Optimal Role Assignment in Heterogeneous Juror Panels.

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Applied Mathematics at Delft University of Technology

Nam Dang

June 1, 2017

Faculty of Electrical Engineering, Mathematics and Computer Science · Delft University of Technology

Abstract

In this thesis we study the optimal role assignment in jury voting. Jurors have to decide between two states of nature. Jurors cannot directly observe the state of nature, but only a noisy signal, that is correlated with the true state of nature. Some jurors are better than the other, and higher ability jurors receive signals that more likely lead to the correct choice. Not all jurors vote simultaneously, and jurors that vote later are informed on what previous jurors have voted. We want to know what the optimal role assignment, if we have jurors with different ability levels.

In a two juror advisor-decider scheme, the first juror has the role of advisor and the second juror is the decider. The advisor passes his vote to the decider, and the decider's choice is the final decision the of the whole decision process. For random signals with linear-, β- and Gaussian probability densities, the best jurors should be the decider.

For three member casting vote schemes, the first two jurors vote simultaneously. If the decision is not unanimous, then there is no majority decision, and the casting juror breaks the tie. Besides receiving his own signal, the casting juror is informed on what the other jurors have voted. For random signals with linear-, β- and Gaussian densities, reliability is maximised, when the median ability juror has the casting vote.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Many decisions are decided by a voting mechanism. The reason why we vote is because some decisions are plagued with uncertainty and to minimise those effects, we take inputs from multiple parties. In this thesis two type of uncertainties are considered.

The first kind of uncertainty is when we have to decide on choices that are subjective by nature. This means that there is no “true” choice. Think a group of friends voting where they should eat or citizens voting for a presidential candidate. This type of voting is called *preference voting*, where voters a voter’s preference is a ranking of all possible alternatives. In an election, all voters pass their preference, and depending on the preferences, an outcome is generated. We consider two types of outcomes: a social preference, which is a ranking of all alternative, or a social choice, a single winner. An issue concerning many elections, is that voters are incentivised to vote strategically. In Chapter 2 we show that there is no “fair and reasonable” voting mechanism where voters do not vote strategically. These are the results from Arrow’s theorem and the Gibbard-Satterthwaite’s Theorem.

For the second type of uncertainty, the “true” choice does exist. The decision makers have to choose between two possible *states of nature*. They cannot directly observe the state of nature, but only a noisy signal correlated with the true state of nature. Think of a court case where jury members have to vote whether the defendant is guilty. We assume that the truth about the defendants guilt exist, but from the evidence we cannot be 100 percent certain what the correct verdict is. In another example, we consider a boxing match where the jury decides which fighter scored the most points. Jurors have to rely on human observations, which are not totally reliable. This type of voting is called *juror voting*.

The main motive of this thesis is the interesting result in the paper [1] by Alpern and Chen. The paper states that for three member casting vote schemes, the casting juror should be the median ability juror. In the model, jurors receive random signals taking values from a fixed interval. Each juror has a different ability level, where better jurors vote for the correct state with higher probability. Each juror has two possible probability density functions depending on the true state of nature. We want to investigate if we can generalise the result for a larger class of density functions. This is covered in Chapter 4.

Before investigating the three juror mechanism, we apply the same model to a simpler two juror advisor-decider mechanism. This is covered in Chapter 3. The advisor-decider mechanism is called roll call voting. There are similar papers on roll call voting, but with slightly different models. In [2], Alpern and Chen study a three juror model, with discrete signals and discrete abilities. The model in [3] jurors do not know their abilities, and the model in [4] uses, binary signals.

Chapter 2

Preference Voting

Democracies are currently the most popular governance systems in the world. One problem is that during the elections, many voters feel that they have to vote strategically. Instead of voting for their favourite candidate, it can be better to vote for the most popular lesser evil candidate to prevent an even worst candidate to win. The problem with strategic voting is that it is non-transparant and everyone has to make assumptions of what others are voting. It would be nice to have a fair voting mechanism that de-incentivises strategic voting. Maybe voters can give whole ranking of the candidates instead of only their most favourite? If the number of candidates is two, than that is trivial. In case the number of alternatives is three or more, designing such voting mechanisms turns out to be problematic. That is shown by Arrow's theorem and Gibbard-Satterthwaite's theorem, which are the main results of this section. The content of this section is based of [5] and [6].

Consider an election where voters have to pass their *preference* on a set of candidates, which will be called *alternatives* from now on. In many voting procedures, like political elections, a voters preference is a single favourite alternative. Here it is generalised, such that the preference is a ordering of all candidates that ranks each alternative from most favourite to least favourite. The kind of ordering that the voters will give is a *weak order* on A. Let $I = \{1, \ldots, n\}$ be the index set of all voters, A be the set of alternatives and $P(A)$ be the set of all weak orders on A. We will give the following definitions of a weak ordering, a voter's preference and a preference profile.

Definition 2.1 (Weak order)**.** We have that $\succeq$ is a weak order on A, if for any $a, b \in A$, it satisfies:

- *Transitivity*; If $a \succeq b$ and $b \succeq c$, then $a \succeq c$.
- *Totality*; We always have $a \succeq b$ or $b \succeq a$.

The set of all weak orders on A is $P(A)$.

Definition 2.2 ((Weak) Preference)**.** A *weak preference* $\succeq$ is a weak order on A. For $a, b \in A$, if a is preferred over b, it is denoted as: $a \succeq b$.

Definition 2.3 (Indifference $\simeq$)**.** A weak preference $\succeq$ is *indifferent* between a and b if $a \succeq b$ and $b \succeq a$. Indifference between a and b is abbreviated as $a \simeq b$.

Definition 2.4 (Strict preference $\succ$)**.** For $a, b \in A$, we say a is strictly preferred over b, if $b \not\succeq a$. This is also written as $a \succ b$.

Definition 2.5 (Just below $\succ!$)**.** We say b is just below a, if $a \succ b$ and there is no other c, such that $a \succ c \succ b$.

Lemma 2.1. *1. For any $a \in A$, we have $a \simeq a$.*

2. If $a \succ b$, then $a \succeq b$.

3. If $a \succ b$ and $b \succ c$, then $a \succ c$.

4. If $a \simeq b$ and $b \simeq c$, then $a \simeq c$.

5. For all $a, b \in A$ we have either $a \succeq b$ or $a \prec b$.

6. If $a \succ b$ and $b \succeq c$, then $a \succ c$.

Proof. 1. By totality we have $a \succeq a$, so $a \simeq a$.

2. $a \succ b$ means $b \not\succeq a$, so because of totality, we need to have $a \succeq b$.

3. We will proof the equivalent statement: If $c \succeq a$, then $b \succeq a$ or $c \succeq b$. Without loss of generality, assume $c \succeq a$ and $a \succ b$, because if $b \succeq a$, then we are done. By part 2 of this lemma, we have $a \succeq b$, so by transitivity $c \succeq b$.

4. $a \simeq b$ and $b \simeq c$ is short for $a \succeq b$, $b \succeq c$, $b \succeq a$ and $c \succeq b$. Transitivity implies $a \succeq c$ and $c \succeq a$ and that means $a \simeq c$.

5. It is clear by the definition of $\prec$.

6. By transitivity: $b \succeq c$ and $c \succeq a$, implies $b \succeq a$, which is equivalent to: $a \succ b$, implies $c \succ b$ or $a \succ c$. This means that $a \succ b$ and $b \succeq c$ (not $c \succ b$) implies $a \succ c$.

■

Definition 2.6 (Preference profile)**.** A preference profile π is a vector of preferences of all of the voters in I, i.e.

$$\pi = (\succeq_1, \ldots, \succeq_n) \in P(A)^I.$$

Additional useful functions are $\text{top}_i(B)$ and $\text{bot}_i(B)$. These functions give voter i's most and least favourite alternative from a subset of alternatives $B \subseteq A$.

Definition 2.7 (Top)**.** For $B \subseteq A$, $i \in I$ and $\succeq \in \pi$, we have that $a \in \text{top}_i^\pi(B)$, if $a \in B$ and there is no $b \in B$, such that $b \succ_i a$. We then say i has a at the top of B in π.

Definition 2.8 (Bottom)**.** For $B \subseteq A$, $i \in I$ and $\succeq \in \pi$, we have that $a \in \text{bot}_i^\pi(B)$, if $a \in B$ and there is no $b \in B$, such that $a \succ_i b$. We then say i has a at the bottom of B in π.

When all votes are in, the outcome of the election is determined. Depending on the goal of the election, two type of outcomes are considered. The first one is a single alternative (the winner), which is called the *social choice*. The second type of outcome is the social preference, which is an ordering on A. The outcomes are generated by:

- the *social welfare function* $F : P(A)^I \to P(A)$,
- the *social choice function* $f : P(A)^I \to A$.

2-1 Social Preference

Social welfare functions map the voting profile to a social preference. The outcome is a ranking of all alternatives. This can be useful in many situations. Let say there is an art contest and each jury members vote by giving a their preference on all the art pieces. Instead of only giving a the winner, we can give a whole ranking of who is first, second, third, etc. Social welfare functions can be given certain characteristics and several are given below:

- *Unanimity* means that if all voters have the same preference $\succeq$, then the social preference is $\succeq$.
- If F is *Pareto* and all voters strictly prefer a over b, then the social preference strictly prefers a over b.
- F is an *dictatorship* if there is a voter, the *dictator*, whose strict preferences are always the social strict preferences.
- *Independent of irrelevant relationships (IIR)* means that if we have two profiles that have the same preferences between a and b, then the social preferences between a and b, is the same. The social preference between a and b only depends on a voter's preference between a and b. Any other choices are not relevant.
- *Neutrality* means that the social choice does not depend on an alternative's name.
- *Pairwise neutrality* means that if we have two elections and all the preferences between (a,b) in one election are the same as the preferences between (c,d) in the other election, then the social preference between (a,b) is the same as the social preference between (c,d).

The precise definitions are:

Definition 2.9 (unanimity)**.** F is unanimous, if $F(\succeq,\ldots,\succeq) = \succeq$.

Definition 2.10 (Pareto)**.** F is Pareto, if $a \succ_i b$ for all $i \in I$, then $a \succ b$, where $F(\pi) = \succeq$.

Definition 2.11 (dictatorship)**.** F is a dictatorship, if for there is an $i \in I$, such that for any $\pi \in P(A)^I$, we have that $a \succ_i b$, implies $a \succ b$, where $F(\pi) = \succeq$. Voter i is called the *dictator*.

Definition 2.12 (independent of irrelevant relationships (IIR))**.** Consider two preference profiles π and π'. Let there be some $a, b \in A$, such that for all $i \in I$;

$$\begin{aligned} a \succeq_i b &\iff a \succeq'_i b \\ b \succeq_i a &\iff b \succeq'_i a. \end{aligned}$$

If F is IIR, then

$$\begin{aligned} a \succeq b &\iff a \succeq' b \\ b \succeq a &\iff b \succeq' a, \end{aligned}$$

where $F(\pi) = \succeq$ and $F(\pi') = \succeq'$.

Definition 2.13 (neutrality)**.** Let $\sigma(A) = \{\sigma(a_1), \dots, \sigma(a_n)\}$ be an arbitrary permutation of A and let π be some voting profile. The permuted voting profile π^σ of π, is defined such that for any $a, b \in A$ and all $i \in I$, we have that

$$a \succeq_i b \iff \sigma(a) \succeq_i^\sigma \sigma(b),$$
$$a \preceq_i b \iff \sigma(a) \preceq_i^\sigma \sigma(b).$$

If F is *neutral*, then

$$a \succeq b \iff \sigma(a) \succeq^\sigma \sigma(b),$$
$$a \preceq b \iff \sigma(a) \preceq^\sigma \sigma(b).$$

where $F(\pi) =\succeq$ and $F(\pi) =\succeq^\sigma$.

Definition 2.14 (pairwise neutrality)**.** Let π and π' be two preference profiles, such that for $a, b, c, d \in A$,

$$a \succeq_i b \iff c \succeq_i' d,$$
$$a \preceq_i b \iff c \preceq_i' d.$$

If F is pairwise neutral, then

$$a \succeq b \iff c \succeq' d,$$
$$a \preceq b \iff c \preceq' d,$$

where $F(\pi) =\succeq$ and $F\left(\pi'\right) =\succeq'$.

For convenience later on, we also name a special case of pairwise neutrality.

Definition 2.15 (pairwise neutrality for strict preferences)**.** Let π and π' be two preference profiles, where for $a, b, c, d \in A$, the preference relations of (a, b) and (c, d) are only strict. If F is pairwise neutral, then $a \succ_i b \iff c \succ_i' d$ for all $i \in I$, implies $a \succ b \iff c \succ' d$, where $F(\pi) =\succeq$ and $F\left(\pi'\right) =\succeq'$.

Pairwise neutrality and neutrality seem similar to each other. It is easy to see that pairwise neutrality implies neutrality. On the other hand, we can find an example in Example 1 that demonstrates that the converse is not always true.

Example 1 (Selecting teams)**.** *On a school yard, kids are deciding on splitting their group into two teams to play football. Two players are assigned the roll of captain and each of them take turn choosing from the remaining players who they want on their team. Each captain rank the players from most to least skilled. When a captain gets the turn to choose, he will select the best available player according to his ranking. The order in which the players are selected, defines a social choice.*

The captains are the set of voters $I = \{1, 2\}$ and all the other players are the set of alternatives A, where $|A| = m$ is even. A captain i has the preference $\succ_i \in P(A)$.

