual shall be the same.

Principle 2: The amount of information collected regarding each alternative shall be the same.

These are extensions of anonymity and neutrality, respectively. However, the closing up of empty slots when indifferences are involved, strictly speaking, violates both anonymity and neutrality. Accordingly, it will be assumed that "open" slots are permitted for the sake of analysis in this paper. However, even if open slots are not permitted, this does not essentially change the results of this analysis.

Our analysis is ordinal rather than cardinal simply because we are dealing with discrete rather than continuous values. There are always a finite number of slots even when we let the number of slots get very large.

We do not assume interpersonal comparability, but we do assume intrapersonal comparability [i.e.] each individual can compare his or her own preference profiles for various numbers of slots. The individual can project his or her true preference profile onto a grid containing the available number of slots. Therefore, alternatives which are ranked indifferently on a coarser grid may be ranked preferentially on a finer grid. We assume that there is a correspondence among grids of varying coarseness for any particular individual. We assume that each individual places each alternative *individually and independently* on the grid based on his sensitivity to that alternative's place. Thus binary independence is replaced by *unary independence.*

We do not make any assumptions about preference intensity. As shown in Figure 6, whether or not the number of slots equals the number of alternatives, certain preference relationships can be said to be more intense than others. In our analysis no meaning is given to the measurement of this intensity any more than it is in traditional Arrovian analysis. For instance, in Figure 6**A** there are 6 alternatives and 6 slots. It can be said that a is prefered to b more intensely than f is prefered to e simply because there are more slots or places separating a and b than there are separating f and e. Also it may be said that d is prefered to f about as intensely as c is prefered to b. Similarly, in Figure 6**B** a is prefered to b more intensely than e is prefered to c, and a is prefered to d about as intensely as e is prefered to c. Whether or not there are blank or open slots in the one case as opposed to the other does not affect the notion of intensity.

Arrow's BC Example

In light of the previous discussion, Arrow's verbal comments and example regarding the BC must be reconsidered. It has been pointed out by others that this example and IIA, which it is claimed to be in violation of, have little in common. Hansson [1973] states: "This [example] has little to do with [IIA]. [IIA] says something about what happens when the situation (i.e. the individual preferences) changes and the environment is constant. In Arrow's [BC] example the individual preferences stay the same, but the environment is changed." Other writers [Ray, 1973] have discussed the "confusion in the literature on these concepts."

Arrow states: "Suppose an election is held, with a certain number of candidates in the field, each individual filing his list of preferences, and then one of the candidates dies. Surely the social choice should be made by taking each of the individuals' preference lists, blotting out completely the dead candidate's name, and considering only the orderings of the remaining names in going through the procedure of determining a winner. That is, the choice to be made among the set S of surviving candidates should be independent of the preferences of individuals not in S. To assume otherwise would be to make the result of the election dependent on the obviously accidental circumstance of whether a candidate died before or after the date of polling."

In taking each of the individuals' preference lists, blotting out the dead candidate's name, and considering only the orderings of the remaining names in going through the procedure of determining a winner, society is presuming to operate on individual data without contacting the individuals involved and is thereby *taking away* the individuals' freedom of choice. We have shown how a rational individual may change his or her preference ordering dependent on the alternatives available but yet totally consistent with rational criteria. Therefore, society should not order the individuals' lists in the way Arrow proposes. Instead, the individuals should be repolled *according to the coarser grid implied by the lesser number of alternatives* in order to get accurate individual information relative to that grid. This is the basic misconception of IIA. *There are no irrelevant alternatives* since the number of alternatives affects the fineness or coarseness of the grid. If the grid is made invariant with respect to the number of alternatives, then the results will be invariant, but this would necessitate having a number of slots not necessarily equal to the number of alternatives. If the grid is not invariant, as is Arrow's assumption, *individual preference orderings will vary as a function of the number of alternatives and hence the social ordering must vary also.* The accidental circumstance of whether a candidate dies before or after the date of polling will affect the results of the election because it will affect the number of slots over which each individual distributes the remaining candidates.

