Natural Rules for Optimal Debates
(Preliminaries for a Combinatorial Exploration)

Yann Chevaleyre and Nicolas Maudet

Abstract

Two players hold contradicting positions regarding a given issue, which depends on a (fixed) number of aspects or criteria they both know. Suppose, as a third-party, that you want to make a decision based on what will report the players. Unfortunately, what the players can communicate is limited. How should you design the rules of your protocol so as to minimize the mistakes induced by these communication constraints? This paper discusses this model originally due to [2] in a specific case variant, and introduces preliminary results of a combinatorial exploration of this problem.

1 Introduction

The situation is the following: Two debaters have contradicting positions regarding a given issue, which depends on a (fixed) number of aspects, or criteria. The value of these aspects being given, there is common knowledge of the decision rule which will eventually selects the outcome (for instance, the majority). They both know what the "actual" state of the world is (so they both know who should be the actual winner). Unlike the players, a third-party agent is not aware of the real state. Now they exchange arguments (e.g. claiming that a given aspect of the state supports their opinion) during a debate, with the aim of convincing this external observer of their position. Of course, what makes the problem interesting is that there is a limitation on the number of communications they can make.

This problem introduced by Glazer and Rubinstein in [2] is a *mechanism design* problem: Designing the rules of the debates such that the probability for the observer to reach the "right" (the one that would be taken with full knowledge of the state) decision is actually maximal.

Basically, a debate consists of two elements:

- *procedural rule*– specifies the protocol constraining the arguments that the debater agent can raise (here some assumptions are made: an agent can just raise arguments supporting his favoured outcome, and nothing else);
- *persuasion rule*– specifies how the observer should make his decision based on the arguments advanced during the debate.

As far as the procedural rules are concerned, the authors discuss three canonical types of debates: (i) only one debater is allowed to speak (*single-speaker* *debate*); (ii) two debaters argue simulatenously (*simultaneous debate*), and (iii) debaters raise sequentially arguments (*sequential debate*). In [2], the authors investigate the three types of debate in the restricted 5-aspects setting (where the numbers of arguments to be communicated is limited to 2), and show in particular that the optimal rule in this context is necessarily sequential. In this preliminary work, we want to initiate the investigation of the extremal behaviour of this problem (when n becomes very large), and we start with the simple case where only one player is allowed to raise arguments (*single-speaker debate*).

The rest of the paper is as follows. In the next section we introduce the basic definitions that will be used throughout this paper. Section 3 then presents the analysis of different sorts of "natural" persuasion rules that a designer may wish to use in order to make his decision . By natural we mean that they can be simply stated in natural language by the designer. We provide an analytical analysis of two very simple rules ("give me any set of size k", and "give me that set"), and offer some preliminary insights of the behaviour of the rules that fall within the vast region in between. These latest findings are mostly supported by experimentations. Section 4 concludes and draws some connections with related works.

2 Basic Definitions

In this section we introduce more formally the problem as stated by [2], sometimes slightly deviating from the original version to introduce are own notations.

A *state* is a binary vector $\{0,1\}^n$, and each player (0,1) "controls" the bits (arguments) of his colour (that is, he cannot lie and cannot play the bits of the other player). We say that a state is an *objectively winning state* for agent x if a fully-informed designer would declare x winner in that state. For instance, the state $\{0,1,1,1,1\}$ means that the first argument is in favour of agent 0, while all the others are supporting agent's 1 view. This is an objectively winning situation for agent 1 (we assume the majority rule).

Typically, only k bits of communication will be allowed in our debates (with $k < n/2$ for obvious reasons as we consider the majority rule). A *persuasion rule* is defined in extension as a set

$$E = \{S_1, S_2, \dots, S_n\}$$

where each set S_i is a subset of $[n]$ of size k (k-subset). Such a rule must be interpreted as follows: "I would declare you winner if you can raise all the arguments contained in S_1, or all the arguments contained in S_2, etc.". For instance, the persuasion rule $E = \{\{1,2\},\{2,3\}\}$ means that the agent must either show arguments 1 and 2, or 2 and 3 (but 1 and 3 is not sufficient) to be declared winner. In this paper we will be interested in persuasion rules that can be simply stated in natural language (typically because they exploit some properties of the k-subsets composing the rules).