On turn 1, captain 1 selects player $a_1 = top_1(A)$ and on turn 2, captain 2 selects player $a_2 = top_2(A \setminus \{a_1\})$. Note that the index here is not the player's name, but it is an indication on which turn he or she is selected. Subsequently, on turn j captain i selects player $a_j = top_i(A \setminus \{a_1, a_2, \dots, a_{i-1}\})$, where *captain 1 can choose on the odd turns and captain 2 on the even turns. When all players are chosen, the preference is then* $F(\succ_1, \succ_2) = \succ$ *and* $a_1 \succ a_2 \succ \ldots \succ a_m$. *Note that* F *is neutral, because nowhere does the player's name play a role in the construction of the social preference.*

Take for example $A = \{a, b, c, d\}$ *and*

$$a \succ_1 b \succ_1 c \succ_1 d$$
$$c \succ_2 a \succ_2 b \succ_2 d.$$

The social preference is $a \succ c \succ b \succ d$. *Alternatively, we can have the following preference profile:*

$$a \succ_1' b \succ_1' c \succ_1' d$$
$$d \succ_2' c \succ_2' b \succ_2' a.$$

The social preference is then $a \succ' d \succ' b \succ' c$. *Note that for all* $i \in I$, *we have that* $b \succ_i c \iff b \succ_i' c$, *but* $c \succ b$ *and* $b \succ' c$, *so* F *is not IIR. If we set* $a = c$ *and* $b = d$ *in definition 2.14, we see that pairwise neutrality implies IIR, so a social welfare function that is not IIR cannot be pairwise neutral.*

We have given an example of a social choice function that is neutral, but not pairwise neutral by showing that it is not IIR. IIR and neutrality are actually sufficient for pairwise neutrality.

Proposition 2.2. *If* F *is neutral and IIR, then* F *is pairwise neutral.*

Proof. Assume F is neutral and IIR. Take some $a, b, c, d \in A$ and $\pi, \pi' \in P(A)^I$ such that for all $i \in I$:

$$\begin{aligned} a \succeq_i b &\iff c \succeq_i' d, \\ a \preceq_i b &\iff c \preceq_i' d. \end{aligned} \tag{2-1}$$

Take a permutation $\sigma(A)$ of A, such that $\sigma(a) = c$ and $\sigma(b) = d$. The permuted voting profile π^σ of π satisfies:

$$\begin{aligned} a \succeq_i b &\iff c \succeq_i^\sigma d, \\ a \preceq_i b &\iff c \preceq_i^\sigma d. \end{aligned} \tag{2-2}$$

for all $i \in I$. By neutrality we have that:

$$\begin{aligned} a \succeq b &\iff c \succeq^\sigma d, \\ a \preceq b &\iff c \preceq^\sigma d. \end{aligned} \tag{2-3}$$

where $F(\pi) = \succ$ and $F(\pi^\sigma) = \succ^\sigma$. Note that:

$$\begin{aligned} a \succeq_i' b &\iff c \succeq_i^\sigma d, \\ a \preceq_i' b &\iff c \preceq_i^\sigma d, \end{aligned} \tag{2-4}$$

so IIR implies

$$\begin{aligned} a \succeq b' &\iff c \succeq^\sigma d, \\ a \preceq b' &\iff c \preceq^\sigma d, \end{aligned} \tag{2-5}$$

where $F(\pi') = \succ'$. Formulas (2-3) and (2-4) imply

$$a \succeq b \iff c \succeq' d,$$
$$a \preceq b \iff c \preceq' d.$$

■

Other properties that look similar are unanimity and Pareto. It is clear that Pareto, implies unanimity. The converse is true if there is also IIR.

Proposition 2.3. *If F is unanimous and IIR, then F is Pareto.*

Proof. Consider $a, b \in A$ and $\pi \in P(A)^I$, where $a \succ_i b$ for all $i \in I$. Take a preference $\succeq_i$ in π. Define a new profile π^* where all voters have the preference $\succeq_i$, so $\pi^* = (\succeq_i, \ldots, \succeq_i)$. By unanimity we have that $a \succ^* b$, where $F(\pi^*) = \succeq^*$. All the preference relationships between a and b are the same in π and π^*, so by IIR $a \succ b$, where $F(\pi) = \succeq$. ■

2-1-1 Arrow's Theorem

Arrows theorem demonstrates that certain social welfare functions, which one might consider "fair and reasonable" cannot exist. The theorem is first introduced by Arrow in [7]. In this case, "fair and reasonable" means that F is unanimous, non-dictatorial and IIR. Before showing Arrow's theorem we will start of with the *pairwise neutrality* lemma.

Lemma 2.4 (pairwise neutrality)**.** *For $|A| \geq 3$, if F is Pareto efficient and IIR, then F is pairwise neutral for strict preferences.*

Proof. Consider profiles $\pi, \pi' \in P(A)^I$ and alternatives $a, b, c, d \in A$, such that all preference relations of (a, b) and (c, d) are strict and

$$a \succ_i b \iff c \succ'_i d \tag{2-6}$$

for all $i \in I$. We are going the show that $a \succ b \iff c \succ' d$, where $F(\pi) = \succeq$ and $F(\pi') = \succeq'$. The proof covers the four possible cases:

1. Alternatives (a, b, c, d) are all different from each other.
2. Either $a = c$ or $b = d$.
3. Either $b = c$ or $a = d$.
4. $b = c$ and $a = d$.

We skip the case $a = c$ and $b = d$, because that is the IIR condition.

Case 1: (a, b, c, d) **are all different.** Define new preference profiles $\pi^\bullet, \pi^\star \in P(A)^I$, such that for all $i \in I$:

$$\begin{aligned}
&a \succ^\bullet_i b \overset{\text{def}}{\iff} a \succ_i b \overset{\text{hyp}}{\iff} c \succ'_i d \overset{\text{def}}{\iff} c \succ^\bullet_i d, \\
&a \succ^\star_i b \overset{\text{def}}{\iff} a \succ_i b \overset{\text{hyp}}{\iff} c \succ'_i d \overset{\text{def}}{\iff} c \succ^\star_i d, \\
&c \succ^\bullet_i a, \quad b \succ^\bullet_i d, \\
&c \prec^\star_i a, \quad b \prec^\star_i d.
\end{aligned}$$

Preferences between other alternatives are arbitrary. The only difference between $\pi^\bullet$ and $\pi^\star$ are in the preference relations of (a,c) and (b,d). Since $a \succ_i b \iff c \succ'_i d$, the contradictions

$$\begin{gathered} c \succ^\bullet_i a \succ^\bullet_i b \succ^\bullet_i d \succ^\bullet_i c, \\ c \prec^\star_i a \prec^\star_i b \prec^\star_i d \prec^\star_i c \end{gathered}$$

do not occur.

By IIR, the social choices are:

$$a \succ b \iff a \succ^\bullet b \iff a \succ^\star b \tag{2-7}$$
$$c \succ' d \iff c \succ^\bullet d \iff c \succ^\star d. \tag{2-8}$$

By Pareto efficiency of F, we have that:

$$c \succ^\bullet a, \quad b \succ^\bullet d, \tag{2-9}$$
$$c \prec^\star a, \quad b \prec^\star d. \tag{2-10}$$

If we assume $a \succ b$, then by (2-7) and (2-9), we have:

$$c \succ^\bullet a \succ^\bullet b \succ^\bullet d,$$

and by transitivity $c \succ^\bullet d$. With formula (2-8) we conclude that $a \succ b \Rightarrow c \succ' d$. Conversely, if we assume $a \prec b$, then with similar arguments, we can find $a \prec b \Rightarrow c \prec' d$.

Case 2: either $a = c$ or $b = d$. We either have $a \succ_i b \iff a \succ'_i d$ or $a \succ_i b \iff c \succ'_i b$. The proof is almost the same as the previous one. Without loss of generality, assume that $a = c$ and $b \neq d$. Define preference profiles $\pi^\bullet, \pi^\star \in P(A)^I$, such that for all $i \in I$,

$$\begin{aligned} & a \succ^\bullet_i b \overset{\text{def}}{\iff} a \succ_i b \overset{\text{hyp}}{\iff} c \succ'_i d \overset{\text{def}}{\iff} c \succ^\bullet_i d, \\ & a \succ^\star_i b \overset{\text{def}}{\iff} a \succ_i b \overset{\text{hyp}}{\iff} c \succ'_i d \overset{\text{def}}{\iff} c \succ^\star_i d, \\ & b \succ^\bullet_i d, \quad b \prec^\star_i d. \end{aligned}$$

By IIR we have that

$$\begin{gathered} a \succ b \iff a \succ^\bullet b \iff a \succ^\star b \\ c \succ' d \iff c \succ^\bullet d \iff c \succ^\star d. \end{gathered}$$

By Pareto efficiency of F, we have that:

$$b \succ^\bullet d, \quad b \prec^\star d.$$

With similar reasoning as before, we can show that $a \succ b \iff a \succ' d$.

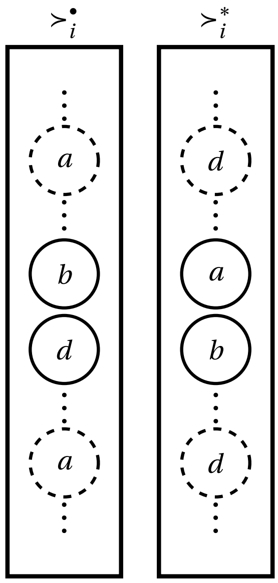

Figure 2-1: The ballot of preferences $\succ_i^\bullet$ and $\succ_i^*$. Higher alternatives are more preferred.

Case 3: Either $b = c$ or $a = d$. We have one of these two situations: $a \succ_i b \iff b \succ_i' d$ or $a \succ_i b \iff c \succ_i' a$. Without loss of generality assume $b = c$ and $a \neq d$.

Define a new profile $\pi^\bullet$, where for all $i \in I$, we have that: $b \succ_i^\bullet !d$ (d just below b), $a \succ_i b \iff a \succ_i^\bullet b$, and all other preferences are arbitrary. See figure 2-1. Note that we either have $a \succ_i^\bullet b \succ_i^\bullet d$ or $b \succ_i^\bullet d \succ_i^\bullet a$, so we have:

$$a \succ_i b \iff a \succ_i^\bullet b \iff a \succ_i^\bullet d. \tag{2-11}$$

The proof of case 2 implies:

$$a \succ b \iff a \succ^\bullet d. \tag{2-12}$$

Next, define another profile π^* where for all $i \in I$, we have that: $a \succ_i^* !b$ (b just below a), $a \succ_i^\bullet d \iff a \succ_i^* d$, and all other preferences are arbitrary. See figure 2-1. With the same reasoning as before, we have that:

$$a \succ_i^\bullet d \iff b \succ_i^* d \tag{2-13}$$

$$a \succ^\bullet d \iff b \succ^* d. \tag{2-14}$$

Formulas (2-11) and (2-13) imply:

$$b \succ_i' d \iff a \succ_i b \iff a \succ_i^\bullet d \iff b \succ_i^* d,$$

so by IIR: $b \succ' d \iff b \succ^* d$. Formulas (2-12) and (2-14) imply:

$$a \succ b \iff b \succ^* d,$$

so we have:

$$a \succ b \iff b \succ' d.$$

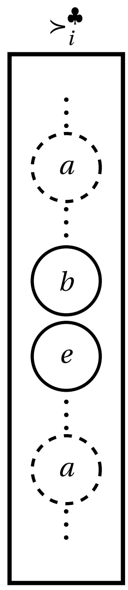

Figure 2-2: The ballot of preference $\succ_i^{\clubsuit}$. Higher alternatives are more preferred. Alternative e is just below b, so we have either $a \succ_i^{\clubsuit} b \succ_i^{\clubsuit} e$ or $b \succ_i^{\clubsuit} e \succ_i^{\clubsuit} a$.

Case 4: $b = c$ and $a = d$. We show that if $a \succ_i b \iff b \succ_i' a$, then $a \succ b \iff b \succ' a$. Take some $e \in A$ unequal to a and b. We define in an iterative fashion three new preference profiles $\pi^{\clubsuit}, \pi^{\diamondsuit}, \pi^{\spadesuit} \in P(A)^I$ the following way:

1. For all $i \in I$, we have that $b \succ_i^{\clubsuit} !e$ (e just below b) and $a \succ_i^{\clubsuit} b \iff a \succ_i b$. The preferences between any other alternatives are arbitrary.

   Looking at Figure 2-2, we can see that if $a \succ_i b$, then $a \succ_i^{\clubsuit} b \succ_i^{\clubsuit} e$ and if $b \succ_i a$, then $b \succ_i^{\clubsuit} e \succ_i^{\clubsuit} a$. So we have for all $i \in I$,

   $$a \succ_i b \iff a \succ_i^{\clubsuit} e, \tag{2-15}$$

   which implies

   $$a \succ b \iff a \succ^{\clubsuit} e. \tag{2-16}$$

2. For each preference $\succ_i^{\diamondsuit}$, we have that $a \succ_i^{\diamondsuit} !b$ (b just below a) and $a \succ_i^{\diamondsuit} e \iff a \succ_i^{\clubsuit} e$. The preferences between any other alternatives are arbitrary.