If the individuals cannot be repolled when the number of candidates changes, we may still be able to extract useful information about the new social profile by viewing the reduced or expanded profiles probabilistically. Upping the ante a bit from the traditional Arrovian analysis, it is required for the purposes of this paper, that not only should the winner be the same in both cases, but the entire social ordering should be the same [i.e.] the social ordering computed from individual orderings with the dead candidate's name removed should be the same as the social ordering computed from individual orderings with the dead candidate's name included and then blotted out of (or removed from) the social ordering. When this is not the case, it will be considered that an error has occurred. For instance, if the social ordering were aPbPcPd and c dies, the new social ordering recomputed from individual data but without repolling and with c removed should be aPbPd. If, instead, it turned out to be aPbId or aPdPb, for example, it could be said that one or two errors, respectively, had occurred.

Consider Arrow's example which, supposedly, condemns the BC. "In particular, suppose that there are three voters and four candidates, x, y, z, and w. Let the weights for the first, second, third, and fourth choices be 4, 3, 2, and 1, respectively. Suppose that individuals 1 and 2 rank the candidates in the order x, y, z, and w, while individual 3 ranks them in the order z, w, x, and y. Under the given electoral system, x is chosen. Then, certainly, if y is deleted from the ranks of the candidates, the system applied to the remaining candidates should yield the same result, especially since, in this case, y is inferior to x according to the tastes of every individual; but, if y is in fact deleted, the indicated electoral system would yield a tie between x and z."

However, under various scenarios (depending on the reformulated preferences of the three voters), x can either be prefered to z or be indifferent to z when y is removed. Consider Figure 7. The initial assumption is that there are four slots and two voters have xPyPzPw and one has zPwPxPy. Now when y is removed, there are three slots and the two voters that had xPyPzPw could have either xPzPw or xPzIw. Because of the uncertainty as to the precise location of z in its slot when there are four slots, z could be as arbitrarily close to y as possible without falling into the slot with y or as arbitrarily close to w as possible without falling into the slot with w. From either of these positions, when y is eliminated and the number of slots reduced to 3, z could fall into the middle or the bottom slot, respectively. Referring to Figure 7 again, voter 3's ordering, zPwPxPy, could be any of the following when y is removed:
1) zPwIx, 2) zPwPx, 3) zIwPx. Note that there are two different ways of having zIwPx. In this case, the BC can be assigned in two different ways depending on which slot is left empty.

If the voters cannot be repolled to determine which slot an alternative will fall into when some alternatives are removed (which reconfigures the grid), assumptions can be made based on the probability that a specific alternative will wind up in a specific slot. For instance, with reference to Figure 8 and assuming that an alternative is equally likely to be anywhere in its slot, for the profile $xP_iyP_izP_iw$, we can compute the probability that z will wind up in the middle slot when y is removed to be 2/3. Therefore, the probability of the individual profile xP_izP_iw is 2/3. Similarly, the probability that z will wind up in the bottom slot when y is removed is 1/3, and, consequently, the probability of the profile xP_izI_iw is 1/3. Note that x always winds up in the top slot, and w always winds up in the bottom slot.

The profile $zP_iwP_ixP_iy$ is slightly more complicated. w and/or x can wind up in the center slot. w can also wind up in the top slot, and x can wind up in the bottom slot. The probability that w winds up in the middle slot equals the probability that x winds up in the middle slot which is 2/3. The probability that w winds up in the top slot equals the probability that x winds up in the bottom slot equals 1/3. Let b signify a “blank” or “open” slot. Therefore the probability of the individual profile zPwIxPb equals 4/9. The probability of the individual profile zPwPx equals 2/9; the probability of the profile zIwPbPx is 1/9; and the probability of the profile zIwPxPb is 2/9. Therefore, the probability of the profile zIwPx is 3/9 = 1/3.