The *error ratio* (ϵ) induced by a rule is the number of states where you would take an erroneous decision when compared to what a fully-informed designer would do (n_{err}), normalised over all possible states. If you take a closer look at the notion of error, it actually occurs that two types of errors can be distinguished:

- *minority errors*, correponding to states where you would declare an agent winner, although this agent doesn't hold a winning position
- *majority errors*, corresponding to states where you would declare an agent loser, although the state is objectively winning for him.

Take the example given above, and assume a 5-bits debate. In states $\{1,1,0,0,0\}$ and $\{0,1,1,0,0\}$, agent 1 can convince the designer despite the state being objectively losing for him. On the other hand, in states $\{0,1,0,1,1\}$, $\{1,0,1,0,1\}$, $\{1,0,1,1,0\}$, and $\{1,0,1,1,1\}$ agent cannot convince the designer that its position is winning. This makes 6 errors overall (2 in favour of agent 1, 4 in favour of the other agent). Although, as correctly noticed by a reviewer, one type of error is the dual of the other (a minority error for one agent is a majority error for the other agent; or, to put it differently, any error is either a minority error for one agent or a minority error for the other agent), it is still useful to distinguish both types. The main reason is that it provides some information concerning which agent is favoured by a given rule.

In the following we will also make use of some additional notions. We say that a persuasion rule is *covered* by a state vector when at least one of its composing rule is covered by that state vector, that is when any argument required by that set is in that set. In these terms, the optimization problem we are faced with is to find the persuasion rule that will minimize the covering over the set of vectors containing $[k, \frac{n}{2}[$ bits (objectively losing situations), while maximizing the covering over the universe of vectors containing $n/2$ or more bits (objectively winning situations). We shall note these two measures respectively c_m and c_M from now on.

It is worth noting that in general (for $k \leq n/2$) the following holds:

$$n_{err} = c_m + (2^{n-1} - c_M)$$

The number of errors is simply the number of covering minority states, added up to the number of majority states (2^{n-1}) that are not covered by the rule.

3 Natural Rules

In this section we discuss the properties of some natural persuasion rules. Natural must be understood here as the fact that they can be simply stated in natural language by the designer (which does not necessarily imply that E will exhibit a simple structure in its extensive form). We refer the reader to [4] for an enlightning discussion on that topic. There are many "natural" rules you

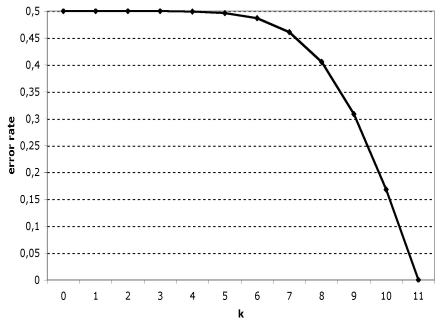

Figure 1: Error ratio induced by the "give me any set of size k" rule ($n = 20$)

can possibly come up with, and some examples are given in [2], like for instance "give me k adjacent bits". In what follows, we first discuss two very (arguably, the most) simple rules, before moving on to the general case lying between these two extreme rules.

3.1 "Give me *any* of size k"

We start with what may be the simplest rule, simply enonced as follows: "give me any subset of size k". Or maybe even more naturally as "give me k bits", without any further constraints. In other words, the set E would consist of the set exhausting any k-subsets of $\{0,1\}^n$. What would be the error induced by this rule? Note first that the majority error is bound to be 0 when $k \in [1, \frac{n}{2}]$. In general, the overall number of errors would then be equal to the number of losing situations covered by the rule (c_m). Take t as being the number of bits to still be placed to make a losing situation once you have covered the rule. There are

$$n_{err} = c_m = \sum_{t=k}^{\lfloor n/2 \rfloor} \binom{n}{t}$$

such situations, that is, the number of errors is given by the sum of binomial coefficients from k to $n/2$. This means that this rule is pretty ineffective: only when the number of bits allowed to be communicated becomes very close to $n/2$ does it give a good error ratio (see Fig. 1). And indeed, if you were allowed to ask the agent to communicate any number of bits, this is the perfect rule you would of course use: by requesting the agent to put forward $n/2$ aspects in favour of his view, you are sure that no agent can fool you in a losing state, while not missing any winning state at the same time.