   Similarly as in step 1, for all $i \in I$, we have that

   $$a \succ_i^{\clubsuit} e \iff b \succ_i^{\diamondsuit} e, \tag{2-17}$$

   which implies

   $$a \succ^{\clubsuit} e \iff b \succ^{\diamondsuit} e. \tag{2-18}$$

3. Each preference $\succ_i^{\spadesuit}$ is defined such that $e \succ_i^{\spadesuit} !a$ (a just below e) and $b \succ_i^{\spadesuit} e \iff b \succ_i^{\diamondsuit} e$. The preferences between any other alternatives is arbitrary.

   Similarly as in step 1, for all $i \in I$, we have that

   $$b \succ_i^{\diamondsuit} e \iff b \succ_i^{\spadesuit} a, \tag{2-19}$$

which implies

$$b \succ^{\diamondsuit} e \iff b \succ^{\spadesuit} a. \tag{2-20}$$

Combining (2-15), (2-17) and (2-19) we get

$$b \succ_i' a \iff a \succ_i b \iff a \succ_i^{\clubsuit} e \iff b \succ_i^{\diamondsuit} e \iff b \succ_i^{\spadesuit} a \tag{2-21}$$

and combining (2-16), (2-18) and (2-20) we get

$$a \succ b \iff a \succ^{\clubsuit} e \iff b \succ^{\diamondsuit} e \iff b \succ^{\spadesuit} a. \tag{2-22}$$

By (2-21) and IIR we have that $b \succ' a \iff b \succ^{\spadesuit} a$ and using (2-22) we conclude

$$b \succ^{\spadesuit} a \iff a \succ b \iff b \succ' a.$$

■

Theorem 2.5 (Arrow's Theorem)**.** *For $|A| \geq 3$, every social welfare function F that satisfies Pareto efficiency and IIR is a dictatorship.*

Before giving the full proof of Arrow's theorem, we give the proof for three candidates two voters, which is easier to understand.

Proof. Three candidates, two voters and strict preferences only. Consider the set of alternatives $A = \{a, b, c\}$ and let the voter set be $I = \{1, 2\}$. Additionally, we assume that all voters only have strict preferences between candidates.

Consider the profile $\pi^A = \{\succ_1^A, \succ_2^A\}$, where both voters have the same preference: $a \succ_i^A b \succ_i^A c$, for $i \in I$. By Pareto efficiency, the social preference is: $a \succ^A b \succ^A c$. See Figure 2-3.

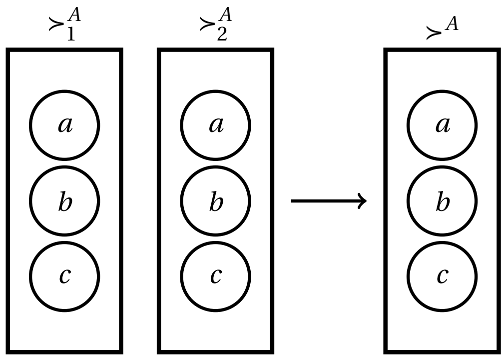

Figure 2-3: Profile π^A

Define a new profile $\pi^B = \{\succ_1^B, \succ_2^B\}$, where $\succ_1^B = \succ_1^A$ and $b \succ_2^B a \succ_2^B c$. Compared to profile π^A, the second voter's preference between a and b is switched. There are two possibilities: $a \succ^B b$ or $b \succ^B a$. With out loss of generality, we assume that $a \succ^B b$. See Figure 2-4. In this case, we show that voter 1 is the dictator.

If it were, $b \succ^B a$, then we can use the begin profile $\pi^{A'}$, where $b \succ_i^{A'} a \succ_i^{A'} c$, for $i \in I$, to get an equivalent situation. In this case, we would proof that voter 2 is the dictator.

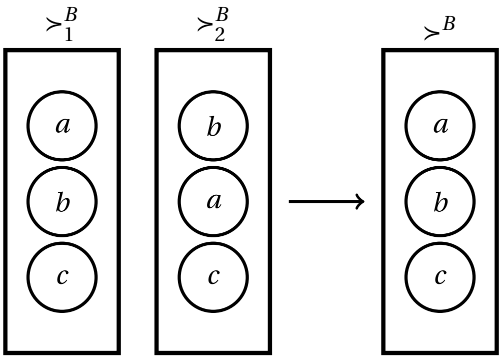

Figure 2-4: Profile π^B

Define the profiles $\pi^A, \pi_A^B, \ldots, \pi^F$, the following way: Voter 1 has the preference $\succ_1^A = \succ_1^B = \ldots = \succ_1^F$, where $\succ_1^A$ is the same as above i.e., $a \succ_1^A b \succ_1^A c$. The preferences of voter 2 are shown in Figure 2-5. Note that these are all possible preferences a voter can have. Voter 1 is the dictator, if his preference is always the same as the social preference i.e., $F(\pi^X) = \succ^X = \succ_1^X$, for all $X \in \{A, B, \ldots, F\}$. As discussed before, this is already true for π^A and π^B.

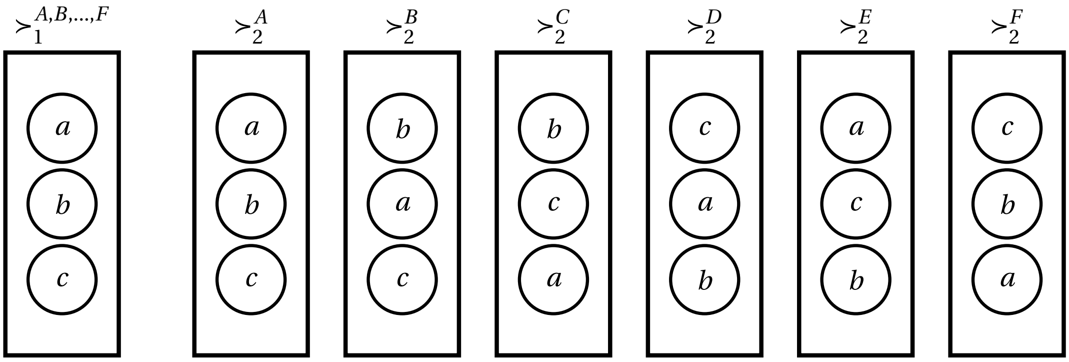

Figure 2-5: Preferences for voter 1 and all possible preferences for voter 2.

Let us now look at π^C, and compare it to π^B. For all $i \in I$, we have that, $a \succ_i^B b \iff a \succ_i^C b$, so IIR implies $a \succ^B b \iff a \succ^C b$, and thus $a \succ^C b$. By Pareto efficiency, we have that $b \succ^C c$. Transitivity implies $a \succ^C b \succ^C c$.

Now, consider π^D, and compare it to π^C. Notice that:

$$a \succ_1^C b \text{ and } b \succ_1^D c$$
$$b \succ_2^C a \text{ and } c \succ_2^D b,$$

so this means $a \succ_i^C b \iff b \succ_i^D c$. The pairwise neutrality Lemma 2.4 implies $a \succ^C b \iff b \succ^D c$, and thus $b \succ^D c$. We have $a \succ^D b$ by Pareto efficiency, so transitivity implies $a \succ^D b \succ^D c$.

Similarly, by using IIR, Pareto and transitivity, we can show for π^E and π^F, that:

$$a \succ^E b \succ^E c$$
$$a \succ^F b \succ^F c.$$

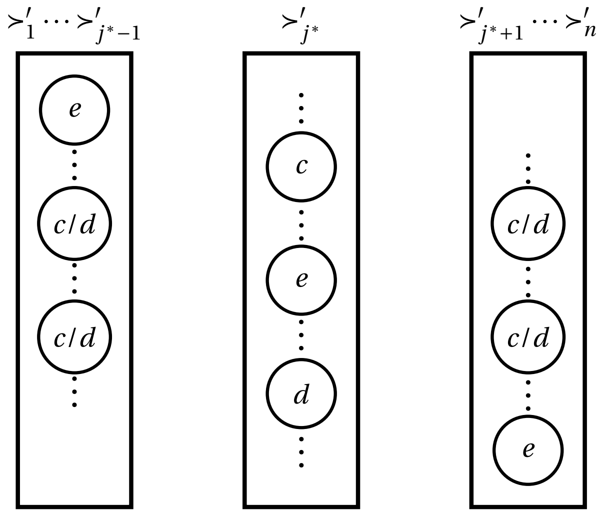

Figure 2-6: Illustration of the modified profile $\succ'$.

■

Proof. Arrow's Theorem. Take any $a \neq b \in A$. Let $\pi^j \in P(A)^I$ be some preference profile where the first j voters strictly prefers a over b and the rest strictly prefers b over a, so $a \succ_i^j b$ if $i \leq j$ and $b \succ_i^j a$ if $i > j$. Take a sequence of preference profiles $\pi^0, \ldots, \pi^n$, where each profile satisfies the above property. Let $F(\pi^j) = \succeq^j$ be the social choice. By Proposition 2.3, we know that F is Pareto and therefore $a \prec^0 b$ and $a \succ^n b$. This means that there has to be some voter $j^* \in I$, such that $a \preceq^{j^*-1} b$ and $a \succ^{j^*} b$. We will now show that j^* has to be the dictator.

Take any $c \neq d \in A$ and let $\pi \in P(A)^I$ be an arbitrary preference profile. Voter j^* is a dictator if $c \succ_{j^*} d$ implies $c \succ d$, where $F(\pi) = \succ$. Assume that $c \succ_{j^*} d$. Since $|A| \geq 3$, we can take an $e \in A$ unequal to c or d. We change π to π' the following way. For voters $i < j^*$ we set $e = \text{top}_i(A)$, for voters $i > j^*$ we set $e = \text{bot}_i(A)$ and voter $i = j^*$ has the preference $c \succ'_{j^*} e \succ'_{j^*} d$. See Figure 2-6 for an illustration. For all voters, the preferences between c and d remains the same, so by IIR the social preference between c and d does not change. , i.e.

$$c \succeq d \iff c \succeq' d \tag{2-23}$$

$$d \succeq c \iff d \succeq' c \tag{2-24}$$

Comparing π' and π^{j^*-1}, we can see that;

$$e \succ'_i c \text{ and } a \succ_i^{j^*-1} b \quad \text{if } i \leq j^* - 1,$$
$$c \succ'_i e \text{ and } b \succ_i^{j^*-1} a \quad \text{if } i > j^* - 1.$$

We know that $b \succeq^{j^*-1} a$, so by pairwise neutrality we have $c \succeq' e$. Similarly for π' and π^{j^*}, we have

$$e \succ'_i d \text{ and } a \succ_i^{j^*} b \quad \text{if } i \leq j^*,$$
$$d \succ'_i e \text{ and } b \succ_i^{j^*} a \quad \text{if } i > j^*.$$

We know that $a \succ_i^{j^*} b$, so by pairwise neutrality $e \succ' d$. Transitivity implies $c \succ' d$ and by (2-23) and (2-24), we conclude $c \succ d$. ■

2-2 Social Choice and the Gibbard-Satterthwaite Theorem

In this section we will look at social choice functions $f : P(A) \to A$, where the outcome is a single alternative (the winner). This is similar to presidential elections. Arrow's theorem has powerful implications for social choice functions. By the Gibbard-Satterthwaite theorem [8] [9], it turns out that in "fair" elections, it is impossible to prevent strategic voting.

We will prove this for preferences that are weak orders, but first we restrict ourselves to preferences with no ties between distinct alternatives, which we call *strong preferences*. These preferences are total orders on A:

Definition 2.16 (Total Order)**.** We say that $\succeq$ is a total order on A, if for all $a, b, c \in A$ it satisfies:

- Transitivity; if $a \succeq b$ and $b \succeq c$, then $a \succeq_i c$.
- Totality; we always have $a \succeq b$ or $b \succeq a$.
- Anti-symmetry: if $a \succeq b$ and $b \succeq a$, then $a = b$.

The set of all total orders on A is $\hat{P}(A) \subset P(A)$.

Definition 2.17 (Strong preference $\succ$)**.** A strong preference $\succ$ on the set of alternatives A is a total order on A.

We say a social choice function can be *strategically* if there exists a situation where a voter can get a more favourable outcome by lying about his true preference. In that case, a voter approaches the election like a game and votes strategically to get the best possible outcome. The reason why we vote (on subjective choices) is to take a decision that is the most acceptable for everyone. But if not everyone is signalling their true believes, then this undermines the intention of the election.

Definition 2.18 (manipulability)**.** A social choice function f can be *(strategically) manipulated* by voter i, if there exist for some profile $(\succ_1, \ldots, \succ, \ldots, \succ_n) \in \hat{P}(A)^I$ and some preference $\succ'_i \in \hat{P}(A)$, such that $b \succ_i a$ and

$$f(\succ_1, \ldots, \succ_i, \ldots, \succ_n) = a$$
$$f(\succ_1, \ldots, \succ'_i, \ldots, \succ_n) = b.$$

Looking at the above definition, if $\succ_i$ is i's true preference, then he can get a better outcome by misrepresenting his vote with $\succ'_i$.

An equivalent and alternative definition of non-manipulability is *monotonicity*. Monotonicity of f means that if a voter changes his preference and the social choice changes from a to b, then it has to be that the voter changed his preference from preferring a over b to b over a.

Definition 2.19 (Monotonicity)**.** A social choice function f is *monotone* if

$$f(\succ_1, \ldots, \succ_i, \ldots, \succ_n) = a \neq b = f(\succ_1, \ldots, \succ'_i, \ldots, \succ_n)$$

implies that $a \succ_i b$ and $a \prec'_i b$.