We can compute for each possible combination of profiles for voters 1, 2 and 3 the corresponding Borda counts for each of the alternatives and, therefore, the corresponding social ordering. Also, we can find the social choice or “top slot” of the social ordering and the probability of occurrence of each combination of individual profiles. This information is all presented in Appendix 1. As can be seen, out of the total of 16 cases, xPz in 14 and xIz in only 2. The probability of error P(e) for xPz is the same as the probability that xIz which is12/81. The most likely social profile is xPzPw with a P(c) of 69/81 which accords perfectly with the social profile for m=4 with y struck out of the list. We, therefore, pick as winner that profile which minimizes the probability of error which is xPzPw. This social profile derived from individual profiles with y removed is completely rational in that it is exactly what would be expected by simply removing y from the social profile for m=4.

Alternatively, we can find the expected values of the BC for x, x and w, respectively when y is removed.

E[BC(x)] for voters 1 and 2 = 3

E[BC(w)] for voters 1 and 2 = 1

E[BC(z)] for voters 1 and 2 = (2/3)(2) + (1/3)(1) = 5/3

E{BC(z)] for voter 3 = 3

E[BC(w)] for voter 3 = (1/3)(3) + (2/3)(2) = 7/3

E(BC(x)] for voter 3 = (2/3)(2) + (1/3)(1) = 5/3

The expected values of the social BCs equal the summation of the expected values for the individual BCs. Therefore,

E_{social}[BC(x)] = (2)(3) + 5/3 = 23/3

E_{social}[BC(z)] = (2)(5/3) + 3 = 19/3

E_{social}[BC(w)] = 2 + 7/3 = 13/3

We conclude that the expected values of the BCs for m = 3 are in the same order as the BCs for m = 4. We prove that this will be the case in general in Theorem 1.

Theorem 1: Consider a group of n individual preference orderings over m alternatives and the corresponding social ordering determined by the BC. If one alternative is removed, the expected values of the BCs for the social ordering will be in the same order as the original social ordering with the removed alternative blotted out. Hence, the BC is non-arbitrary.

As an example, consider Figure 9. We choose m=6. The Borda counts, then, are 1 for the lowest ranked alternative and 6 for the highest ranked. Consider any individual profile such as aPbPcPdPePf as shown in 9**A**. There are 6 slots and each alternative occupies a slot. We assume that the “true” location of each alternative is uniformly distributed over its slot as shown in 9**B**. Now if we remove a slot so that there are 5 slots, a will still fall in the top slot and f will still fall in the bottom slot. The other alternatives will have some probability of falling in two different slots. For instance, e with a BC of 2 will fall partly into slot 1 and partly into slot 2 when there are 5 slots. This is shown in 9**C** and **D.** We can compute the probability of e falling into each slot as follows. P(e in slot 1) = 0.2. P (e in slot 2) = 0.8. P(d in slot 2) = 0.4. P(d in slot 3) = 0.6. The expected value for the BC for e is E(BC for e) = (0.2)(1) + (0.8)(2) = 1.8. E(BC for d) = (0.4)(2) + (0.6)(3) = 2.6. The expected values for the other alternatives are given in Fig. 9**E**.

<u>Proof:</u> Let m be the original number of alternatives for some individual preference ordering. Let j be a variable that denotes the place value ($1 \le j \le m$) and m-j+1 denote the BC. There are a total of m places or slots. For instance, when an alternative is in first place, the BC is m, and when an alternative is in m^{th} place, the BC is 1. When one alternative is removed, there is one less slot ($1 \le j \le m-1$). The individual will project his first (last) place preference for m alternatives onto the first (last) place slot for m-1 alternatives since the slots are larger for m-1 than for m. The overlap from the j^{th} slot for m slots ($2 \le j \le m-1$) onto the $(j-1)^{th}$ slot for m-1 slots is . This represents the probability (divided by m) that the alternative in the j^{th} slot (for m slots) will end up in the $(j-1)^{th}$ slot (for m-1 slots). The probability that the alternative in the j^{th} slot will end up in the j^{th} slot is .