3.2 "Give me *that* set"

In that case we assume that the designer can ask the agent to simply give just one set ($|E| = 1$), of arbitrary size k. (We assume n to be odd.) The minority and majority covering are as follows:

$$c_m = \sum_{i=0}^{\lfloor n/2-k \rfloor} \binom{n-k}{i}$$

$$c_M = \sum_{i=\lceil n/2-k \rceil}^{n-k} \binom{n-k}{i}$$

In that case, we have $c_M \geq c_m$.
Observe that $c_M + c_m = 2^{n-k}$, hence we have

$$\begin{aligned} n_{err} &= c_m + 2^{n-1} - (2^{n-k} - c_m) \\ &= 2c_m + (2^{n-1} - 2^{n-k}) \end{aligned}$$

The error ratio is then

$$\begin{aligned} \epsilon &= \frac{2c_m + (2^{n-1} - 2^{n-k})}{2^n} \\ &= \frac{c_m}{2^{n-1}} + \frac{1}{2} - 2^{-k} \end{aligned}$$

We will now show that this is an increasing monotonic function.

Lemma 1 *For odd values of n and for $k \geq 1$, the error rate of the "give me* that *set" rule increases as k grows.*

Proof. Let n be odd. We will show that $\frac{n_{err}-2^{n-1}}{2} = c_m - 2^{n-k-1}$ is a increasing function of k. More precisely, we will show that the value c_m decreases as k grows, but that 2^{n-k-1} decreases faster, thus ensuring that n_{err} increases as k grows. To achieve this, it suffices to show that $c_m^k - c_m^{k+1} \leq 2^{n-k-1} - 2^{n-k-2} \leq 2^{n-k-2}$. In the following, we make use of the binomial formula : $\binom{x}{y} = \binom{x-1}{y-1} + \binom{x-1}{y}$.

$$
\begin{aligned}
c_m^k - c_m^{k+1} &= \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor - k} \binom{n-k}{i} - \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor - k - 1} \binom{n-k-1}{i} \\
&= \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor - k - 1} \left\{ \binom{n-k}{i} - \binom{n-k-1}{i} \right\} + \binom{n-k}{\lfloor \frac{n}{2} \rfloor - k} \\
&= \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor - k - 1} \binom{n-k-1}{i-1} + \binom{n-k}{\lfloor \frac{n}{2} \rfloor - k} \\
&= \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - k - 2} \binom{n-k-1}{i} + \binom{n-k}{\lfloor \frac{n}{2} \rfloor - k} \\
&= \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - k - 2} \binom{n-k-1}{i} + \binom{n-k-1}{\lfloor \frac{n}{2} \rfloor - k - 1} + \binom{n-k-1}{\lfloor \frac{n}{2} \rfloor - k} \\
&= \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor - k} \binom{n-k-1}{i}
\end{aligned}
$$

First, it can be easily verifyied that $\lfloor \frac{n}{2} \rfloor - k \leq \left\lfloor \frac{n-k-1}{2} - 1 \right\rfloor$ for all $k \geq 1$ and $n \geq 1$. Exploiting the fact that $\sum_{i=0}^{\lfloor \frac{x-1}{2} \rfloor} \binom{x}{i} \leq 2^{x-1}$ for any $x \in \mathbb{N}$, and substituting x with $n-k-1$ we can now write the following, which completes the proof.

$$
c_m^k - c_m^{k+1} \leq \sum_{i=0}^{\lfloor \frac{n-k-1}{2} - 1 \rfloor} \binom{n-k-1}{i} \leq 2^{n-k-2}
$$

What does it tell us? Well, simply that if you have only one set to ask, then the smaller subset the better—in other words, just ask one bit. Of course you should not expect a very good error ratio (for instance, for $n = 20$ the error ratio starts at 40% for the singleton set and then tends towards 50% when k grows.)

3.3 The mostly unnatural region in between

So far we have studied two extreme natural cases: the case where only one set is asked, and the case when any k-subset is asked. It would be interesting to observe the behaviour of the persuasion when the number sets composing the persuasion lies in between (although it would be unlikely in general that the obtained rule would be natural). To do that, we first derived an analytical formula (shown in Appendix) representing the error rate in the general case. Unfortunately, deriving upper and lower bounds on such a formula proved to be difficult, and we did not get any satisfying result yet. For this reason, we choose to set up an experimental study, whose most strinking result is reported below ($n = 21$, a number $|E|$ of k-subsets is randomly generated to create a rule). Note that the axis representing the cardinality of E is logarithmic