Proposition 2.6. *A social choice function f is non-manipulable if and only if it is monotone.*

Proof. Assume f is manipulable. Then there is a profile $(\succ_1, \dots, \succ_i, \dots, \succ_n)$ and $\succ'_i$, such that $b \succ_i a$ and

$$f(\succ_1, \dots, \succ_i, \dots, \succ_n) = a \neq b = f(\succ_1, \dots, \succ'_i, \dots, \succ_n),$$

so f is not monotone.

Conversely, assume that f is not monotone. Then there is a profile $(\succ_1, \dots, \succ_i, \dots, \succ_n)$ and preference $\succ'_i$, such that

$$f(\succ_1, \dots, \succ_i, \dots, \succ_n) = a \neq b = f(\succ_1, \dots, \succ'_i, \dots, \succ_n),$$

and $b \succ_i a$ or there is a profile $(\succ_1, \dots, \succ'_i, \dots, \succ_n)$ and $\succ_i$, such that

$$f(\succ_1, \dots, \succ_i, \dots, \succ_n) = a \neq b = f(\succ_1, \dots, \succ'_i, \dots, \succ_n),$$

and $a \succ'_i b$. So f is not incentive compatible.

■

Besides being non-manipulable, we als want elections to be "fair". We say that a fair election should satisfy the following two properties: Each candidate should come up as a winner in some elections and no voter can be a dictator. This made precise by the following two definitions:

Definition 2.20 (Non-imposed)**.** The social choice function f is a *non-imposed*, if for all $a \in A$, there exist a profile $\pi \in \hat{P}(A)^I$, such that $f(\pi) = a$.

Definition 2.21 (Dictatorship)**.** The social choice function f is a *dictatorship* if there exist a voter i (*the dictator*), such that for any $\pi \in \hat{P}(A)^I$, if $a \in \text{top}_i^\pi(A)$, then $f(\pi) = a$.

Note that for strong preferences, $\text{top}_i^\pi(A)$ is a singleton.

2-2-1 Gibbard-Satterthwaite theorem for strong preferences.

The Gibbard-Satterthwaite states that it is impossible to find an social choice function f that is non-manipulable, non-dictatorial and non-imposed, when the number of alternatives is three or more.

The idea behind the proof is as follows; We assume that f is incentive compatible, non-dictatorial and non-imposed. We define the extended social welfare function F_e of f and prove that F_e is a social welfare function, that contradicts Arrow's theorem.

Definition 2.22 ($\succ^S$)**.** Take some preference $\succ \in \hat{P}(A)$ and some subset $S \subset A$. We define the preference $\succ^S \in \hat{P}(A)$ by moving all elements in S to the top and preserving the internal ordering between elements in S and between elements in $A \setminus S$. More precisely, if $a \in S$ and $b \in A \setminus S$, then $a \succ^S b$ and if $a, b \in S$ or $a, b \in A \setminus S$, then $a \succ^S b \iff a \succ b$. An illustrative example is given in Figure 2-7.

Proposition 2.7. *Take any* $\succ_1, \dots, \succ_n \in \hat{P}(A)$ *and any* $S \subset A$. *If* f *is monotone and non-imposed, then*

$$f(\succ_1, \dots, \succ_n) \in S.$$

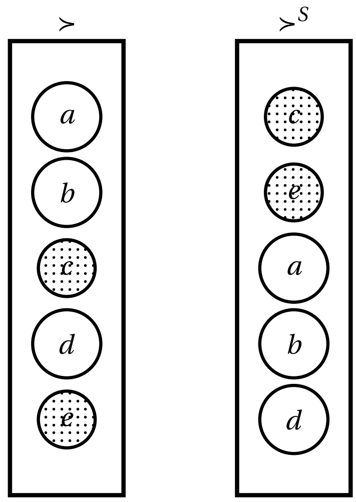

Figure 2-7: An example of $\succ^S$, where $S = \{c, e\}$

.

Proof. Take some $s \in S$. Since f is onto there is some $\succ'_1, \ldots, \succ'_n \in \hat{P}(A)$ such that,

$$f\left(\succ'_1, \ldots, \succ'_n\right) = s.$$

Next, sequentially change $\succ'_1, \ldots, \succ'_n$ to $\succ^S_1, \ldots, \succ^S_n$ element by element. At no point i will the social choice become some $b \in A \setminus S$, i.e.

$$\begin{aligned} f\left(\succ^S_1, \ldots, \succ^S_{i-1}, \succ'_i, \ldots, \succ_n\right) &= a \in S \\ f\left(\succ^S_1, \ldots, \succ^S_i, \succ'_{i+1}, \ldots, \succ_n\right) &= b \in A \setminus S. \end{aligned}$$

If it did, then by monotonicity this would imply that $b \succ^S_i a$, but by definition we have that $a \succ^S_i b$ for all i. ■

Definition 2.23 (F_e)**.** We define the extended social welfare function F_e of f. Let $F_e(\succ_1, \ldots, \succ_n) = \succ$. For $a, b \in A$, we have that $a \succ b$ if and only if $f\left(\succ^{\{a,b\}}_1, \ldots, \succ^{\{a,b\}}_n\right) = a$.

Lemma 2.8. *For $|A| \geq 3$, if f is monotone and non-imposed, then F_e is a social welfare function. Additionally, for $\pi \in \hat{P}(A)^I$, we have that $F_e(\pi)$ is a strong preference.*

Proof. We show that $F_e(\succ_1, \ldots, \succ_n) = \succ$ is a total order i.e. antisymmetric, transitive and total.

- Antisymmetry. Assume $a \succ b$ and $b \succ a$. We have that

$$a \succ b \iff f\left(\succ^{\{a,b\}}_1, \ldots, \succ^{\{a,b\}}_n\right) = a \tag{2-25}$$

$$b \succ a \iff f\left(\succ^{\{a,b\}}_1, \ldots, \succ^{\{a,b\}}_n\right) = b \tag{2-26}$$

  and therefore $a = b$, since the arguments of f in (2-25) and (2-26) are the same.

- Totality. By proposition 2.7, we know that $f\left(\succ^{\{a,b\}}_1, \ldots, \succ^{\{a,b\}}_n\right) \in \{a, b\}$ for any $a, b \in A$. This means that for any a, b it is well defined whether it is either $a \succ b$ or $b \succ a$.

- Transitivity. Assume f is not transitive, then for some distinct $a,b,c \in A$, we would have $a \succ b$, $b \succ c$ and $a \not\succ c$. By totality, $a \not\succ c$ implies $c \succ a$. Without loss of generality, assume that
$$f\left(\succ_1^{\{a,b,c\}},\dots,\succ_n^{\{a,b,c\}}\right) = a.$$
We sequentially change $\succ_i^{\{a,b,c\}}$ to $\succ_i^{\{a,c\}}$. By monotonicity we have that
$$f\left(\succ_1^{\{a,c\}},\dots,\succ_n^{\{a,b\}}\right) = a,$$
which implies $a \succ c$.

■

Theorem 2.9 (Gibbard-Satterthwaite for strong preferences)**.** *For* $|A| \geq 3$*, the social choice function* f *is incentive compatible (monotone) and non-imposed if and only if* f *is a dictatorship.*

Proof. Assume f is a dictatorship. If we have
$$f(\succ_1,\dots,\succ_i,\dots,\succ_n) = a \neq a' = f(\succ_1,\dots,\succ_i',\dots,\succ_n),$$
then i is the dictator and a is the top choice in $\succ_i$ and a' is the top choice in $\succ_i'$, so $a \succ_i a'$ and $a' \succ_i' a$ and thus f is monotone.

We prove the converse by showing that if f is incentive compatible, non-imposed and not a dictatorship, then the extended social welfare function F_e will be unanimous, IIR and not a dictatorship, which contradicts Arrow's theorem.

- Pareto; If all voters i have the preference $a \succ_i b$, then $\succ_i^{\{a,b\}} = \left(\succ_i^{\{a,b\}}\right)^{\{a\}}$, so by Proposition 2.7;
$$f\left(\succ_1^{\{a,b\}},\dots,\succ_n^{\{a,b\}}\right) = a$$
and therefore $a \succ b$, where $\succ = F_\mathrm{e}(\succ_1,\dots,\succ_n)$.
- Independent of irrelevant relationships; Let $\succ_1,\dots,\succ_n \in P(A)$ and $\succ_1',\dots,\succ_n' \in P(A)$ be two preference profiles such that $a \succ_i b \iff a \succ_i' b$. By Property 2.7 we know that
$$f\left(\succ_1^{\{a,b\}},\dots,\succ_n^{\{a,b\}}\right) \in \{a,b\} \ni f\left(\succ_1'^{\{a,b\}},\dots,\succ_n'^{\{a,b\}}\right).$$
Now, sequentially change $\succ_1,\dots,\succ_n$ to $\succ_1',\dots,\succ_n'$. Since f is monotone, and $a \succ_i b \iff a \succ_i'$ b, we have that $f\left(\succ_1'^{\{a,b\}},\dots,\succ_j'^{\{a,b\}},\succ_{j+1}^{\{a,b\}},\dots,\succ_n^{\{a,b\}}\right)$ remains constant for all $j \in I$ and thus
$$f\left(\succ_1^{\{a,b\}},\dots,\succ_n^{\{a,b\}}\right) = f\left(\succ_1'^{\{a,b\}},\dots,\succ_n'^{\{a,b\}}\right).$$
- Non-dictatorial; Select an arbitrary voter $i \in I$. Since f is not a dictatorship, there exists an $a \in A$ and $\boldsymbol{\succ} \in P(A)^I$, such that $a = \mathrm{top}_i(A)$ and $f(\boldsymbol{\succ}) = b \neq a$. Now, sequentially change $\succ_1,\dots,\succ_n$ to $\succ_1^{\{a,b\}},\dots,\succ_n^{\{a,b\}}$. Since f is monotone,and $a \succ_i b \iff a \succ_i^{\{a,b\}} b$, we have that $f\left(\succ_1^{\{a,b\}},\dots,\succ_j^{\{a,b\}},\succ_{j+1},\dots,\succ_n\right)$ remains constant for all $j \in I$, so $f\left(\succ_1^{\{a,b\}},\dots,\succ_n^{\{a,b\}}\right) = b$. This implies $b \succ a$, with $F_\mathrm{e}(\succ_1,\dots,\succ_n) = \succ$ and therefore i cannot by a dictator in F_e. Since i was arbitrary, F_e is not a dictatorship.

■

2-2-2 Gibbard-Satterthwaite for weak preferences

With the Gibbard-Satterthwaite theorem for strong preferences, we can proof the extension for weak preferences. For weak preferences, the definitions for manipulability, monotonicity, dictatorships and non-imposed are practically the same, but with $P(A)$ instead of $\hat{P}(A)$:

Definition 2.24 (manipulability)**.** A social choice function f can be *(strategically) manipulated* by voter i, if there exist for some profile $(\succeq_1,\ldots,\succeq_i,\ldots,\succeq_n) \in P(A)^I$ and some preference $\succeq_i' \in P(A)$, such that $b \succ_i a$ and

$$f(\succeq_1,\ldots,\succeq_i,\ldots,\succeq_n) = a$$
$$f(\succeq_1,\ldots,\succeq_i',\ldots,\succeq_n) = b.$$

Definition 2.25 (Monotonicity)**.** A social choice function f is *monotone* if

$$f(\succeq_1,\ldots,\succeq_i,\ldots,\succeq_n) = a \neq b = f(\succeq_1,\ldots,\succeq_i',\ldots,\succeq_n)$$

implies that $a \succeq_i b$ and $a \preceq_i' b$.

Proposition 2.10. *A social choice function f is non-manipulable if and only if it is monotone.*

Proof. Assume f is manipulable. Then there is a profile $(\succ_1,\ldots,\succ_i,\ldots,\succ_n)$ and $\succ_i'$, such that $b \succ_i a$ and

$$f(\succ_1,\ldots,\succ_i,\ldots,\succ_n) = a \neq b = f(\succ_1,\ldots,\succ_i',\ldots,\succ_n),$$

so f is not monotone.

Inversely, assume that f is not monotone. Then there is a profile $(\succeq_1,\ldots,\succeq_i,\ldots,\succeq_n)$ and preference $\succeq_i'$, such that

$$f(\succeq_1,\ldots,\succeq_i,\ldots,\succeq_n) = a \neq b = f(\succeq_1,\ldots,\succeq_i',\ldots,\succeq_n),$$

and $b \succ_i a$ or there is a profile $(\succeq_1,\ldots,\succeq_i',\ldots,\succeq_n)$ and $\succeq_i$, such that

$$f(\succeq_1,\ldots,\succeq_i,\ldots,\succeq_n) = a \neq b = f(\succeq_1,\ldots,\succeq_i',\ldots,\succeq_n),$$

and $a \succ_i' b$. So f is not incentive compatible. ∎

Definition 2.26 (Non-imposed)**.** The social choice function f is a *non-imposed*, if for all $a \in A$, there exist a profile $\pi \in P(A)^I$, such that $f(\pi) = a$.

Definition 2.27 (Dictatorship)**.** The social choice function f is a *dictatorship* if there exist a voter i (*the weak dictator*), such that for any $\pi \in P(A)^I$, if $a \in \text{top}_i^\pi(A)$, then $f(\pi) = a$.