Therefore, the expected value of the BC for m-1 slots for one individual is

$E[BC_{ind}(j)|(m-1) \text{ slots}] =$, $1 \le j \le m$.

This is equivalent to , $1 \le j \le m$.

This represents a linear transformation: , $0 \le k \le m-1$

Therefore, the expected values of the BCs for m-1 slots are linear transformations of the values of the BCs for m slots. The social BC for the i^{th} alternative ($1 \le i \le m$) is the sum of the individual BCs for that alternative over all n individuals:

BC_{social} (i^{th} alt.)|m slots = . Since the expected value of the sum is equal to

the sum of the expected values, the expected value of the social BC for the j^{th} alternative with one alternative removed is as follows:

E[BC_{social} (i^{th} alt.)|m-1 slots] = |m-1 slots].

where p_k(i) = the number of k^{th} place votes for the i^{th} alternative over all individuals

E[BC_{social} (i^{th} alt.)|m-1 slots] =

= + n

Therefore, the expected value of the social BC for each alternative is the original BC multiplied by a constant plus another constant. This is sufficient to show that the ordering of expected values of the BCs with one alternative removed is the same as the original ordering of the BCs. QED.

Returning to Arrow's example of the BC, when alternatives are removed, the BC can be taken with the original grid since that information is already available. There is no set rule that it must be taken with a maximum count (or grid) of 3 when information is available for a grid of 4. Arrow's assumption amounts to throwing available information away. Assuming a grid with 4 slots and y removed, x would have a count of 10, z a count of 6 and w a count of 4 yielding xPzPw while the original result would have included a count of 7 for y yielding xPyPzPw and striking out y in this profile yields the expected result.

One final point regarding Arrow's example is that the results for 3 alternatives are derived *relative* to the results for 4 alternatives. However, the results for 4 alternatives, themselves are also relative and not absolute because they were derived relative to the true preference values. In general, when alternatives are removed or added, individual and social orderings are subject to change and, therefore, must be considered to be probabilistic.

Paradox of Voting

The paradox of voting occurs in the simplest example when there are three voters having, respectively, the following three preference orderings: aP_1bP_1c, bP_2cP_2a and cP_3aP_3b. On the binary level, the following relationships are assumed to occur: 2 votes for aP_ib and 1 for bP_ia yielding aPb, 2 votes for bP_ic and 1 for cP_ib yielding bPc and 2 votes for cP_ia and 1 for aP_ic yielding cPa. There is no single 3-level social ordering that is consistent with the three binary orderings in the sense that, when one of the alternatives is blotted out ([e.g.] when the candidate dies), the resulting binary orderings are consistent with the known binary orderings. However, in light of the present discussion, if the ternary preference orderings are specified by each individual, the binary orderings that each individual would specify are unknown. With reference to Figure 10, it can be seen that a specification of aP_1bP_1c at the ternary level is consistent with a specification of aI_1b, bP_1c, aP_1c or aP_1b, bI_1c, aP_1c at the binary level. bP_2cP_2a is consistent with bI_2c, cP_2a, bP_2a or bP_2c, cI_2a, bP_2a at the binary level. cP_3aP_3b is consistent with cI_3a, aP_3b, cP_3b or cP_3a, aI_3b, cP_3b.

We now prove that the BC which yields aIbIc for the social ordering is consistent with the most likely social orderings at the binary stage. There are 8 cases in all. Appendix 2 lists the cases and the results for each case. No matter which alternative is removed, the probability of a tie for m=2 equals 1/2 while the probability that xPy equals 1/4 and the probability that yPx also equals 1/4. Therefore, the most likely social ordering for m=2 is aIb with c removed, bIc with a removed and aIc with b removed. Therefore, this case is in accordance with Theorem 1 and the most likely results for the binary case are the same as the result for the ternary case with the appropriate alternative removed from the ternary ordering. The BC yields entirely rational and consistent results.

