

On measures of cohesiveness under dichotomous opinions: some characterizations of Approval Consensus Measures

J. C. R. Alcantud, R. de Andrés Calle, J. M. Cascón

*Facultad de Economía y Empresa, University of Salamanca,
37007 Salamanca, Spain
{jcr, rocioac, casbar}@usal.es*

Abstract

The measurement of the degree of agreement in a group has recently attracted considerable attention by researchers from various fields. In this paper we consider situations where each member of a population classifies a list of options as either “acceptable” or “non-acceptable” (as in job committees or elections by approval voting); either “agree” or “disagree” (as in polls or surveys); either “guilty” or “not guilty” (as in jury courts); etc. In order to measure the cohesiveness that the expression of such dichotomous opinions conveys, we propose the novel concept of approval consensus measure (ACM), which does not refer to any priors of the agents like preferences or other decision-making processes. Then we give axiomatic characterizations of two generic classes of ACMs. Finally, we focus on the 2012 presidential elections in USA as a real scenario to put in practice these two proposals.

Keywords: Dichotomous opinions, Consensus measures, Tanimoto similarity index, 2012 USA presidential elections

JEL classification: D70.

1. Introduction

Social Choice and Decision Making Theories try to give answers to many daily real situations. The study, analysis, and testing of the way that individual preferences produce a “representative” collective choice is an important research area. In particular, the measurement of the degree of agreement

in a group has attracted growing attention although its study is essentially complex and controversial (because it involves the treatment of individual opinions, aggregation procedures or voting rules; it is conditional on the context; etc), a handicap that is common to virtually all branches of the Social Sciences. Both aspects -collective choice and group cohesiveness- have been the subject of a joint treatment in [2] and [3].

In this paper we focus on the study of the degree of cohesiveness in situations where a group of agents express their opinions in a dichotomous way, e.g., because they intend to apply *approval voting* (AV) in order to reach a collective decision. Many successful theoretical and empirical works on AV have studied this voting system from various points of view (see [10] for a comprehensive analysis). We emphasize the interest of this particular field of application for two main reasons. In the first place, because AV adapts to many real-world situations. Since the publication of Brams and Fishburn [6, 7], many organizations and scientific societies use AV, to wit, the Mathematical Association of America (MAA), the American Mathematical Society (AMS), the Institute for Operational Research and Management Sciences (INFORMS), the American Statistical Association (ASA), the Institute of Electrical and Electronic Engineers (IEEE), and other smaller societies such as the Society for Judgment and Decision Making, the Society for Social Choice and Welfare, etc. (see [10]). In the second place, because the conclusion that ‘the Approval Voting method is more likely to lead to a consensus vote than polarizing the electorate’ (Alós-Ferrer and Granić [4, p. 173]) makes dichotomous assessments worth investigating in relation with the measurement of cohesiveness in a society.

Summing up on these arguments, we believe that the practical interest of consensus models in social choice and decision making (as the recent reports in [9]) makes it pertinent to perform a separate study of the degree of consensus or cohesiveness when agents express themselves through dichotomous opinions. In our proposal we proceed according to the following starting points. Experts have to classify the alternatives e.g., as either “acceptable” or “non-acceptable” (as in job committees or elections by AV); either “agree” or “disagree” (as in polls or surveys); either “guilty” or “not guilty” (as in jury courts); etc. Then, and following the approach initiated in Bosch [5], we propose and characterize some classes of measures of the consensus or cohesiveness that such expression of the individual assessments conveys. We generically refer to them as *approval consensus measures* (ACMs), a novel concept that we introduce in this paper in order to deal with such kind of

situations.

ACMs play a role in Social Choice alike to that of “standard deviation” for data sets. In our context, in the presence of a voting or polling situation we can have a numerical indicator of how closely the votes or opinions are clustered. The precise meaning of the term “closeness” produces alternative candidate indicators. The suitability or unsuitability of a given index is dependent on the context, and its descriptive analysis helps to discriminate its field of application. In this seminal work we are interested in two essential approaches, to wit, the *s-Simple* and the *s-Simple Tanimoto* ACMs.

We first give an axiomatic characterization of *s-Simple* ACMs (that measure the probability that a randomly chosen contraction of the set of candidates to a subset with s elements produces unanimity). Then we reproduce the analysis for a modification of the class of *s-Simple* ACMs, that intends to lessen the influence of irrelevant alternatives (i.e., those whom nobody approves of). To that purpose we draw inspiration from the *Tanimoto similarity index* (see [12] and [13]) in order to define and characterize *s-Simple Tanimoto* ACMs. This similarity index and others have been applied in different fields, especially in Biogenetics (see [8], [11] and [14] as a sample). As a result, this variation of the *s-Simple* ACMs verifies an independence of irrelevant alternatives property, which supposes a distinctive feature of *s-Simple Tanimoto* ACMs. Finally, we focus on the forecasts made by polling agencies about the 2012 presidential elections in USA as a real scenario to put in practice these two proposals.

The paper is structured as follows. Section 2 is devoted to present standard basic terminology, as well as our original proposal of measurement of consensus for the specific setting with dichotomous opinions, the approval consensus measure. In Section 3 and Section 4 we set forth the characterization and properties of two families of approval consensus measures, namely *s-Simple* and *s-Simple Tanimoto* ACMs, respectively. Next, in Section 5 we present our illustrative real example. Finally, in Section 6 we give some concluding remarks.

2. Notation and definitions

Let $X = \{x_1, \dots, x_k\}$ be the finite set of k options, alternatives or candidates. It is assumed that X contains at least two alternatives, i.e., that the cardinality of X is greater or equal than 2, $|X| \geq 2$. Abusing notation, on

occasions we refer to option x_s as option s for convenience. A population of agents or experts is a finite subset $\mathbf{N} = \{1, 2, \dots, N\}$ of natural numbers.

We consider that each expert can vote for or approve of as many options, alternatives or candidates as he/she wishes, thus showing extreme and dichotomous opinions. As is standard, in order to formalize these assessments we can take three alternative positions.