Theorem 2.11 (Gibbard-Satterthwaite for weak preference profiles)**.** *For $|A| \geq 3$, if the social choice function f is non-manipulable (monotone) and non-imposed, then f is a weak dictatorship.*

Proof. Let $\hat{f}$ be f restricted to $\hat{P}(A)^I$, i.e. $\hat{f} : \hat{P}(A)^I \to A$ and for all $\pi \in \hat{P}(A)^I$, we have $f(\pi) = \hat{f}(\pi)$. First, we will show using Theorem 2.9 that $\hat{f}$ is a dictatorship, with some dictator j. We then show that j is the dictator of f.

It is clear that if f is non-manipulable, then $\hat{f}$ is also non-manipulable. We show that $\hat{f}$ is non-imposed, by showing that for all $i \in I$ and some $\hat{\pi} \in \hat{P}(A)^I$, if $a \in \text{top}_i^{\hat{\pi}}$, then $\hat{f}(\hat{\pi}) = a$. Suppose not, then there exist some $\hat{\pi} \in \hat{P}(A)^I$, where x is the top preference of all voters and $\hat{f}(\hat{\pi}) = f(\hat{\pi}) \neq x$. Since f is non-imposed, there exists a profile π, such that $f(\pi) = x$. Now we change the preferences in π to the ones in $\hat{\pi}$ one by one, i.e.

$$\begin{aligned}
\pi_0 &= \pi = (\succeq_1, \dots, \succeq_n) \\
\pi_1 &= \left(\hat{\succ}_1, \succeq_2, \dots, \succeq_n\right) \\
&\vdots \\
\pi_n &= \hat{\pi} = \left(\hat{\succ}_1, \dots, \hat{\succ}_n\right).
\end{aligned}$$

There has to exists a smallest $k \in (1, \dots, n)$, where $f(\pi_{k-1}) = x \neq y = f(\pi_k)$. But this would contradict monotonicity of f, because $x \hat{\succ}_k y$. Now, theorem 2.9 implies that $\hat{f}$ is a dictatorship, with some dictator j.

In the second part, we prove that j is also the dictator of f. Suppose j is not the dictator, then there exists a profile $\pi \in P(A)^I$, such that $f(\pi) = x$ and $x \notin \text{top}_j^{\pi}(A)$. One by one, make x the single top preference for every voter except j, i.e.

$$\begin{aligned}
\pi_1 &= \left(\succeq_1^{\{x\}}, \succeq_2, \dots, \succeq_j, \dots, \succeq_n\right) \\
\pi_2 &= \left(\succeq_1^{\{x\}}, \succeq_2^{\{x\}}, \dots, \succeq_j, \dots, \succeq_n\right) \\
&\vdots \\
\pi_{n-1} = \pi^x &= \left(\succeq_1^{\{x\}}, \dots, \succeq_{j-1}^{\{x\}}, \succeq_j, \succeq_{j+1}^{\{x\}}, \dots, \succeq_n^{\{x\}}\right).
\end{aligned}$$

By monotonicity, x remains the winner at every consecutive π_k, where $k = 1, \dots, n-1$.

Similarly in the next step, let $\hat{\succ}$ be the strict preference that results from arbitrary breaking all ties in $\succeq$. One by one, we break all ties in π^x for all voters except j, i.e.

$$\begin{aligned}
\pi_1^x &= \left(\hat{\succ}_1^{\{x\}}, \dots, \succeq_{j-1}^{\{x\}}, \succeq_j, \succeq_{j+1}^{\{x\}}, \dots, \succeq_n^{\{x\}}\right) \\
&\vdots \\
\pi_{n-1}^x &= \left(\hat{\succ}_1^{\{x\}}, \dots, \hat{\succ}_{j-1}^{\{x\}}, \succeq_j, \hat{\succ}_{j+1}^{\{x\}}, \dots, \hat{\succ}_n^{\{x\}}\right).
\end{aligned}$$

By monotonicity, x remains the winner for every consecutive π_k^x, where $k = 1, \dots, n-1$. If we now break all ties in $\succeq_j$, then by Theorem 2.9, the winner will become some $y \in \text{top}_j^{\pi}$, but this contradicts monotonicity, because if the winner becomes y instead of x, then we should have $x \succeq_j y$ and $x \hat{\prec}_j y$.

∎

Chapter 3

Two Juror Verdicts

The model in this chapter is based on the model in [1], which uses a three juror casting vote scheme. Before searching for general results for three jurors, we look at a simpler two juror advisor-decider model.

3-1 The Advisor-Decider Mechanism

Let the state of nature be either A or B. The state is unknown and we want to find out which is most likely to be true. In an advisor-decider decision process, there are two jurors where one is the *advisor* and the other is the *decider*. The adviser passes a vote A or B to the decider. The decider, with the knowledge of the advisor's choice and his own insights, gives his vote. The decider's vote is the final choice of this whole decision process.

An example of this, is a football game where the assistant referee flags if a player is off-side. Combined with his own observation of the game, the main referee makes the final call for the refereeing team.

A juror cannot observe the state of nature directly, but only a realisation of a random signal S_i. The random signal S_i takes values in a continues interval $\mathscr{S} \subseteq \mathbb{R}$, and it represents the noisy observations of the jurors. The signal S_i has different probability density functions, depending on the true state of nature:

$$\begin{aligned} f_i(s_i), &\quad \text{if nature is } A \\ g_i(s_i), &\quad \text{if nature is } B. \end{aligned}$$

The corresponding cumulative distributions are denoted as $F_i(s_i)$ and $G_i(s_i)$ respectively. We restrict ourselves to density functions that satisfy:

Definition 3.1.

1. $f_i(s_i)$ is continuous in $\mathscr{S}$.

2. $f_i(s_i) = g_i(-s_i)$.

3. $g_i(s_i)/f_i(s_i)$ is non-increasing in s_i.

Throughout the chapter, we will often make use of the following fact:

$$F_i(s_i) = 1 - G_i(-s_i). \tag{3-1}$$

We can show it by noting that property 2 implies: $F_i(s_i) = C - G_i(-s_i)$. Taking the limit of s_i to infinity, implies $C = 1$.

The jurors make a decision in what is called a *Bayesian decision process.* There exist extensive literature on such processes, see [10], but we will deal with our example only. In addition to the observation s_i, each jurors holds prior believes. Juror i's prior belief that the probability nature is A, is π_0. We assume that the jurors are unbiased, so without any information A and B are equally probable; The default prior probability that nature is A is $\pi_0 = 1/2$.

After i receives the observation s_i, then his posterior believe that nature is A is:

$$P(A|s_i) = \frac{\pi_i f_i(s_i)}{\pi_i f_i(s_i) + (1-\pi_i)g_i(s_i)}.$$

The juror will then:

- Vote A, if $P(A|s_i) > 1/2$.
- Vote B, if $P(A|s_i) < 1/2$.
- Make a random choice between A and B with equal probability, if $P(A|s_i) = 1/2$.

We can see that $P(A|s_i) = \frac{\pi_i}{\pi_i+(1-\pi_i)g_i/f_i}$ is non-decreasing in s_i, because g_i/f_i is non-increasing in s_i. The higher s_i is, the stronger he believes that nature is A. We do not have to worry about the case when $f_i(s_i) = 0$, because the jurors will never receive such signals.

Jurors can have different probability density functions. We say that a juror i has a higher *ability* then j, if it is more likely for i to receive a higher signal, when nature is A, i.e. $1 - F_i(s) \geq 1 - F_j(s)$. By symmetry, it is more likely for a higher ability juror to receive a lower signal, when nature is B. The probability densities for each juror is public knowledge.

Assume two jurors i and j, where i is the advisor and j is the decider. Since the advisor has no other information, besides a realisation s_i of his random signal, his prior is $\pi_i = \pi_0$. With his signal s_i and prior p_i, the decider makes a choice between A and B.

After the advisor's vote, the decider is informed on which state i voted on, but not the advisor's signal value s_i. After the advisor voted some $X \in \{A, B\}$, the decider j will update his prior:

$$\pi_{j,X} = P(A|i \text{ voted } X) \tag{3-2}$$

With the updated prior and a realisation s_j, the decider makes a final decision between A and B.

The probability that the decider votes A, while the true state is A, is:

$$\begin{aligned} q_A =& P(\mathrm{A} \cap j \text{ votes } A) \\ =& P(A)P(j \text{ votes } A|A) \\ =& P(A)\left[P(i \text{ votes } A \cap j \text{ votes } A|A) + P(i \text{ votes } B \cap j \text{ votes } A|A)\right] \\ =& P(A)\left[P(i \text{ votes } A)P(j \text{ votes } A|A, i \text{ votes } A) + P(i \text{ votes } B)P(j \text{ votes } A|A, i \text{ votes } B)\right]. \end{aligned}$$

Similarly, The probability that the decider votes B, while nature is B, is q_B. The *Quality* is Q, and it is the probability a correct answer is given:

$$Q = q_A + q_B. \tag{3-3}$$

3-2 Linear Density Functions

We start by analysing the problem with linear density functions, which are the simplest kind of density functions. Alpern and Chen use the same kind of functions in [1]. Consider a juror i, who has an ability $a_i \in [0,1]$. In this case, a higher ability juror has a higher a_i. The juror's signal is a random variable S_i, that takes values in the signal space $\mathscr{S} = [-1,1]$. The linear density functions are:

$$\begin{aligned} f_i(s) &= \frac{1 + a_i s}{2}, \quad \text{if nature is A} \\ g_i(s) &= \frac{1 - a_i s}{2}, \quad \text{if nature is B} \end{aligned}$$

The corresponding cumulative distribution functions are then:

$$\begin{aligned} F_i(s) &= \frac{(s+1)(a_i s - a_i + 2)}{4}, \quad \text{if nature is A} \\ G_i(s) &= \frac{(s+1)(a_i - a_i s + 2)}{4}, \quad \text{if nature is B.} \end{aligned}$$

Examples of linear density functions are plotted in Figure 3-1.

Let there be two jurors i and j, where i has the role of advisor and j is the decider. The abilities i and j, are denoted as a and b respectively. The advisor has a prior $\pi_i = \pi_0 = 1/2$. If i receives a realisation s_i of his random signal, then he believes the probability that nature is A is:

$$\begin{aligned} P(A|s_i) =& \frac{\pi_0 f_i(s_i)}{\pi_0 f_i(s_i) + (1-\pi_0) g_i(s_i)} \\ =& \frac{f_i(s_i)}{f_i(s_i) + g_i(s_i)} \end{aligned}$$

Note that if $a = 0$, then $f_i(s_i) = g_i(s_i) = 1/2$ and $P(A|s_i) = 1/2$. In this case the advisor makes a random choice between A and B, with equal probability. The lowest ability juror is no better than a coin flip. For jurors with an ability $a > 0$, we have that $P(A|s_i)$ is strictly increasing in s_i. That is because $f_i(s_i)$ is strictly increasing and $g_i(s_i)$ is strictly decreasing, which implies that $g_i(s_i)/f_i(s_i)$ is strictly decreasing. This implies that there is some threshold t_i, where $P(A|t_i) = 1/2$ and $P(A|t_i) > 1/2$, for $s_i > t_i$. For the advisor, this is when $f_i(t_i) = g_i(t_i)$, and that is only when $t_i = 0$.

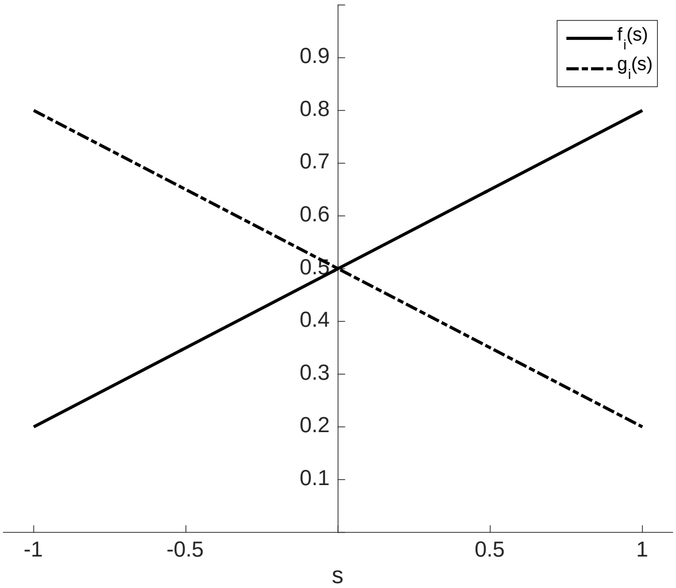

Figure 3-1: Example of linear distribution functions with $a_i = 0.6$. When nature is A the distribution is $f_i(s)$, and when nature is B the distribution is $g_i(s)$.

Assume that nature is A. The probability that $s_i > t_i = 0$ and that the juror will give the correct verdict is the area under the curve $f_i(s)$, from $s = t_i = 0$ to $s = 1$. The size of the area increases as a is higher. A juror with the maximum ability of $a = 1$ gives the correct with probability $3/4$.

When the advisor votes A, then the decider's prior is updated:

$$\begin{aligned}\pi_{j,A}(a) =& P(A|s_i > 0) = \frac{\pi_0 (1 - F_i(0))}{\pi_0 (1 - F_i(0)) + (1 - \pi_0)(1 - G_i(0))} \\ =& 1 - F_i(0) \\ =& \frac{2+a}{4}.\end{aligned} \tag{3-4}$$

In case $a = 0$, the prior is $\pi_{j,A}(0) = 1/2$, so having an advisor with the weakest ability, is like having no advisor at all. Similarly, we find that updated prior is $\pi_{j,B}(a) = 1 - \pi_{j,A}(a) = \frac{2-a}{4}$, if the advisor voted B.