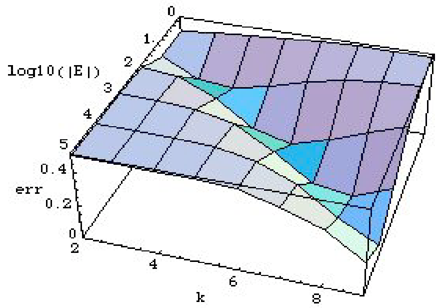

Figure 2: Error ratio of randomly generated rules of size $|E|$, depending on k

($log_{10}|E|$). This is so because we observed that the value of k for which the error is minimized depends logarithmically on the size of E. During all our experiments, we noticed that, while measuring the error rate as a function of k, having set the other parameters of the simulation, the error rate always decreases until k reaches a particular value which we will refer to as k_{opt} (this value depends on the other parameters), and then increases again. This can also apply to the extremal persuasion rules described above : consider the "give me that set" rule ($|E|$=1). Its error rate is best at $k = 1$, and then increases. Thus, k_{opt} for this rule equals to one. On the contrary, the error rate of the "give me any set" rule ($|E| = \binom{n}{n/2}$) always decreases as a function of k, until k reaches $\frac{n}{2}$. Thus, setting $k_{opt} = \frac{n}{2}$ also fits our framework. Thus, finding the value of k_{opt} is highly relevant to our problem. Further experimentations not reported here strongly suggest that this optimal value, for $n = 21$, is $2.log_{10}|E| + 2$. Fig. 2 shows the output of that particular experiment: for instance, when $log_{10}(|E|) = 1$ (10 k-subsets) we have $k_{opt} = 2$, when $log_{10}(|E|) = 2$, $k_{opt} = 4$, and when $log_{10}(|E|) = 5$, we finally have $k_{opt} = 9$ (the error ratio is then 8%).

3.4 Partitions of "Give me k bits within that set" sets

We now briefly discuss a case of historical interest, which represents a special (rather natural, see) family of rules for which the arguments can be, in some sense, clustered. Recall that in the case n=5, it has been proven by [2] that the optimal rule for this kind of debate is the rule consisting of asking the player to raise two bits either within the set $\{1, 2, 3\}$ or within $\{4, 5\}$. This rule could of course be represented in extension as $E = \{\{1, 2\}, \{2, 3\}, \{1, 3\}, \{4, 5\}\}$, but its appealing naturalness lies in the use of the IN-like operator which allows its compact representation, together with the fact that the obtained rule is a partition. To give us some hints as to whether that OR of IN partitions exhibited a good behaviour when the value of n becomes larger, we have conducted limited experiments. As means of example, the preliminary experiments that we ran with such partitions (with $n = 21$ and $k = 3$) give an error rate of 47% with 3 IN-subsets (we report here the optimal error ratio found after the random generation of such rules), of 36% with 5 IN-subsets, and 26% for 6 subsets, to eventually reach 21% for the partition consisting of exactly 7 subsets of size 3.

What these experiments show is that the error ratio constantly decreases when the number of IN-subsets augments. Although this may seem somewhat surprising at first sight, you have to notice that when the number of IN-subsets augments, the cardinality of E (defined in extension as k-subsets) actually decreases. This confirms the observations made in the previous subsection. Overall, the best rule possible belonging to this class seems then to be the rule composed of $\lfloor n/2 \rfloor$ IN-subsets of size k (or $k + 1$ for some number of agents $< n$) each, although we need more evidence to be able to firmly conclude. Note, however, that this is indeed in line with the result reported in [2].

4 Related and Future Works

Our ambition with this preliminary work is to initiate the study of the extremal behaviour of a mechanism problem introduced in [2]. The first results that we obtain here mainly concern two very simple kind of rule: "give me any set of size k", and "give me that set". Although the "give me any set of size k" rule is the only perfect rule when the communication is unrestricted, we show here that it is pretty ineffective in general when we put some limit on the number of bits to be transmitted. As for the "give me that set" rule, our result remarkably shows that the best strategy in that case is to simply ask the agent to report *one* bit (even if you are allowed more bits to be transmitted), as this is the optimal value of k in the case of E containing a single subset. These results are complemented with some experiments which show, for the instances of the problem that we studied, that the value of k for which the error is minimized depends logarithmically on the size of E. Finally, we briefly focused on the case of partitions of IN-subsets (where you ask the agent to raise k bits within that set, whatever the bits), which happens to be the generalization of the rule proven to be optimal for $n = 5$ by [2].