1. Let $\mathcal{P}(X)$ be the set of all subsets of X . For any expert $i \in \mathbf{N}$, let $B_i \in \mathcal{P}(X)$ be the set of alternatives that he/she approves of. We write $\mathbb{P} = \mathcal{P}(X)^N$ for the set of all the assessments on X , i.e., an element $B_1 \times \dots \times B_N \in \mathbb{P}$ captures the sets of alternatives that the respective agents approve of.

2. We can also capture the dichotomous opinions of expert $i \in \mathbf{N}$ on X by means of $A_i \in \{0, 1\}^k$, i.e., component j of A_i is 1 if and only if expert i approves of alternative j . We write $\mathbb{V} = \{0, 1\}^k \times \dots \times \{0, 1\}^k$ for the set of all dichotomous experts' opinions on X , thus $A_1 \times \dots \times A_N \in \mathbb{V}$ captures the sets of alternatives that the respective agents approve of.

The elements of \mathbb{P} can be identified with elements of \mathbb{V} in a trivial manner.

3. An *approval profile* of the society \mathbf{N} on the set of alternatives X is an $N \times k$ matrix

$$M = \begin{pmatrix} M_{11} & \dots & M_{1k} \\ \vdots & \ddots & \vdots \\ M_{N1} & \dots & M_{Nk} \end{pmatrix}_{N \times k}$$

where M_{ij} is the opinion of the expert i over the alternative x_j , in the sense

$$M_{ij} = \begin{cases} 1 & \text{if expert } i \text{ approves the alternative } x_j, \\ 0 & \text{otherwise.} \end{cases}$$

We write $\mathbb{M}_{N \times k}$ for the set of all $N \times k$ matrices. Again, the elements of \mathbb{P} , resp. of \mathbb{V} , can be identified with elements of $\mathbb{M}_{N \times k}$ in a trivial manner.

For one thing, and relating to past notation, row i of the profile M can be identified with $A_i \in \{0, 1\}^k$ thus it describes the dichotomous assessment of expert i over the alternatives. For another, column j of the profile M captures the experts' assessments on the alternative j . We denote it by M^j .

Any permutation σ of the experts $\{1, 2, \dots, N\}$ determines a profile M^σ by permutation of the rows of M : row i of the profile M^σ is row $\sigma(i)$ of

the profile M . Similarly, any permutation π of the alternatives $\{1, 2, \dots, k\}$ determines a profile ${}^\pi M$ by permutation of the columns of M : column j of the profile ${}^\pi M$ is column $\pi(j)$ of the profile M .

For each approval profile M , its restriction to a *subprofile* on the alternatives in $I \subseteq X$, denoted M^I , arises from exactly selecting the columns of M that are associated with the respective alternatives in I (in the same order). In particular, and dropping brackets for simplicity, $M^{\{j\}} = M^j$ is column j of M , and $M^{i,j}$ is the two-column submatrix of M that consists of columns i and j (in the same order). Abusing notation, M^{-j} is the restriction of M to the subprofile on $X - \{x_j\}$. An s -restricted profile of M is the restriction of M to a subprofile on s alternatives.

Any partition $\{I_1, \dots, I_t\}$ of $\{1, 2, \dots, k\}$, that we identify with a partition of X , generates a *decomposition of M* into subprofiles M^{I_1}, \dots, M^{I_t} .

For each approval profile M on k alternatives, by n_0 we denote the number of alternatives that all agents disapprove of, and by n_1 we denote the number of alternatives that all experts approve of. Similarly, $n = n_0 + n_1$ denotes the number of alternatives the agents are unanimous on, i.e., the number of columns of M that are constant. Strictly speaking, the notation $n(M)$, $n_0(M)$, $n_1(M)$ should be used in order to emphasize the dependence of these amounts on the approval profile. We believe that dropping the reference to M does not cause any confusion thus we omit it.

The following example illustrates the use of the previous notation.

Example 1. Let $X = \{x_1, x_2, x_3, x_4\}$ be a set of four alternatives, thus $k = 4$ (we also use $X = \{1, 2, 3, 4\}$ for simplicity). We suppose a population of $N = 3$ agents, i.e. $\mathbf{N} = \{1, 2, 3\}$, and the following approval profile M :

$$M = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

This abstract specification may capture different real situations. For example, the agents could be a three-member court deciding if each of the judgements x_1 to x_4 (like ‘the defendant did a certain action’ or ‘the defendant is liable’) is true or false. Or they could be three performance appraisal reviewers evaluating if an employee or department has reached certain goals or not. Whatever the case, here we are not interested in how such assessments are used subsequently.

The information in M can also be expressed by setting $B_1 = \{x_1, x_3\}$, $B_2 = \{x_1, x_2\}$ and $B_3 = \{x_1, x_2, x_3\}$, or alternatively, $A_1 = (1, 0, 1, 0)$, $A_2 = (1, 1, 0, 0)$ and $A_3 = (1, 1, 1, 0)$. Besides, $n_0 = n_1 = 1$.

Suppose a permutation σ of the experts $\mathbf{N} = \{1, 2, 3\}$ given by

$$\sigma: \mathbf{N} \rightarrow \mathbf{N}$$

1	3
2	2
3	1

, then row 1 of M^σ is row $\sigma(1) = 3$ of M , and so forth,

therefore $M^\sigma = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$.

Suppose a permutation π of the alternatives $X = \{1, 2, 3, 4\}$ given by

$$\pi: \mathbf{X} \rightarrow \mathbf{X}$$

1	2
2	3
3	4
4	1

, then column 1 of ${}^\pi M$ is column $\pi(1) = 2$ of M , and so forth,

therefore ${}^\pi M = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$.

Let $I_1 = \{1, 4\}$ and $I_2 = \{2, 3\}$, then M is decomposed into the following two subprofiles:

$$M^{I_1} = M^{1,4} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \text{ and } M^{I_2} = M^{2,3} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

An approval profile M is *unanimous* if the set of approved alternatives is the same across experts. In matrix terms, if the columns of $M \in \mathbb{M}_{N \times k}$ are constant. Equivalently: if $n = k$, or if $A_1 = \dots = A_N$. For convenience, $(1)_{N \times k}$ denotes the $N \times k$ matrix whose cells are universally equal to 1.