The decider, with prior π_j and signal s_j, believes the probability that nature is A is:

$$P\left(A|s_j\right) = \frac{\pi_{j,A} f_j(s_j)}{\pi_{j,A} f_i(s_j) + (1 - \pi_{j,A}) g_j(s_j)}.$$

When the decider has the lowest ability $b = 0$, then:

$$P\left(A|s_j\right) = \begin{cases} \pi_{j,A}(a) \geq \frac{1}{2}, & \text{if } i \text{ voted } A \\ \pi_{j,B}(a), \leq \frac{1}{2} & \text{if } i \text{ voted } B. \end{cases}$$

This means that the decider with the lowest ability, will always follow the advisor's vote.

If j's ability is $b > 0$, then, as before, we can find a threshold t_j, such that $P\left(A|s_j\right) > 1/2$, if $s_j > t_j$. If the advisor voted A, this is:

$$t_{j,A} = \frac{1 - 2\pi_{j,A}}{b} = -\frac{a}{b2}, \tag{3-5}$$

and if he voted B, this is $t_{j,B} = -t_{j,A}$. Note that if $a/b \geq 2$, then $t_j, A \leq -1$ and $t_j, B \geq 1$, so when the advisor's ability is twice as high as the decider's ability, the decider will always follow the advisor. The decider's signal is irrelevant. Take the same jurors, and switch their roles, so the better juror votes last. The threshold is some value close to zero, and the decider will follow the advisor, if the magnitude of his signal is small. In this case, the decider's signal is relevant in the decision process. Intuitively, using all available information should give a better change of getting the correct verdict. Later on, we will see that letting the better juror be the decider, is indeed optimal.

The probability that the correct verdict is given by the decider, is the quality:

$$\begin{aligned}
Q(a,b) =& \pi_0 \left[P(s_i > 0|A) P\left(s_j > t_{j,A}|A\right) + P(s_i < 0|A) P\left(s_j > t_{j,B}|A\right) \right] \\
&+ (1-\pi_0) \left[P(s_i < 0|B) P\left(s_j < t_{j,B}|B\right) + P(s_i > 0|B) P\left(s_j < t_{j,A}|B\right) \right] \\
=& \pi_0 \left[(1 - F_i(0)) \left(1 - F_j\left(t_{j,A}\right)\right) + F_i(0) \left(1 - F_j\left(t_{j,B}\right)\right) \right] \\
&+ (1-\pi_0) \left[G_i(0) G_j\left(t_{j,B}\right) + (1 - G_i(0)) G_j\left(t_{j,A}\right) \right] \\
=& \left[1 - F_i(0)\right] \left(1 - F_j\left(t_{j,A}\right)\right) + F_i(0) \left(1 - F_j\left(t_{j,B}\right)\right)
\end{aligned}$$

$$Q(a,b) = \begin{cases} \frac{2+a}{4} & \text{if } \frac{a}{b} \geq 2 \text{ or } b = 0, \\ \frac{a^2+4b^2+8b}{16b} & \text{if } 0 \leq \frac{a}{b} < 2. \end{cases} \tag{3-6}$$

For ability levels a and b, we want to find out which juror should be the decider, if we want to maximise the quality $Q(a,b)$. Let the difference in quality when switching roles between advisor and decider be

$$\Delta(a,b) = Q(a,b) - Q(b,a). \tag{3-7}$$

We can prove the following proposition:

Proposition 3.1. *Let there be two jurors i and j, with ability levels a and b respectively. If $a \leq b$, then $\Delta_{ij} = Q_{ij} - Q_{ji} \geq 0$. To maximise the quality, the worst juror should be the advisor and the best juror should be the decider.*

Proof. We can assume that $b \neq 0$, because if $b = 0$, then also $a = 0$, so the role assignment does not matter.

We will consider two cases for the rest of the proof. In the first case we considers: $b/a \geq 2$ or $a = 0$. We then have:

$$\begin{aligned}
\Delta_{ij} =& \frac{a^2+4b^2+8b}{16b} - \frac{2+b}{4} \\
=& \frac{a^2}{16b} \\
\geq& 0.
\end{aligned}$$

The last case is when:

$$1 \leq b/a < 2. \tag{3-8}$$

The quality difference is:

$$\begin{aligned}
\Delta_{ij} =& \frac{a^2+4b^2+8b}{16b} - \frac{b^2+4a^2+8a}{16a} \\
=& \frac{a^3 - 4a^2b + 4ab^2 - b^3}{16ab}
\end{aligned} \tag{3-9}$$

If we rewrite $b = a+\epsilon$, then under (3-8), ϵ has to satisfy $0 \le \epsilon < a$. We then find:

$$a^3 - 4a^2b + 4ab^2 - b^3 = \epsilon\left(a^2 - \epsilon^2 + a\epsilon\right) \ge 0,$$

since $\epsilon \le a$. ■

There are limitations with the linear distribution in this section. One limitation is that the two best possible jurors only have 13/16 change of getting the correct verdict. In the next section we are going to look at a different distribution, where it is possible that the probability of getting the correct verdict goes to 1.

3-3 β-Density Functions

With the β-distribution, we use the signal space $\mathscr{S} = [-1,1]$. A juror i with ability level $a_i \in [1,\infty)$ has the probability density function:

$$\begin{aligned} f_i(s_i) &= \frac{a_i}{2}\left(\frac{1+s_i}{2}\right)^{a_i-1} && \text{if nature is } A \\ g_i(s_i) &= \frac{a_i}{2}\left(\frac{1-s_i}{2}\right)^{a_i-1} && \text{if nature is } B, \end{aligned}$$

and cumulative distribution function

$$\begin{aligned} F_i(s_i) &= \left(\frac{1+s_i}{2}\right)^{a_i} && \text{if nature is } A \\ G_i(s_i) &= 1-\left(\frac{1-s_i}{2}\right)^{a_i} && \text{if nature is } B. \end{aligned}$$

Examples of the beta distribution functions for different ability levels are shown in Figure (3-2). The advantage of using beta density functions over the linear ones, is that we can model very competent jurors. With a prior $\pi_i = 1/2$, the probability that i gives the correct verdict goes to one as the ability level goes to infinity, while in the model with linear densities, the best juror is only correct with probability 3/4.

Similar to the previous section, we have two jurors i and j, where i is the advisor and j is the decider. The ability levels for i and j are a and b respectively. Both jurors receive signal values s_i and s_j. The advisor's prior is $\pi_i = \pi_0$, and the probability he believes that nature is A is:

$$\begin{aligned} P(A|s_i) &= \frac{\pi_0 f_i(s_i)}{\pi_0 f_i(s_i) + (1-\pi_0)g_i(s_i)} \\ &= \frac{(1+s_i)^{a-1}}{(1+s_i)^{a-1} + (1-s_i)^{a-1}} \end{aligned} \tag{3-10}$$

For the lowest ability level, $a = 1$, we have that $P(A|s_i) = 1/2$ and the advisor will randomly select between A and B with equal probability.

If $a > 1$, then $P(A|s_i)$ is strictly increasing. This means that there is threshold $t_i = 0$, where $P(A|t_i = 0) = 1/2$, and the advisor votes A, if $s_i > t_i$, and he votes B, if $s_i < t_i$. Since the distribution is continuous, the probability that $s_i = 0$ is zero.

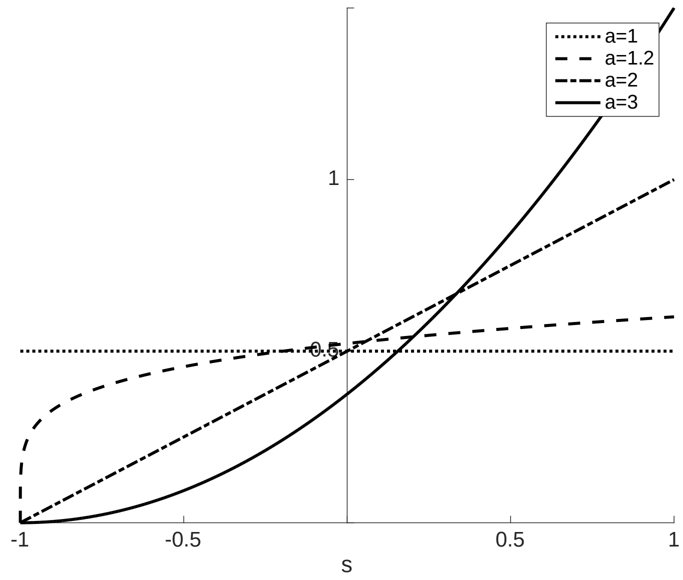

Figure 3-2: Example of the beta distribution functions $f(s)$ with different a.

If the advisor i voted A, then the decider's updated prior is:

$$\begin{aligned}\pi_{j,A} =& P(A|s_i > 0) = \frac{\pi_0\,(1-F_i(0))}{\pi_0\,(1-F_i(0)) + (1-\pi_0)\,(1-G_i(0))} \\ =& 1 - F_i(0) \\ =& 1 - 2^{-a}.\end{aligned} \tag{3-11}$$

When the advisor has the lowest ability level $a = 1$, then the decider's prior is $\pi_{j,A} = 1/2$, which is like having no advisor at all. As the advisor's ability level gets higher the prior increases, and $\pi_{j,A} \to 1$ as $a \to \infty$. Similarly, if the advisor voted B then the decider's prior is $\pi_{j,B} = 2^{-a}$.

For some prior π_j, the decider believes the probability that nature is A is:

$$P\left(A|s_j\right) = \frac{\pi_j (1+s_j)^{b-1}}{\pi_j (1+s_j)^{b-1} + (1-\pi_j)(1-s_j)^{b-1}}. \tag{3-12}$$

For $b > 1$, Formula (3-12) is strictly increasing, so we can find a threshold, which is a signal when $P\left(A|s_j\right) = 1/2$. If the advisor voted A, then the threshold is:

$$\begin{aligned}t_{j,A} =& \frac{\left(\frac{1-\pi_A}{\pi_A}\right)^{\frac{1}{b-1}} - 1}{\left(\frac{1-\pi_A}{\pi_A}\right)^{\frac{1}{b-1}} + 1} \\ =& \frac{\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}} - 1}{\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}} + 1}.\end{aligned} \tag{3-13}$$

Note the following facts:

$$t_{j,A}\begin{cases} = 0, & \text{if } a = 0, \\ \to -1, & \text{as } a \to \infty \\ \to -1, & \text{as } b \to 1^+ \\ \to 0, & \text{as } b \to \infty \end{cases} \tag{3-14}$$

where $b \to 1^+$ is the right-handed limit.

For $b = 1$, then the decider has the lowest ability, and we have have that:

$$P\left(A|s_j\right) = \pi_j = \begin{cases} 1-2^{-a} & \text{if } i \text{ voted A} \\ 2^{-a} & \text{if } i \text{ voted B.} \end{cases}$$

This means that the decider will always follow the advisor's advice, when $a > 1$, or he will make a random choice with equal probability, when $a = 1$.

When $b > 1$ The quality is the probability that a correct verdict is given by the decider, which is:

$$\begin{aligned}
Q_{b>1}(a,b) =& \pi_0\left[P(s_i>0|A)P\left(s_j>t_{j,A}|A\right)+P(s_i<0|A)P\left(s_j>t_{j,B}|A\right)\right] \\
&+(1-\pi_0)\left[P(s_i<0|B)P\left(s_j<t_{j,B}|B\right)+P(s_i>0|B)P\left(s_j<t_{j,A}|B\right)\right] \\
=& \pi_0\left[(1-F_i(0))\left(1-F_j\left(t_{j,A}\right)\right)+F_i(0)\left(1-F_j\left(t_{j,B}\right)\right)\right] \\
&+(1-\pi_0)\left[G_i(0)G_j\left(t_{j,B}\right)+(1-G_i(0))\,G_j\left(t_{j,A}\right)\right] \\
=& \left[1-F_i(0)\right]\left(1-F_j\left(t_{j,A}\right)\right)+F_i(0)\left(1-F_j\left(t_{j,B}\right)\right) \\
=& 1-\left(\frac{\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}}{\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}+1}\right)^b+2^{-a}\left[\frac{\left(\frac{1}{2^a-1}\right)^{\frac{b}{b-1}}-1}{\left(\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}+1\right)^b}\right] \\
=& 1-2^{-a}\left[\frac{(2^a-1)\left(\frac{1}{2^a-1}\right)^{\frac{b}{b-1}}+1}{\left(\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}+1\right)^b}\right] \\
=& 1-\frac{2^{-a}}{\left(\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}+1\right)^{b-1}} \\
=& 1-\frac{2}{b}f_j(-t_{j,A})F_i(0)
\end{aligned}$$

When $b = 1$, the decider is correct if the advisor is correct, so the quality is:

$$\begin{aligned}
Q_{b=1}(a,b) =& \pi_0\left[P(s_i>0|A)P\left(s_j>t_{j,A}|A\right)+P(s_i<0|A)P\left(s_j>t_{j,B}|A\right)\right] \\
&+(1-\pi_0)\left[P(s_i<0|B)P\left(s_j<t_{j,B}|B\right)+P(s_i>0|B)P\left(s_j<t_{j,A}|B\right)\right] \\
=& \pi_0(1-F_i(0))+(1-\pi_0)G_i(0) \\
=& 1-F_i(0) \\
=& 1-2^{-a}.
\end{aligned}$$

The overall quality is:

$$Q(a,b) = \begin{cases} Q_{b=1}(a,b) & \text{if } b = 1 \\ Q_{b>1}(a,b) & \text{if } b > 1, \end{cases} \tag{3-15}$$

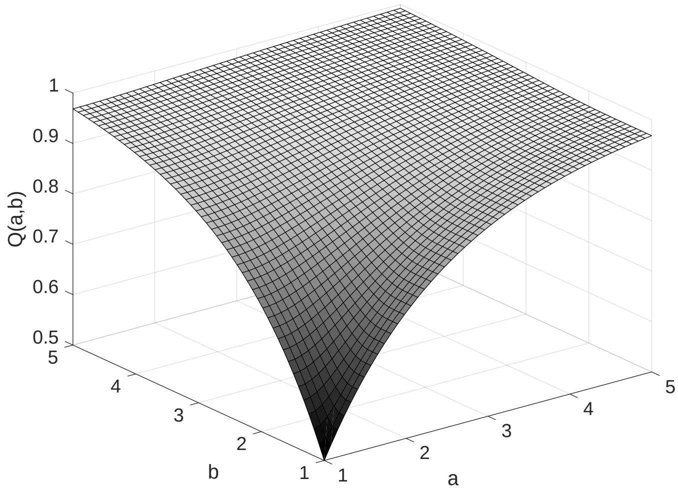

Figure 3-3: The quality for $a, b \in [1,5]$.

and a plot is given in Figure 3-3. When 2 jurors of the lowest abiltiy, then the quality is 1/2, and the quality goes to one as the jurors get better.