Simple Proof of Arrow's GPT

Arrow's General Possibility Theorem (GPT) or Impossibility Theorem as it's also known states that no SWF exists which meets his stated rational and ethical criteria. There have been various attempts and comments regarding simplification of the proof. [Feldman, 1974], [Stevens, 1976]. It can be shown that this theorem can be proven using only completeness, transitivity, universality, Pareto, binary independence, and neutrality. The dictatorship condition is not needed Furthermore, it can be done in just a few steps. The steps are as follows:

1) Let there be m alternatives and n voters.

2) By universality, a SWF must provide a solution for every possible domain element in order for it to exist. Therefore, it must satisfy the domain element m=3, n=2 for which voter 1 has aP_1bP_1c, voter 2 has bP_2cP_2a.

3) By binary independence and Pareto, bP_1c and bP_2c produce bPc.

4) aP_1c and cP_2a must produce aIc by binary independence and neutrality.

5) aP_1b and bP_2a must produce aIb by binary independence and neutrality.

6) aIbPcIa is intransitive.

7) Therefore, no SWF exists. QED

From the above analysis, it can be concluded that Arrow's GPT is a more sophisticated and complex restatement of the voting paradox and the dictatorship condition is not needed.

Probability of Error in Social Ordering

As has already been established, when the number of alternatives is changed, there may be quite rational and predictable changes in the individual preference orderings and hence in the social ordering produced by any SWF. Additionally, even if the individuals re-express their preferences after new alternatives have been introduced or withdrawn, there is still a probability of error due to the discrepancy between the individual's true preferences and the expressed preferences.

Other writers have considered the special case in which n approaches infinity [Chichilnisky, Heal, 1997], [Fishburn, 1970], and the case in which both m and n approach infinity. [Grafe, Grafe, 1983], [Pazner, Wesley, 1977]. As m', the number of slots, gets very large, we might expect that the slot assigned to any particular alternative by any particular individual will vary only slightly and that the social ordering will become more stable as alternatives are added or removed. The probability of error in the SWF should decrease. We might expect that the probability of error will decrease with m' and increase with n.

We next derive an expression for the individual probability of error, P_i (e). We will assume that an error occurs in an individual preference profile when, due to grid constraints, an individual ranks two alternatives as indifferent when a finer grid would reveal that one is preferred to the other. For the purposes of this analysis, we assume that as the grid becomes finer, all indifferences will eventually be discriminated into preferences. Also we assume infinite sensitivity on the part of the individuals. Let the two alternatives be x and y. Therefore, an error occurs when xI_iy is expressed and the individual's true preference is xP_iy or yP_ix. This amounts to the individual's placing x and y in the same slot when a finer grid would reveal that they are actually placed in neighboring slots according to the individual's true preference profile.

Assuming that each alternative is equally likely to be in any slot, the probability of error, P_i (e), is the probability that at least two alternatives will fall in the same slot given that there are m alternatives and m' slots. Alternatively, the probability that the individual profile is correct, P_i (c), is the probability that there is at most one alternative per slot. Let's assume initially that m' = m. The probability of a particular alternative being in a particular slot is 1/m. The probability of m alternatives being in m particular slots is $1/m^m$. The number of permutations of m alternatives is m! Therefore, the probability that m alternatives are in m separate slots, P_i (c), is $m!/m^m$. =0. Therefore, P_i (e) =1.