An *extension* of an approval profile M of the society \mathbf{N} on $X = \{x_1, \dots, x_k\}$ is an approval profile \tilde{M} on $\tilde{X} = \{x_1, \dots, x_k, x_{k+1}, \dots, x_{k'}\}$ such that the restriction of \tilde{M} to the first k alternatives of \tilde{X} coincides with M .

An *expansion* of an approval profile M of the society \mathbf{N} on $X = \{x_1, \dots, x_k\}$ is an approval profile \bar{M} of a society $\bar{\mathbf{N}} = \{1, \dots, N, N + 1, \dots, \bar{N}\}$ on $X = \{x_1, \dots, x_k\}$ such that the restriction of \bar{M} to the first N experts of $\bar{\mathbf{N}}$ coincides with M .

Inspired by the concept of a consensus measure in Bosch [5], we now define approval consensus measures as follows:

Definition 1. An *approval consensus measure* (also ACM for simplicity) is a mapping $\mu : \mathbb{M}_{N \times k} \rightarrow [0, 1]$ that assigns a number $\mu(M) \in [0, 1]$ to each approval profile M , with the following properties:

- i) $\mu(M) = 1$ if and only if M is unanimous.
- ii) *Anonymity*: $\mu(M^\sigma) = \mu(M)$ for each permutation σ of the agents and $M \in \mathbb{M}_{N \times k}$
- iii) *Neutrality*: $\mu(\pi M) = \mu(M)$ for each permutation π of the alternatives and $M \in \mathbb{M}_{N \times k}$

Remark 1. The minimal set of conditions i) - iii) in Definition 1 does not guarantee a good performance of the evaluation, and ACMs with a weird behavior can be constructed easily. For example, suppose $k > 1$ and for each approval profile M define

$$w(M) = \begin{cases} 1 & \text{if } n = k, \\ 9/10 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Because w is defined as a function of n , anonymity and neutrality are trivial. Condition i) is verified by definition. However its behavior is clearly odd: it says that for example, for $k = 1000$ the ‘consensus’ is higher when no option produces a unanimous reaction (i.e. $n = 0$) than in the case where $n = 999$ options are unanimously accepted.

In view of this drawback, the researcher must attempt to avoid misspecifications by proposing restrictive forms for ACMs. Their respective normative behavior must be studied for the purpose of validation. This is the aim of the following sections.

3. The class of s -Simple ACMs. Characterization and properties

In this section we introduce and analyze a class of approval consensus measures that depend on a parameter $k \geq s > 0$. Intuitively, they measure the probability that a randomly chosen s -restricted profile of a given profile is unanimous. Formally:

Definition 2. The s -Simple approval consensus measure is the mapping $\mathcal{C}_s : \mathbb{M}_{N \times k} \rightarrow [0, 1]$ given by

$$\mathcal{C}_s(M) = \frac{n(n-1)\dots(n-(s-1))}{k(k-1)\dots(k-(s-1))} = \frac{\binom{n}{s}}{\binom{k}{s}} = \frac{C_n^s}{C_k^s} \quad (1)$$

for each approval profile M on k alternatives, where n denotes the cardinality of the set of alternatives the agents are unanimous on.

It is not difficult to check that these expressions produce approval consensus measures. The key part is that their values lie in $[0, 1]$, and particularly that $\mathcal{C}_s(M) \geq 0$ throughout. This holds true because when $n \geq s$ the numerator is a product of positive numbers, and when $n < s$ one of the factors in the numerator is 0.

Example 2. In the situation of Example 1 one has $n = 2$ thus $\mathcal{C}_1(M) = \frac{2}{4}$, $\mathcal{C}_2(M) = \frac{1}{6}$, and $\mathcal{C}_3(M) = \mathcal{C}_4(M) = 0$.

We proceed to explore some general properties of the class of s -Simple approval consensus measures. Afterwards we provide an axiomatic characterization in our main result in Theorem 1. Because particular instances of this class of approval consensus measures have specific interpretations we also perform additional *ad-hoc* analyses of focal cases.

Let M denote an approval profile. The following properties hold true:

1. $\mathcal{C}_1(M) \geq \mathcal{C}_2(M) \geq \dots \geq \mathcal{C}_k(M)$ because $\mathcal{C}_s(M) = 0$ when $s > n$, and for each index s such that $s < k$ and $s \leq n$ one has

$$\frac{\mathcal{C}_{s+1}(M)}{\mathcal{C}_s(M)} = \frac{\frac{\binom{n}{s+1}}{\binom{k}{s+1}}}{\frac{\binom{n}{s}}{\binom{k}{s}}} = \frac{n-s}{k-s} \leq 1 \text{ since } n \leq k$$

2. Reversal invariance: If we define the complementary approval profile M^c of M according to $A_i^c = X - A_i$ (in matrix terms: $M^c = (1)_{N \times k} - M$) then $\mathcal{C}_s(M) = \mathcal{C}_s(M^c)$ because the set of alternatives the agents are unanimous on does not change. This property is the analogue of Axiom 5 in Alcantud et al. [2] for the current setting.
3. The consensus measure does not change if agents are replicated in any number. This is a strengthened version of the replication axiom (for the analysis of societies with linear orders on the alternatives) in Alcalde and Vorsatz [1].
4. Suppose that an agent $N + 1$ is added to the society \mathbf{N} , and that this agent approves, resp. disapproves, of all the alternatives in X . Then the consensus measure does not rise.

We argue for the case where agent $N + 1$ disapproves of all the alternatives in X , the other case being symmetrical. The expanded approval profile \bar{M} that thus arises has n_0 unanimously disapproved alternatives and 0 unanimously approved alternatives. Therefore a comparison must be made between the following two numbers:

$$\mathcal{C}_s(M) = \frac{C_{n_0+n_1}^s}{C_k^s}, \mathcal{C}_s(\bar{M}) = \frac{C_{n_0+0}^s}{C_k^s}$$

thus $\mathcal{C}_s(M) \geq \mathcal{C}_s(\bar{M})$ if and only if $C_{n_0+n_1}^s \geq C_{n_0}^s$.