We now want to know which role assignment is optimal to maximise the quality. Assume now that $a \leq b$, so i has the same or has less ability then j. We can assume that $b > 1$, because role assignment does not matter when $a = b = 1$. The quality difference when i is the advisor or when j is the advisor is:

$$\begin{aligned}\Delta(a,b) =& Q(a,b) - Q(b,a)\\ =& \frac{2^{-b}}{\left(\left(\frac{1}{2^b-1}\right)^{\frac{1}{a-1}}+1\right)^{a-1}} - \frac{2^{-a}}{\left(\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}+1\right)^{b-1}}\\ =& \frac{2^b\left(\left(\frac{1}{2^b-1}\right)^{\frac{1}{a-1}}+1\right)^{a-1} - 2^a\left(\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}+1\right)^{b-1}}{2^{a+b}\left(\left(\frac{1}{2^b-1}\right)^{\frac{1}{a-1}}+1\right)^{a-1}\left(\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}+1\right)^{b-1}}.\end{aligned} \tag{3-16}$$

Numerical computations suggest that $\Delta(a,b) \geq 0$, when $a \leq b$, and it is strictly higher for some values. We have tested it for values $a, b \in [1,20]$ that lie on a two dimensional equidistant mesh of 100 by 100. See figure 3-4 for an example plot with a smaller range. Numerical results suggest the following conjecture:

Conjecture 1. *If $a \leq b$, then $\Delta(a,b) \geq 0$. In other words, the optimal role assignment that maximises the quality, is one where the juror with lesser ability is the advisor.*

The analytical proof that $\Delta(a,b) \geq 0$ is still an open problem. To give a start, note that the denominator of (3-16) is positive, we have that $\Delta_{ij} \geq 0$ if and only if:

$$2^a\left(\left(\frac{1}{2^a-1}\right)^{\frac{1}{b-1}}+1\right)^{b-1} - 2^b\left(\left(\frac{1}{2^b-1}\right)^{\frac{1}{a-1}}+1\right)^{a-1} \geq 0. \tag{3-17}$$

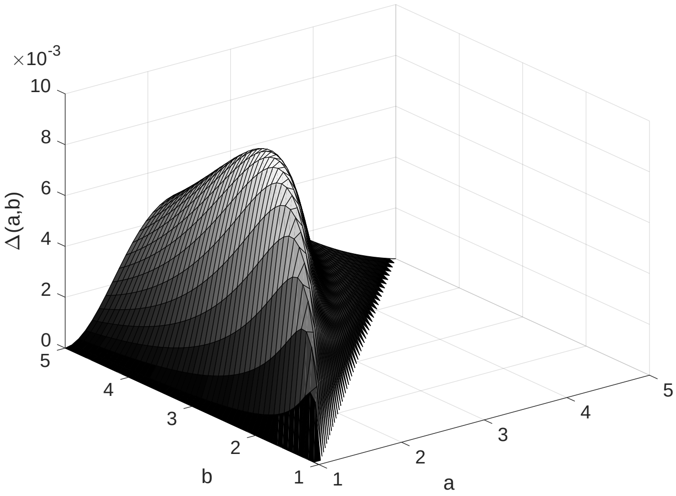

Figure 3-4: The quality difference for $a, b \in [1, 5]$ and $a \leq b$.

3-4 Gaussian Density

In this section, we use Gaussian density functions. When nature is A, the random signals are Gaussian distributed with mean 1 and when nature is B, the mean is -1. The juror's ability is represented by the variance, where jurors with higher ability have lower variance.

Let the signal space be $\mathscr{S} = \mathbb{R}$. Juror i has an ability level $a_i \in (0, \infty)$ and his signal is a random variable S_i, with distribution:

$$S_i \sim \mathcal{N}(1, a_i^2), \qquad \text{if nature is } A$$
$$S_i \sim \mathcal{N}(-1, a_i^2), \qquad \text{if nature is } B.$$

The density and cumulative distribution functions, when nature is A, are denoted as $f_i(s_i)$ and $F_i(s_i)$ respectively. When nature is B, these are denoted as $g_i(s_i)$ and $G_i(s_i)$. An example plot is given in Figure 3-5. Note that higher ability jurors have a lower a_i. This is different compared to the models with linear and β density functions, where higher a_i represent better jurors.

Let π_i be i's prior probability that nature is A. The probability that nature is A, when i receives a realisation of his random signal s_i is:

$$P(A|s_i) = \frac{f(s)\pi_i}{\pi_i f(s_i) + (1-\pi_i)g(s_i)}.$$

We can see that $P(A|s_i)$ is strictly increasing in s_i, because

$$\frac{g_i(s_i)}{f_i(s_i)} = e^{\frac{-4s_i}{2a_i^2}}$$

is strictly decreasing in s_i. This means that there exists some threshold t_i, such that $P(A|t_i) = 1/2$ and the juror would believe that $P(A|s_i) > 1/2$, if $s_i > t_i$.

It can be found that the solution for t_i is:

$$t_i = \frac{a_i^2}{2} \ln\left(\frac{1-\pi_i}{\pi_i}\right). \tag{3-18}$$

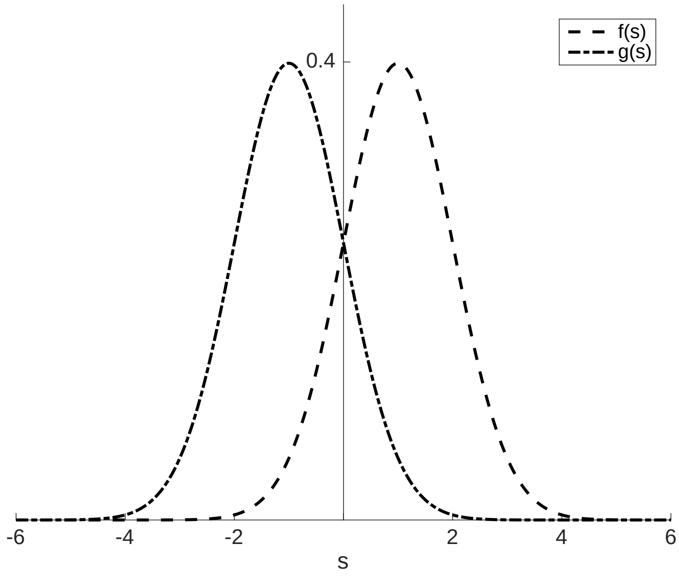

Figure 3-5: Example of probability density functions $f(s)$ and $g(s)$

Note the following facts:

$$t_i \begin{cases} = 0, & \text{if } \pi_i = \frac{1}{2} \\ \to -\infty, & \text{as } \pi_i \to 1 \\ \to \infty, & \text{as } \pi_i \to 0. \end{cases}$$

Consider 2 jurors where i is the advisor and j is the decider. Their abilities are a and b respectively. The advisor's prior is $\pi_i = \pi_0 = 1/2$, and therefore his threshold is $t_i = 0$.

When the decider j get advised A, then his prior is:

$$\begin{aligned}\pi_{j,A} = P(A|s_i > 0) =& \frac{\pi_0(1-F_i(0))}{\pi_0(1-F_i(0)) + (1-\pi_0)(1-G_i(0))} \\ =& 1 - F_i(0) \\ =& \frac{1}{2}\mathrm{erfc}\left(\frac{1}{-a\sqrt{2}}\right),\end{aligned}$$

and his threshold is:

$$t_{j,A} = \frac{b^2}{2}\ln\left(\frac{2-\mathrm{erfc}\left(\frac{1}{-a\sqrt{2}}\right)}{\mathrm{erfc}\left(\frac{1}{-a\sqrt{2}}\right)}\right).$$

Similarly, we can find that $\pi_{j,B} = 1 - \pi_{j,A}$ and $t_{j,B} = -t_{j,A}$, if the advisor voted B. The probability that a correct verdict is given by the decider, is the quality:

$$\begin{aligned}Q(a,b) =& \pi_0\left[P(s_i > 0|A)P\left(s_j > t_{j,A}|A\right) + P(s_i < 0|A)P\left(s_j > t_{j,B}|A\right)\right] \\ & + (1-\pi_0)\left[P(s_i < 0|B)P\left(s_j < t_{j,B}|B\right) + P(s_i > 0|B)P\left(s_j < t_{j,A}|B\right)\right] \\ =& \pi_0\left[(1-F_i(0))\left(1-F_j\left(t_{j,A}\right)\right) + F_i(0)\left(1-F_j\left(t_{j,B}\right)\right)\right] \\ & + (1-\pi_0)\left[G_i(0)G_j\left(t_{j,B}\right) + (1-G_i(0))\,G_j\left(t_{j,A}\right)\right] \\ =& \left[1-F_i(0)\right]\left(1-F_j\left(t_{j,A}\right)\right) + F_i(0)\left(1-F_j\left(t_{j,B}\right)\right)\end{aligned} \tag{3-19}$$

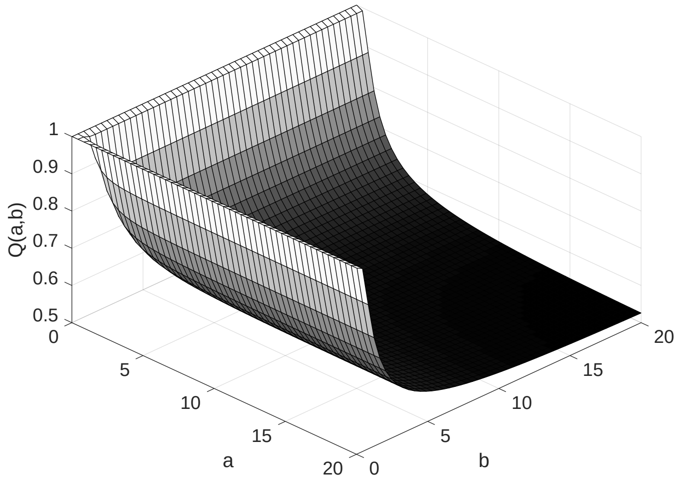

Figure 3-6: Plot of the quality.

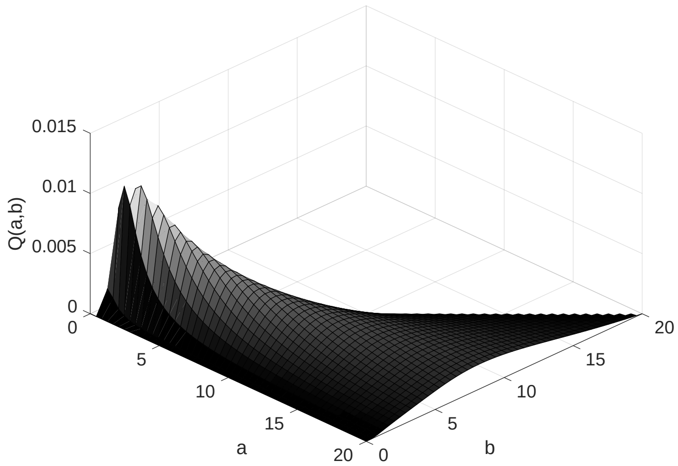

Figure 3-7: Plot of the quality difference when $a \geq b$.

A plot of the quality is shown in Figure 3-7. Note how the quality goes to one as the a and b gets lower, and the quality goes to 1/2 as the a and b gets higher.