There are many ways to develop the line of research initiated with this preliminary work, the first being to refine our understanding of the behaviour of the type of rules discussed here. In particular, we have not precisely studied the influence of the way the subsets of E intersect with each other. We also aim at studying the other sort of debates introduced in [2], in particular the case of sequential debate which look very interesting.

There are also many possible connections to be made with others areas of research. We just mention here two obvious ones, as a way of conclusion.

As it happens, a persuasion rule is a *set system*, a combinatorial object well studied in the combinatorics litterature, see for instance [1]. However, to the best of our knowledge, the kind of properties that we study here are not classicaly investigated by that community.

Another area of research which seems (at first sight at least) pretty concerned with the problem discussed here is communication complexity. Communication complexity is concerned with the minimal number of bits that need to be exchanged in order to mutually compute some given function [3]. One main difference lies in the fact that agents are cooperative, whereas in our context they try to manipulate the designer to get the result of their wish. Also, communication complexity is typically concerned with finding bounds on the number of bits to be exchanged to be able to compute the function without any possible mistake, whereas here we assume to start with some communication constraints and try to design the rule so as to minimize the errors necessarily induced by these constraints.

Appendix: A Formula for the General Case

In this Appendix, we present the analytical formula of the total number of errors in the general case. As quoted in Section 3.3, we did not manage to derive satisfying bounds for this formula yet.

Suppose that $E = \{e_1, \dots, e_q\}$ where each e_i is a k-subset. In the following, $|\cup F|$ stands for $|\cup_{f \in F} f|$. Let us first compute c_m, the minority coverage.

$$
\begin{aligned}
c_m &= \sum_{i=1}^{q} \sum_{x=0}^{\lfloor \frac{n}{2} \rfloor - k} \binom{n-k}{x} - \sum_{i<j} \sum_{x=0}^{\lfloor \frac{n}{2} \rfloor - |e_i \cup e_j|} \binom{n - |e_i \cup e_j|}{i} + \dots \\
&= \sum_{F \subseteq E, F \neq \emptyset} \left[(-1)^{|F|-1} \sum_{x=0}^{\lfloor \frac{n}{2} \rfloor - |\cup F|} \binom{n - |\cup F|}{x} \right]
\end{aligned}
$$

$$
c_M = \sum_{F \subseteq E, F \neq \emptyset} \left[(-1)^{|F|-1} \sum_{x=\lceil \frac{n}{2} \rceil - |\cup F|}^{n-|\cup F|} \binom{n - |\cup F|}{x} \right]
$$

By adding both coverages, we get $c_m + c_M = \sum_{F \subseteq E, F \neq \emptyset} (-1)^{|F|-1} 2^{n-|\cup F|}$. The error is thus $n_{err} = c_m + 2^{n-1} - c_M = 2c_m + 2^{n-1} - \sum_{F \subseteq E, F \neq \emptyset} (-1)^{|F|-1} 2^{n-|\cup F|}$. Simplifying, we get the following general formula:

$$
n_{err} - 2^{n-1} = 2 \sum_{F \subseteq E, F \neq \emptyset} (-1)^{|F|} H_n(|\cup F|)
$$

Where $H_n(x) = 2^{n-x-1} - \sum_{t=0}^{\lfloor \frac{n}{2} \rfloor - x} \binom{n-x}{t}$.

References

[1] B. Bollobas. *Combinatorics: Set Systems, Hypergraphs, Families of Vectors, and Combinatorial Probability.* Cambridge University Press, 1986.

[2] J. Glazer and A. Rubinstein. Debates and decisions: On a rationale of argumentation rules. *Games and Economic Behaviour*, 36:158–173, 2001.

[3] E. Kushilevitz and N. Nisan. *Communication Complexity.* Cambridge University Press, 1997.

[4] A. Rubinstein. *Economics and Language.* Cambridge University Press, 2000.

Yann Chevaleyre
LAMSADE
75775 Paris Cedex 16, France
Email: `yann.chevaleyre@lamsade.dauphine.fr`

Nicolas Maudet
LAMSADE
75775 Paris Cedex 16, France
Email: `maudet@lamsade.dauphine.fr`