5. Suppose that an alternative $k + 1$ is added to the set of alternatives X , and that this alternative is unanimously approved, resp. disapproved, by all agents. If M is not unanimous, and the introduction of the new alternative does not affect the agents' assessments of the original alternatives, then the consensus measurement of the enlarged sets is strictly higher than the original one.

The argument is as follows. Let \tilde{M} be the extended profile, we want to show that $\mathcal{C}_s(M) < 1$ implies $\mathcal{C}_s(M) < \mathcal{C}_s(\tilde{M})$. Equivalently,

$$\mathcal{C}_s(M) = \frac{C_n^s}{C_k^s} < \mathcal{C}_s(\tilde{M}) = \frac{C_{n+1}^s}{C_{k+1}^s}$$

which after some algebra is equivalent to $k > n$. This is true because M is not unanimous.

6. Convergence to full unanimity can be established if we repeatedly introduce alternatives with the property that all agents agree on their

acceptability. Formally: Suppose that alternatives $k + 1, \dots, k + t$ are added to the set of alternatives X , and that each alternative is either unanimously approved or unanimously disapproved by all agents. If the introduction of new alternatives does not affect the agents' assessments of past sets of alternatives, then the consensus measurement of the extended approval profiles $\tilde{M}^{(t)}$ approaches 1 when t tends to infinity.

The argument is as follows. If $n = k$ the claim is trivial, so we assume $n < k$. Since $\tilde{M}^{(t)}$ is the profile after enlarging the set of alternatives with the $k + 1, \dots, k + t$ new alternatives, we want to show that $\lim_{t \rightarrow \infty} (\mathcal{C}_s(\tilde{M}^{(t)})) = \lim_{t \rightarrow \infty} (\frac{C_{k+t}^s}{C_{n+t}^s}) = 1$ for each $s \leq k, n < k$. Because

$$\lim_{t \rightarrow \infty} \left(\frac{C_{n+t}^s}{C_{k+t}^s} \right) = \lim_{t \rightarrow \infty} \frac{\frac{(n+t)!}{(n+t-s)!}}{\frac{(k+t)!}{(k+t-s)!}} = \lim_{t \rightarrow \infty} \left(\frac{(k+t-s)}{(k+t)} \dots \frac{(n+t-s+1)}{(n+t+1)} \right)$$

and this is a finite product of constantly $k - n$ sequences that converge to 1 (when t tends to infinity), the thesis ensues.

7. The computation of the s -Simple ACM of a profile on k alternatives is the average of the corresponding measures of $(k - 1)$ -restricted profiles:

Proposition 1. Let $X = \{x_1, \dots, x_k\}$ be a set with k alternatives, and let M be an approval profile on X . Then

$$\mathcal{C}_s(M) = \frac{1}{k} \sum_{j=1}^k \mathcal{C}_s(M^{-j})$$

Proof. If the agents are unanimous on exactly n alternatives of X , the neutrality property of approval consensus measures permits to assume

that these alternatives are $\{x_1, \dots, x_n\}$. Then

$$\begin{aligned}
\frac{1}{k} \sum_{j=1}^k \mathcal{C}_s(M^{-j}) &= \frac{1}{k} \left(\sum_{j=1}^n \mathcal{C}_s(M^{-j}) + \sum_{j=n+1}^k \mathcal{C}_s(M^{-j}) \right) = \\
&= \frac{1}{k} \left(n \frac{C_{n-1}^s}{C_{k-1}^s} + (k-n) \frac{C_n^s}{C_{k-1}^s} \right) = \\
&= \frac{1}{k} \frac{n \frac{(n-1)!}{s!(n-1-s)!} + (k-n) \frac{n!}{s!(n-s)!}}{\frac{(k-1)!}{s!(k-1-s)!}} = \\
&= \frac{\frac{n!(n-s)}{s!(n-s)!} + \frac{n!(k-n)}{s!(n-s)!}}{\frac{k!(k-s)}{s!(k-s)!}} = \frac{\frac{n!}{s!(n-s)!}}{\frac{k!}{s!(k-s)!}} = \frac{C_n^s}{C_k^s} = \mathcal{C}_s(M)
\end{aligned}$$

□

3.1. Necessary and sufficient conditions for the s -Simple ACM

We characterize the s -Simple ACM in terms of the following two properties:

Definition 3. An approval consensus measure μ verifies:

- i) s -triviality if and only if for each approval profile M on X and each $I \subseteq X$ with cardinality s ,

$$\mu(M^I) = \begin{cases} 1 & \text{if the agents are unanimous on every alternative in } I, \\ 0 & \text{otherwise.} \end{cases}$$

- ii) s -reducibility if and only if for each approval profile M on X ,

$$\mu(M) = \frac{1}{C_k^s} \sum_{\substack{I \subseteq X \\ |I|=s}} \mu(M^I)$$

For a given s , s -triviality means that the application of μ to any s -restricted profile behaves in a trivial manner: the profiles have an approval consensus measure of 1 exactly when they are unanimous, the alternative being approval consensus measure of 0. And s -reducibility means that the approval consensus measure of a profile is the average of the approval consensus measures of its s -restricted profiles.

Theorem 1. Let μ be an approval consensus measure on X . Then $\mu = \mathcal{C}_s$ if and only if μ verifies s -triviality and s -reducibility.

Proof. Let $X_u \subseteq X$ denote the set of alternatives for which the agents have an unanimous opinion, thus $n = |X_u|$.

Let us first prove that \mathcal{C}_s verifies s -triviality and s -reducibility for each $k \geq s > 0$. For each approval profile M on X and each $I \subseteq X$ with cardinality s , let $n_I \leq s$ denote the number of alternatives of I for which the agents have an unanimous opinion. Then $\mathcal{C}_s(M^I) = \frac{\binom{n_I}{s}}{\binom{s}{s}}$ equals 1 iff $n_I = s$, i.e., iff M^I is unanimous; and it equals 0 otherwise.