Now, let m' > m. The probability that one particular alternative is in one particular slot is 1/m'. The probability that m alternatives are in m particular slots is $1/(m')^m$. The number of permutations of m alternatives in m' slots is m'(m'-1)(m'-2)$\cdots$(m'-m-1). The probability that the m alternatives are in any m separate slots is

P_i (c) =

and

Therefore, as the number of slots increases, we can force the probability of error arbitrarily close to 0. The probability of error over the entire population of n voters would be nP_i (e), and this too could be limited to any arbitrary value by increasing m'. The probability of error for the social preference profile, P(e), assuming no errors introduced from other sources, would be equal to nP_i (e) which again could be made to approach zero as $m' \rightarrow \infty$. The probability of error for the SWF can be made arbitrarily as low as possible by increasing the amount of information per alternative from each individual. A more in depth analysis could take into account finite sensitivity. In this case, it would seem that the value of m' necessary to achieve any particular P(e) would be significantly less.

Beyond the BC — New Directions

The BC will always result in one and exactly one social ordering, and this is one of its virtues. However, since the final positions of the alternatives in the social profile represent an average of their positions over all the individuals, widely varying individual profiles may result in the same social profile produced by the BC. Also, the BC in and of itself does not embody any particular ethical criteria. It may be desirable to use another SWF which is ethically based and which produces initially a set of possible SWFs that can be winnowed further according to another set of criteria.

Consider the SWF that measures the mismatch between each possible social profile and the aggregated individual profiles. By measuring the "distance" between each alternative in the "trial" social profile and the individual profiles and summing, we get a measure of the mismatch between the "trial" profile and the aggregated individual profiles. Consider the trial social profile bac and the individual profile abc. The "distance" from a in the individual profile to a in the trial profile is 1. Similarly for b. Since c is in the same position in both profiles, the distance is 0. Therefore, the total mismatch is 2. The mismatch for a trial profile abc would be zero and for cba, 4. By summing over the entire population a measure of the mismatch for each possible trial social profile can be obtained. Then, that set of profiles which minimizes the mismatch can be selected for further consideration. The next criterion that might be applied would be to select that subset of profiles that maximized some equality criterion, for example. Finally, this set can be winnowed further by applying a Rawlsian criterion [Rawls, 1971] such as selecting the subset which minimizes the mismatch between the trial SWF and the least well-off individual. This criterion can be reapplied until the trial set is reduced to one social profile.

Summary and Conclusions

We have shown that the BC is not arbitrary when considered from a probabilistic point of view but, *au contraire*, extremely rational. We have proven that the expected value of the social profile produced by the BC when one alternative is removed is identical to the original profile with the removed alternative blotted out. A consequence of the proven rationality of the BC is the resolution of the voting paradox. It is shown that the BC for the social profile which is aIbIc is completely consistent with the most likely social profiles at the binary level which are aIb, aIc and bIc, respectively. Also, a simplified proof of Arrow's GPT is provided.

We have shown that a rational individual will order alternatives with regard to the number of slots that are available by projection from his or her true preferences rather than by binary comparisons. When the grid is relatively coarse, (few slots available) some alternatives may fall into the same slot whereas, when the grid is relatively fine (many slots available), the individual is able to discriminate among alternatives that were previously classified as indifferent. We assume that the individual's true preference profile is placed on a much finer grid than the one society makes available for the individual's input. The fineness of the individual's true preference grid is limited only by the ability to discriminate or sensitivity. In the traditional model, the number of slots equals the number of alternatives, but more information can be provided if the number of slots is increased beyond the number of alternatives.

The source of error in the social profile is linked to the finite amount of information that can be gathered from each individual and due to the necessary uncertainty involved in any individual preference profile that can be considered by society due to limited information gathering and processing resources. This error is manifested in a preference profile which diverges in one or more places from the profile that would result if infinite information were available. It can be driven arbitrarily close to zero providing that society is willing to pay the cost of collecting and processing the additional information. In the real world, we anticipate that this information will eventually be collected via the internet and processed via computers. Therefore, there will be an associated cost per bit. In the final analysis, it comes down to how much precision society deems necessary vs. the price society is willing to pay for it.