To check for s -reducibility we just need to observe that $\sum_{\substack{I \subseteq X \\ |I|=s}} \mathcal{C}_s(M^I) = C_n^s$ because s -triviality implies that this sum is precisely the number of subsets of X with cardinality s , such that the agents have an unanimous opinion on their alternatives. In other words: It is the number of combinations of n distinct elements taken s at a time. Formally:

$$\sum_{\substack{I \subseteq X \\ |I|=s}} \mathcal{C}_s(M^I) = \sum_{\substack{I \subseteq X_u \\ |I|=s}} \mathcal{C}_s(M^I) + \sum_{\substack{I \not\subseteq X_u \\ |I|=s}} \mathcal{C}_s(M^I) = C_n^s + 0$$

Conversely, let μ be an approval consensus measure that verifies s -triviality and s -reducibility. Due to s -triviality,

$$\sum_{\substack{I \subseteq X \\ |I|=s}} \mu(M^I) = |\{I \subseteq X_u : |I| = s\}| = C_n^s$$

and now s -reducibility yields

$$\mu(M) = \frac{C_n^s}{C_k^s} = \mathcal{C}_s(M)$$

□

Focal instances of s yield special cases of s -simple ACMs that we proceed to investigate further. The case $s = 1$ will be called the *simple* ACM. The case $s = k$ will be called the *trivial* ACM: it is equal to 1 when the profile is unanimous, and 0 otherwise. In between we have the case $s = 2$, that we call the *Pareto* ACM.

3.2. The simple ACM. Further properties

The expression of the 1-simple (or simple) ACM is $\mathcal{C}_1(M) = \frac{n}{k}$ for each approval profile M on a set with k alternatives. It measures the probability that all agents unanimously approve/disapprove of a randomly selected alternative. Theorem 2 below gives an alternative characterization of \mathcal{C}_1 in terms of the following property:

Definition 4. An approval consensus measure μ verifies *convexity* if and only if for each approval profile M on X , and each decomposition of M into two subprofiles M_1 and M_2 , with k_1 and k_2 columns respectively,

$$\mu(M) = \frac{k_1 \mu(M_1) + k_2 \mu(M_2)}{k}$$

Convexity means that the measure of a profile is a weighted average of the measures of any decomposition into subprofiles, their weights being given by their respective relative sizes. A routine checking shows that \mathcal{C}_1 verifies convexity.

Theorem 2. Let μ be an approval consensus measure on X . The following statements are equivalent:

1. μ is the simple approval consensus measure.
2. μ verifies 1-triviality and 1-reducibility.
3. μ verifies 1-triviality and convexity.

Proof. Due to Theorem 1, we only need to check that conditions 1 and 3 are equivalent. We already know that 1 implies 3. Let us assume that μ verifies 1-triviality and convexity. We proceed by induction on k to prove $\mu(M) = \frac{n}{k}$ for each approval profile M on a set with k alternatives. The case $k = 1$ holds by 1-triviality. Assume that the statement is true for sets with k alternatives or lesser. Let $X = \{x_1, \dots, x_k, x_{k+1}\}$ be a set with $k + 1$ alternatives, and let M be an approval profile on X . Denote by M_1 the restriction of the profile M to $\{x_1, \dots, x_k\}$. Convexity assures

$$\mu(M) = \frac{k \mu(M_1) + \mu(M^{k+1})}{k + 1}$$

and the induction hypothesis yields

$$\mu(M) = \frac{k \mathcal{C}_1(M_1) + \mathcal{C}_1(M^{k+1})}{k + 1}$$

Because \mathcal{C}_1 verifies convexity, the latter expression boils down to $\mu(M) = \mathcal{C}_1(M)$. \square

3.3. The Pareto ACM. Further properties

The expression of the 2-Simple ACM is $\mathcal{C}_2(M) = \frac{n(n-1)}{k(k-1)}$ for each approval profile M on a set with k alternatives. It measures the probability that a shrink of the set of alternatives to two randomly chosen alternatives produces full consensus among the agents. This is related to a possible adapted variation of Bosch’s Pareto measure [5, pp. 81-82], which “is based on the number of pairs on which the voters agree”: here we interpret that the voters agree on x_i and x_j when they are unanimous on x_i and also on x_j , i.e., when the voters have exactly the same opinion about which of the two alternatives must be approved.

A routine checking proves the following relationships between \mathcal{C}_1 and \mathcal{C}_2 for each approval profile M on a set with k alternatives:

$$\mathcal{C}_2(M) = \mathcal{C}_1(M) \frac{k \mathcal{C}_1(M) - 1}{k - 1}$$

$$\mathcal{C}_1(M) = \frac{1 - \sqrt{4 \mathcal{C}_2(M) k (k - 1) + 1}}{2k}$$

Convexity explains the behavior of the simple approval consensus measure when a new alternative is added without affecting the agent’s opinion on the original alternatives. The following result gives the corresponding exact relation for the Pareto ACM, which can be checked in a straightforward manner:

Proposition 2. Let $X = \{x_1, \dots, x_k\}$ be a set with k alternatives, and let M be an approval profile on X such that the agents are unanimous on n alternatives. Denote by \tilde{M} an extension of the profile M to $\tilde{X} = \{x_1, \dots, x_k, x_{k+1}\}$. Then

$$\mathcal{C}_2(\tilde{M}) = \frac{k(k-1)\mathcal{C}_2(M) + 2n\mathcal{C}_1(\tilde{M}^{k+1})}{k(k+1)} = \frac{C_k^2 \mathcal{C}_2(M) + n \mathcal{C}_1(\tilde{M}^{k+1})}{C_{k+1}^2}$$

4. The class of s -Simple Tanimoto ACMs. Characterization and properties

In this section we introduce and explore a variant of s -Simple ACMs that satisfies an independence of irrelevant alternatives criterion. It is inspired in the Tanimoto similarity index (see [12] and [13]). At this point we have

to introduce some additional notations. An alternative x_j is called *irrelevant* on profile M if all agents disapprove of it, i.e. $M_{ij} = 0$ for all $i \in \mathbf{N}$, otherwise it is relevant. An approval profile M is *irreducible* if it lacks irrelevant alternatives. This means that each alternative is approved of by at least one agent. Given a non-unanimous approval profile M , we denote by M_R its unique irreducible subprofile, i.e., M_R arises from M after removing irrelevant alternatives.

We now introduce the class of Simple Tanimoto approval consensus measures that depend on a parameter s . Intuitively, for each non-unanimous profile they measure the probability that a randomly chosen s -set of relevant alternatives are approved by all agents. Formally:

Definition 5. The s -Simple Tanimoto ACM is the mapping $\mathcal{T}_s : \mathbb{M}_{N \times k} \rightarrow [0, 1]$ given by

$$\mathcal{T}_s(M) = \begin{cases} 1 & \text{if } M \text{ is unanimous} \\ 0 & \text{if } M \text{ is not unanimous and } k - n_0 < s \\ \frac{C_{n_1}^s}{C_{k-n_0}^s} & \text{otherwise} \end{cases} \quad (2)$$

for each approval profile M on k alternatives.

Since $k - n_0 \geq n_1$ it is immediate to check that the above expression provides an approval consensus measure.

Example 3. Continuing with Examples 1 and 2, one has $\mathcal{T}_1(M) = \frac{1}{3}$ and $\mathcal{T}_2(M) = \mathcal{T}_3(M) = \mathcal{T}_4(M) = 0$.

Now we proceed as in the previous section: we enumerate some general properties of the class of s -Simple Tanimoto ACMs and then we give their axiomatic characterization. Therefore, let M denote an approval profile.

1. $\mathcal{T}_1(M) \geq \mathcal{T}_2(M) \geq \dots \geq \mathcal{T}_k(M)$. If M is unanimous, $k - n_0 < s$ or $n_1 < s$, then the chain of inequalities is trivial. Otherwise $s \leq n_1 < k - n_0$ yields

$$\frac{\mathcal{T}_s(M)}{\mathcal{T}_{s-1}(M)} = \frac{n_1 - s + 1}{k - n_0 - s + 1} < 1$$

2. $\mathcal{C}_s(M) \geq \mathcal{T}_s(M)$. We only discuss non-trivial cases, that is, $s \leq n_1 < k - n_0$. Exploiting the inequality $\frac{x+1}{y+1} > \frac{x}{y}$ with $x, y > 0$ and $x < y$, we obtain

$$\frac{n-i}{k-i} > \frac{n-n_0-i}{k-n_0-i} = \frac{n_1-i}{k-n_0-i} \quad \text{for } 0 \leq i \leq s-1$$

From this, the assertion is straightforward.

3. The consensus measure does not change if agents are replicated in any number.
4. Suppose that M is not unanimous, and that an agent $N+1$ is added to the society \mathbf{N} . If this agent disapproves of all the alternatives in X , then the consensus measure is zero. It suffices to observe that the number of approved alternatives by unanimity is zero.

On the other hand, if the new agent approves of all the alternatives in X the consensus measure does not rise. Let \bar{M} be the expanded approval profile, then note that $\bar{n}_0 = 0 \leq n_0$, $\bar{n}_1 = n_1$ and $\bar{k} = k$. We have either $k - n_0 < s$ or $k - n_0 \geq s$. In the first case, since M is not unanimous, we get $\bar{n}_1 = n_1 < k - n_0 < s$ and then $\mathcal{T}_s(\bar{M}) = 0 = \mathcal{T}_s(M)$. In the second case, a simple computation arrives at:

$$\begin{aligned} \mathcal{T}_s(\bar{M}) &= \frac{n_1(n_1-1)\dots(n_1-s+1)}{k(k-1)\dots(k-s+1)} \\ &\leq \frac{n_1(n_1-1)\dots(n_1-s+1)}{(k-n_0)(k-n_0-1)\dots(k-n_0-s+1)} = \mathcal{T}_s(M) \end{aligned}$$

5. Suppose that M is not unanimous, that an alternative $k+1$ is added to set of alternatives X , and that this alternative is unanimously approved of by all agents. If the introduction of the new alternative does not affect the agents' assessments of the original alternatives, then the consensus measurement of the enlarged sets does not decrease.

Let \tilde{M} be the extended approval profile, we then have $\tilde{n}_0 = n_0$, $\tilde{n}_1 = n_1 + 1$ and $\tilde{k} = k + 1$. We distinguish two cases:

- Case $0 < k - n_0 < s$ or $n_1 < s$. This implies $\mathcal{T}_s(M) = 0$ and then $\mathcal{T}_s(\tilde{M}) \geq \mathcal{T}_s(M)$.
- Case $k - n_0 \geq s$ and $n_1 \geq s$. This implies $\tilde{k} - \tilde{n}_0 > s$ and $\tilde{n}_1 > s$.

A simple computation gives

$$\begin{aligned}\mathcal{T}_s(\tilde{M}) &= \frac{C_{\tilde{n}_1}^s}{C_{\tilde{k}-\tilde{n}_0}^s} = \frac{C_{n_1+1}^s}{C_{k-n_0+1}^s} \\ &= \frac{n_1+1}{k-n_0+1} \frac{k-n_0+1-s}{n_1+1-s} \mathcal{T}_s(M)\end{aligned}\tag{3}$$

Now, using the fact that $\frac{x}{y} > \frac{x+1}{y+1}$ when $x, y > 0$ and $x > y$, we infer

$$\frac{k-n_0+1-s}{n_1+1-s} > \frac{(k-n_0+1-s)+1}{(n_1+1-s)+1} > \dots > \frac{k-n_0+1}{n_1+1}\tag{4}$$

Combining (3) and (4) we obtain the desired assertion $\mathcal{T}_s(\tilde{M}) > \mathcal{T}_s(M)$.