Finally, we have shown that the same result that applies to the communication of information in communication systems applies to the aggregation of preferences in social choice: there will always be some probability of erroneous results when the information gathered is finite, but that error can be made as low as desired by increasing the amount of information collected.

References

1. Arrow, KJ (1963) Social Choice and Individual Values. John Wiley & Sons Inc., New York.
2. Chichilnisky, G, Heal, G (1997) Social choice with infinite populations: construction of a rule and impossibility results. Social Choice and Welfare 14: 303-318.
3. Feldman, AM (1974) A very unsubtle version of Arrow's Impossibility Theorem. Economic Inquiry 12: 534-546.
4. Fine, BJ (1995) On the relationship between true preference and actual choice. Social Choice and Welfare 12: 353-361.
5. Fishburn, PC (1970) Arrow's impossibility theorem: concise proof and infinite voters. Journal of Economic Theory 2:103-106.
6. Fishburn, PC (1971) Should social choice be based on binary comparisons. Journal of Mathematical Sociology 1:133-142.
7. Fishburn, PC (1984) Probabilistic Social Choice Based on Simple Voting Comparisons. Review of Economic Studies 51(4): 683-692.
8. Grafe, J, Grafe, F (1983) On Arrow-type impossibility theorems with infinite individuals and infinite alternative. Economic Letters 11: 75-79.
9. Hammond, PJ (1982) Utilitarianism, uncertainty and information, in Sen AK, Williams, B (eds), "Utilitarianism and Beyond," Cambridge University Press, Cambridge, 85-102.
10. Hansson, B (1973) The independence condition in the theory of social choice. Theory and Decision 4:25-49.
11. Heller, WP, Starr, RM, Starrett, DA (1986) Uncertainty, information, and communication. Cambridge University Press, Cambridge.
12. Intriligator, MD (1973) A Probabilistic Model of Social Choice. Review of Economic Studies 40(4): 553-560.
13. Luce, RD, Raiffa, H (1958) Games and Decision. John Wiley and Sons, New York.
14. Ng, Y-K (1975) Bentham or Bergson? Finite sensibility, utility functions, and social welfare functions. Review of Economic Studies 42(4): 545-569.

15. Pazner, EA, Wesley, E (1977) Stability of social choices in infinitely large societies. Journal of Economic Theory 14:252-262.
16. Rawls, J (1971) A Theory of Justice. the Belknap Press of Harvard University Press, Cambridge, MA.
17. Ray, P (1973) Independence of Irrelevant alternatives. Econometrica 41: 978-991.
18. Saari, DG (1990) The Borda Dictionary. Social Choice and Welfare 7: 279-317.
19. Saari, DG (1994) Inner consistency or not inner consistency; a reformulation is the answer. Social Choice and Welfare, In: Barnett W, Moulin H, Salles M, Schofield N (eds) Social Choice and Welfare.
20. Saari, DG (1995) Basic Geometry of Voting. Springer-Verlag, Berlin.
21. Saari, DG (1997) Informational Geometry of Social Choice. Social Choice and Welfare 14: 211-232.
22. Saari, DG (1998) Connecting and resolving Sen's and Arrow's Theorems. Social Choice and Welfare 15: 239-261.
23. Sen, AK (1993) Internal Consistency of Choice. Econometrica, 61: (3) 495-521.
24. Shannon, CE (1948) The Mathematical Theory of Communication. Bell System Technical Journal 27: 379-423, 623-56.
25. Stevens, DN (1976) A very unsubtle version of Arrow's impossibility theorem: a subtle comment. Economic Inquiry 14: 297-299.
26. Svensson, L-G (1985) The utilitarian criterion, finite sensibility and the weak majority preference principle: a further analysis. Social Choice and Welfare 2: 23-35.
27. von Neumann, J, Morgenstern, O, (1944) Theory of Games and Economic Behavior. Princeton University Press, Princeton.
28. Williamson, OE, Sargent, TJ (1967) Social choice: a probabilistic approach. Economic Journal 77: 797-813.