6. The s -Simple Tanimoto ACM satisfies an independence of irrelevant alternatives criterion. Suppose that M is not unanimous, that an alternative $k+1$ is added to set of alternatives X , and that this alternative is unanimously disapproved of by all agents. Then the consensus measurement does not change. This is trivial because $\tilde{n}_1 = n_1$ and $\tilde{k} - \tilde{n}_0 = k - n_0$.
7. Convergence to full unanimity can be established if we repeatedly introduce alternatives that are unanimously approved. Formally: Suppose that alternatives $k+1, \dots, k+t$ are added to the set of alternatives X , and that each alternative is unanimously approved by all agents. If the introduction of new alternatives does not affect the agents' assessments of past sets of alternatives, then the consensus measurement of the extended approval profiles $\tilde{M}^{(t)}$ approaches 1 when t tends to infinity. The argument is analogous to that of Section 3 and so omitted.

We now characterize the s -Simple Tanimoto ACMs. We first introduce the following definitions.

Definition 6. An approval consensus measure μ verifies:

- i) *Independence of irrelevant alternatives* if and only if $\mu(M) = \mu(M_R)$ for any non-unanimous profile.

- ii) *s-reducibility on irreducible profiles* if μ is a *s*-reducible measure on the set of irreducible approval profiles, that is: for each irreducible approval profile M on X

$$\mu(M) = \frac{1}{C_k^s} \sum_{\substack{I \subseteq X \\ |I|=s}} \mu(M^I)$$

- iii) *s-nullity* if and only if the consensus measurement of any non-unanimous profile that approves of less than s alternatives is zero.

The first property reveals that the unanimously disregarded alternatives do not play any role in the consensus measurement. The second one is a weak version of *s*-reducibility at Section 3.2. Combining both properties we infer that the consensus measurement of a profile M only depends on its *s*-restricted profiles that are irreducible. The last property means that the consensus only can be positive if the agents approve of at least s alternatives.

Theorem 3. Let μ be an approval consensus measure on X . Then $\mu = \mathcal{T}_s$ if and only if μ verifies *s*-triviality, independence of irrelevant alternatives, *s*-reducibility on irreducible profiles and *s*-nullity.

Proof. Given an approval profile, let $X_u(M)$, $X_1(M) \subseteq X(M)$ be the set of alternatives for which the agents have an unanimous opinion and the set of approved alternatives by unanimity, respectively. Thus $n = |X_u(M)|$, and $n_1 = |X_1(M)|$.

Let us first prove that \mathcal{T}_s verifies the four properties above. For each approval profile M and each $I \subseteq X$ with cardinality s , let n_0^I and n_1^I be the number of alternatives of I that all agents disapprove of and approve of, respectively. If M^I is unanimous then it is obvious that $\mathcal{T}_s(M^I) = 1$. If M^I is not unanimous, we have either $n_0^I > 0$ or $n_1^I = 0$. In the first case it must be the case that $|I| - n_0^I < s$ and then $\mathcal{T}_s(M^I) = 0$. In the second case we deduce $n_1^I < s$, therefore $C_{n_1^I}^s = 0$. Thus $\mathcal{T}_s(M^I) = 0$ irrespective of the case.

To prove that \mathcal{T}_s satisfies independence of irrelevant alternatives it suffices to note that an approval profile M and its associated irreducible profile M_R have the same number of approved alternatives and the same number of approved by unanimity alternatives.

Note that *s*-nullity is a simple consequence of the definition of \mathcal{T}_s .

We finally check for *s*-reducibility on irreducible profiles. Let M be an irreducible profile, since $n_0 = 0$ and $X_u(M) = X_1(M)$ we deduce by

s -triviality that

$$\mathcal{T}_s(M) = \frac{1}{C_k^s} \sum_{\substack{I \subseteq X(M) \\ |I|=s}} \mathcal{T}_s(M^I)$$

Conversely, let μ be an approval consensus measure that satisfies the four properties above. We can assume that M is not unanimous, because in other case any measure provides consensus one. By independence of irrelevant alternatives we have

$$\mu(M) = \mu(M_R),$$

where the cardinality of the set of evaluated alternatives in M_R is $k - n_0$. If $k - n_0 < s$ by s -nullity we deduce $\mu(M) = 0 = \mathcal{T}_s(M)$.

We now analyze the case $k - n_0 \geq s$. Due to s -reducibility on irreducible profiles we infer

$$\mu(M_R) = \frac{1}{C_{k-n_0}^s} \sum_{\substack{I \subseteq X(M_R) \\ |I|=s}} \mu(M_R^I)$$

Since $X_u(M_R) = X_1(M_R)$, using s -triviality we obtain

$$\sum_{\substack{I \subseteq X(M_R) \\ |I|=s}} \mu(M_R^I) = C_{n_1}^s,$$

and we then arrive at the desired assertion

$$\mu(M) = \frac{C_{n_1}^s}{C_{k-n_0}^s} = \mathcal{T}_s(M)$$

□

5. Illustrative real example

In order to perform a practical exploration of our indexes, in this section we develop an illustrative real example to visualize some differences among s -Simple approval consensus measures (s -SACM) and s -Simple Tanimoto consensus measures (s -STACM). Moreover, different values of the parameters are considered in this applied exercise.

We focus on the predictions about the 2012 presidential elections in USA made by polling agencies. For simplicity, we consider the forecast by three

polling outfits, namely, PPP (Public Policy Polling), Rasmussen reports, and SurveyUSA. Although these polls were performed nationwide, we only report on their predictions for the following states: California (CA), Colorado (CO), Connecticut (CT), New Jersey (NJ), Minnesota (MN), Missouri (MO), Ohio (OH) and Washington (WA).

Table 1 gathers the polls results: we adopt the convention that ‘1’, respectively ‘0’, means a prediction that the Democrats, respectively the Republicans, would win the state (see for example Colorado, where the PPP and SurveyUSA agencies predicted a victory of the Democrats but the Rasmussen agency predicted a victory of the Republicans).

Following the notation adopted in the paper, the number of states where the agencies unanimously predicted a Democrat, respectively Republican, victory are $n_1 = 6$ and $n_0 = 1$.

Table 1: Polls results

	State							
	CA	CO	CT	NJ	MN	MO	OH	WA
PPP(D)	1	1	1	1	1	0	1	1
Rasmussen	1	0	1	1	1	0	1	1
Survey USA	1	1	1	1	1	0	1	1
Total	3	2	3	3	3	0	3	3

Source: <http://www.realclearpolitics.com/polls/>

Table 2 shows that the degree of agreement about the electoral predictions among the states for s -SACM and s -STACM decreases as parameter s increases.

Remark 2. We insist that this section is a purely technical exercise. More importantly, we do not advocate for the use of the s -STACM index in our example: it makes no sense to stipulate ‘irrelevant alternatives’ in the current setting. In addition, observe that the choice of the ‘0’ label is no more than a technical code and can be freely attached.

Table 2: Measures of the cohesiveness of the electoral surveys

	s							
	1	2	3	4	5	6	7	8
s -SACM	$\frac{7}{8}$	$\frac{3}{4}$	$\frac{5}{8}$	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0
s -STACM	$\frac{6}{7}$	$\frac{5}{7}$	$\frac{2}{3}$	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	0	0

6. Final comments and concluding remarks

In this paper we explore the problem of measuring the degree of cohesiveness in a setting where experts express their opinions on alternatives in a dichotomous way. To that purpose we define the novel concept of approval consensus measures (ACMs). In this initial contribution to the topic we give the first necessary and sufficient conditions that characterize relevant primary classes of ACMs, namely s -Simple ACMs and s -Simple Tanimoto ACMs. Because we elaborate on approval consensus measures that adapt to simple real-world situations, an obvious future development is the study and characterization of other consensus measures that can be used to analyze complex situations more faithfully. Moreover, our approach (or other similar proposals) could be applied in conflict resolution as a component of a decision aid system.

It may appear that our model should bear comparison with other proposals in the literature on consensus measures. In this regard, there is the immediate intuition that dichotomous opinions on the elements in X can be identified with complete preorders or weak orders (i.e., complete and transitive binary relations) on that set of alternatives. There are at least two reasons to dissent from such view. A general objection is that by expressing a dichotomous opinion only a part of the actual preference can be revealed, provided that the agent has actually used such procedure to achieve his or her classification of the alternatives. This may easily introduce arbitrariness in any subsequent measure based on the revealed preference. A more particular objection is that our model is not limited to choice situations where the agent must prioritize some alternatives, as the example in section 5 has shown. Nevertheless, let us consider the only proposal to measure consensus in collectives with weak orders on the alternatives that we are aware of,

namely García-Lapresta and Pérez-Román [9, pp. 213-234]. Their procedure is based on the adapted version of distance between linear orders to weak orders, which requires to know the entire individual preference of each agent. Here we find a technical reason that differentiates the latter model from ours, which only needs knowing aggregate opinions on each issue. For many situations (e.g., opinion polls or surveys) the use of aggregate data seems more realistic.

Acknowledgements

The authors thank the Editor of Information Sciences, three anonymous referees and Francisco Herrera for their valuable comments and suggestions. J. C. R. Alcantud, R. de Andrés Calle and J. M. Cascón acknowledge financial support by the Spanish Ministerio de Economía y Competitividad (Projects ECO2009-07682 and ECO2012-31933; Project ECO2012-32178; and Project CGL2011-29396-C03-02, respectively). J. M. Cascón acknowledges financial support by Junta de Castilla y León under, Project SA266A12-2.

References

- [1] J. Alcalde and M. Vorsatz. The measurement of consensus: An axiomatic analysis. FEDEA Working Paper 2008-28, 2008.
- [2] J. C. R. Alcantud, R. de Andrés Calle, and J. M. Cascón. Consensus and the act of voting. *Journal of Microeconomics*. Forthcoming.
- [3] J. C. R. Alcantud, R. de Andrés Calle, and J. M. Cascón. A unifying model to measure consensus solutions in a society. *Mathematical and Computer Modelling*, 2011. DOI: 10.1016/j.mcm.2011.12.020.
- [4] C. Alós-Ferrer and D.-G. Granić. Two field experiments on approval voting in Germany. *Social Choice and Welfare*, 39:171–205, 2012.
- [5] R. Bosch. *Characterizations of Voting Rules and Consensus Measures*. PhD thesis, Tilburg University, 2005.
- [6] S. J. Brams and P. C. Fishburn. Approval voting. *American Political Science Review*, 72:831–847, 1978.

- [7] S. J. Brams and P. C. Fishburn. *Approval Voting*. Birkhauser, Boston, 1983.
- [8] M. A. Fligner, J. S. Verducci, and P. E. Blower. A modification of the Jaccard-Tanimoto similarity index for diverse selection of chemical compounds using binary strings. *Technometrics*, 44:110–119, 2002.
- [9] E. Herrera-Viedma, J. L. García-Lapresta, J. Kacprzyk, H. Nurmi, M. Fedrizzi, and S. Zadrozny, editors. *Consensual Processes*. Springer-Verlag, Berlin, 2011.
- [10] J-F. Laslier and M. R. Sanver, editors. *Handbook on Approval Voting*. Springer-Verlag, Berlin Heidelberg, 2010.
- [11] D. S. Rodríguez-Baena, A. J. Pérez-Pulido, and J. S. Aguilar-Ruiz. A biclustering algorithm for extracting bit-patterns from binary datasets. *Bioinformatics*, 27:2738–2745, 2011.
- [12] D. J. Rogers and T. T. Tanimoto. A computer program for classifying plants. *Science*, 132:1115–1118, 1960.
- [13] T. T. Tanimoto. An elementary mathematical theory of classification and prediction. Technical report, IBM, 1958.
- [14] L. Touil, F. Guesmi, K. Fares, and A. Ferchichi. Genetic diversity of some mediterranean populations of the cultivated alfalfa (*Medicago sativa* L.) using ISSR markers. *Biotechnology*, 7:808–812, 2008.