We now want to know what optimal role assignment, when jurors have different abilities. Define the quality difference:

$$\Delta(a, b) = Q(a, b) - Q(b, a).$$

The values $\Delta(a, b)$ are computed for $a, b \in [0.01, 100]$ that lie on an equidistant mesh of size 500 by 500. For $a \geq b$, we find that $\Delta(a, b) \geq 0$. An example plot of $\Delta(a, b)$ on a smaller ability range is shown in Figure 3-7. Numerical results suggest the following conjecture:

Conjecture 2. *If $a \geq b$, then $\Delta(a, b) \geq 0$. In other words, the optimal role assignment that maximises the quality, is one where the juror with lesser ability is the advisor.*

3-5 Conclusion

We have given the advisor-decider mechanism, and analysed it with the linear-, β-, and Gaussian probability density functions. We wanted to find out what the optimal role assignment is for jurors with different ability levels. For linear densities, we have analytically shown that the worst juror should be the advisor. For β-, and Gaussian densities, numerical results also suggest that the worst juror should be the advisor. The analytical proofs are still an open problem. We conjecture that the worst juror should be the advisor for any density function that satisfy the properties in Definition 3.1.

Chapter 4

Three Member Casting Vote Scheme

Let the state of nature be either A or B. Inside a casting vote scheme, three jurors have to decide which state is most likely to be true. The jurors cannot directly observe the state of nature, but only random signals that are correlated with the state of nature. First, two jurors vote independently with respect to each other. If the decisions from these two jurors are not unanimous, than the third juror casts the final vote. A real world example, is the procedure of accepting a paper into a journal. The paper is first reviewed by two referees. If they don't have the same opinion, than the editor makes the final call.

Here we will study the model given by Alpern and Chen in [1]. According to their results, when jurors have different abilities, the median ability juror should have the casting vote, if we want to maximise the quality of the final verdict. Jurors receive random signal from a linear probability density function. We want to find out if this result is also true for a larger class of density functions.

4-1 The Casting Vote Scheme

In this section, we introduce the model of the casting vote scheme. First, two jurors will place their votes independently. We call them the *initial jurors*. If they both vote for the same state, then that state will have the majority vote, and it will be the final verdict. If they do not vote unanimously, then the third juror is the casting voter, who breaks the tie. With the knowledge of what each previous juror has voted, he will make a final decision. The question is: For jurors of different ability levels, who should be the tiebreaker, if we want the maximise the probability of getting the correct verdict?

The jurors cannot directly observe the state of nature, but only a random signal S_i, that is correlated with the state of nature. The random signal S_i takes values in $\mathscr{S}$, and the probability density function depends on the true state of nature, and it is:

$$\begin{aligned} &f_i(s_i), \quad \text{if nature is } A \\ &g_i(s_i), \quad \text{if nature is } B. \end{aligned}$$

The corresponding cumulative distributions are denoted as $F_i(s_i)$ and $G_i(s_i)$ respectively. We restrict ourselves to a class of densities functions that satisfy:

Definition 4.1.

1. $f_i(s_i)$ is continues in $\mathscr{S}$.
2. $f_i(s_i) = g_i(-s_i)$.
3. $g_i(s_i)/f_i(s_i)$ is non-increasing in s_i.

Throughout the chapter, we will often make use of the following fact:

$$F_i(s_i) = 1 - G_i(-s_i). \tag{4-1}$$

Jurors can have different probability density functions. We say that a juror i has a higher *ability* then j, if it is more likely for i to receive a higher signal, when nature is A, i.e. juror i's signal has first order stochastic dominance over j's signal. By symmetry, it is more likely for a higher ability juror to receive a lower signal, when nature is B. The probability densities for each juror is public knowledge.

The jurors make a decision in what is called a *Bayesian decision process.* There exist extensive literature on such processes, see [10], but we will deal with our example only. In addition to the observation s_i, each jurors holds prior believes. Juror i's prior belief that the probability nature is A, is π_0. We assume that the jurors are unbiased, so without any information A and B are equally probable; The default prior probability that nature is A is $\pi_0 = 1/2$.

After i receives the observation s_i, then his posterior believe that nature is A is:

$$P(A|s_i) = \frac{\pi_i f_i(s_i)}{\pi_i f_i(s_i) + (1-\pi_i) g_i(s_i)}.$$

The juror will then:

- Vote A, if $P(A|s_i) > 1/2$.
- Vote B, if $P(A|s_i) < 1/2$.
- Make a random choice between A and B with equal probability, if $P(A|s_i) = 1/2$.

We can see that $P(A|s_i) = \frac{\pi_i}{\pi_i + (1-\pi_i) g_i/f_i}$ is non-decreasing in s_i, because g_i/f_i is non-increasing in s_i. The higher s_i is, the stronger he believes that nature is A. We do not have to worry about the case when $f_i(s_i) = 0$, because the jurors will never receive such signals.

4-1-1 Voting Procedure and Quality of Verdict

Let there be three jurors i, j, and k, where i and j are the initial jurors, and k is the casting juror. The jurors i and j, have the prior $\pi_i = \pi_j = \pi_0 = 1/2$.

The initial jurors receive the values of their random signals and vote for the state they believe is most likely. If the both vote for the same state, then that state is the final verdict. If both jurors vote oppositely, let us say i votes A and j votes B, then the casting juror's updated prior is:

$$\begin{aligned}\pi_j = \pi_{AB} =& P(A|i \text{ votes } A \text{ and } j \text{ votes } B)\\ =& \frac{P(A)P(i \text{ votes } A \text{ and } j \text{ votes } B|A)}{P(i \text{ votes } A \text{ and } j \text{ votes } B)}\\ =& \frac{P(A)P(i \text{ votes } A|A)P(j \text{ votes } B|A)}{P(i \text{ votes } A \text{ and } j \text{ votes } B)}\end{aligned} \tag{4-2}$$

With the updated prior and his own signal, the casting juror will give the final verdict.

Let q_A be the probability that the final verdict is A, on the condition that nature is A:

$$\begin{aligned}q_A =& P(i \text{ and } j \text{ vote } A|\text{A})\\ &+ P(i \text{ votes } A,\ j \text{ votes } B, \text{ and } k \text{ votes } A|\text{A})\\ &+ P(i \text{ votes } B,\ j \text{ votes } A, \text{ and } k \text{ votes } A|\text{A}).\end{aligned}$$

The probability q_B is defined the same way, but for B, instead of A. The quality Q is the probability that a correct verdict is given:

$$Q = \pi_0 q_A + (1-\pi_0) q_B. \tag{4-3}$$

4-2 Linear Density Functions

In this section, we use linear density functions, which are the same used by Alpern and Chen in [1]. Each juror's signal, is a random variable taking values in $\mathscr{S} = [-1,1]$ and has the probability density function:

$$\begin{aligned}f_i(s) &= \frac{1+a_i s}{2} \quad \text{when nature is A}\\ g_i(s) &= \frac{1-a_i s}{2} \quad \text{when nature is B.}\end{aligned}$$

An example plot is given in Figure 4-1. The cumulative distribution functions are:

$$\begin{aligned}F_i(s) &= \frac{(s+1)(a_i s - a_i + 2)}{4} \quad \text{when nature is A}\\ G_i(s) &= \frac{(s+1)(a_i - a_i s + 2)}{4} \quad \text{when nature is B.}\end{aligned}$$

Let there be three jurors i, j and k, with ability levels a, b and c respectively. The initial jurors are i and j, and the tiebreaker is k. The jurors i and j believe that the prior probability nature is A,

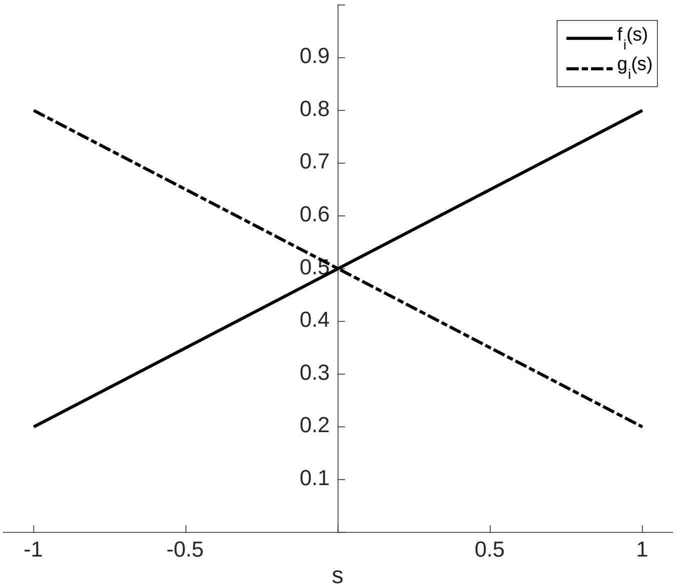

Figure 4-1: Example of linear distribution functions with $a_i = 0.6$. When nature is A the distribution is $f_i(s)$, and when nature is B the distribution is $g_i(s)$.

is $\pi_i = \pi_j = \pi_0 = 1/2$. When an initial juror, let us say i, receives his signal s_i, then he believes the state is A with posterior probability:

$$\begin{aligned} P(A|s_i) &= \frac{\pi_0 f_i(s)}{\pi_0 f_i(s_i) + (1-\pi_0)\, g_i(s_i)} \\ &= \frac{1+as}{2} \end{aligned}$$

We can see that a juror with ability $a = 0$ is no better than a fair coin flip. When $a > 0$, we have $P(A|s_i) > \frac{1}{2}$, if and only if $s_i > 0$. The juror votes A if his signal is $s_i > 0$, and he votes B if his signal is $s_i < 0$. The same holds for juror j.

When i and j both vote for the same state of nature, then that state is the final verdict. If they vote opposite compared to each other, then the casting juror takes both votes and his own signal into consideration to pass a final verdict. The tiebreaker only knows what each initial juror voted and not the signals they receive. Juror j's updated prior, when i votes A and j votes B, is:

$$\begin{aligned} \pi_{AB} &= \frac{\pi_0\,(1-F_i(0))\,F_j(0)}{\pi_0\,(1-F_i(0))\,F_j(0) + (1-\pi_0)\,(1-G_i(0))\,G_j(0)} \\ &= \frac{(2+a)(2-b)}{8-2ab} \\ &= \frac{1}{2} + \frac{a-b}{4-ab} \end{aligned}$$

Similarly, we find that $\pi_{BA} = 1 - \pi_{AB}$. Note the following facts:

$$\pi_{AB} \begin{cases} = 0, & \text{if } a = b, \\ > 1/2, & \text{if } a > b, \\ < 1/2, & \text{if } a < b. \end{cases} \tag{4-4}$$

So when the initial jurors have the same ability and vote oppositely, then the tiebreaker is not better informed than before. When the initial jurors have different abilities, then the updated prior will be biased towards the decision of the better juror.

When i votes A and j votes B, then the tiebreaker believes the probability that nature is A, is:

$$P(A|i \text{ votes } A, j \text{ votes } B, s_k) = \frac{\pi_{AB} f_k(s_k)}{\pi_{AB} f_k(s_k) + (1-\pi_{AB})\, g_k(s_k)}.$$

When $c = 0$, then $P(A|i \text{ votes } A, j \text{ votes } B, s_k) = \pi_{AB}$, so the worst ability tiebreaker always believes his prior, and he follows the decision of the juror with the highest ability.

For $c > 0$, there exist a threshold t_{AB}, such that $P(A|i \text{ votes } A, j \text{ votes } B, s_k) > 1/2$, if $s_k > t_{AB}$. The threshold can be found by solving:

$$\begin{aligned} P(A|i \text{ votes } A, j \text{ votes } B, t_{AB}) &= \frac{\pi_{AB} + c t_{AB} \pi_{AB}}{2c t_{AB} \pi_{AB} - c t_{AB} + 1} \\ &= \frac{1}{2}. \end{aligned}$$

The solution is:

$$t_{AB}(a,b,c) = \frac{1-2\pi_{AB}}{c} = \frac{2(b-a)}{c(4-ab)} \tag{4-5}$$

Similarly, we find that $t_{BA} = -t_{AB}$.

When $\pi_{AB} > 1/2$, then $t_{AB} < 0$. For negative signals s_k, that are close enough to zero, the tiebreaker votes A, while the initial jurors vote B for any negative signal. We also have that $t_{AB} > 0$, if $\pi_{AB} < 1/2$. From (4-4), we know the tiebreaker's prior is biased towards the decision of the better juror. This implies that the threshold is negative if the better juror voted A, and positive if the better juror voted B. In the case that $c \le |1-2\pi_{AB}|$, then $|t_{AB}| \ge 1$. So if the tiebreaker's ability is very low, then he will always follow the decision of the better initial juror.

4-2-1 Quality of Verdict

The quality Q is the probability that a correct verdict is given at the end:

$$Q = \pi_0 q_A + (1-\pi_0) q_B, \tag{4-6}$$

where q_A is the probability that the verdict is A, conditionally that nature is A. The same definition holds for q_B.

For $c > |1-2\pi_{AB}|$, we have that $|t_{AB}| < 1$, and:

$$\begin{aligned} q_A = &(1-F_i(0))\left(1-F_j(0)\right) \\ &+ (1-F_i(0))\, F_j(0)\, (1-F_k(t_{AB})) \\ &+ F_i(0)\left(1-F_j(0)\right)(1-F_k(t_{BA})) \end{aligned} \tag{4-7}$$

and

$$\begin{aligned} q_B &= G_i(0)G_j(0) + (1-G_i(0))\, G_j(0) G_k(t_{AB}) + G_i(0)\left(1-G_j(0)\right) G_k(t_{BA}) \\ &= (1-F_i(0))\,(1-F_b(0)) + F_i(0)\left(1-F_j(0)\right) G_k(-t_{BA}) + (1-F_i(0))\, F_j(0) G_k(-t_{AB}) \\ &= (1-F_i(0))\left(1-F_j(0)\right) + F_i(0)\left(1-F_j(0)\right)(1-F_k(t_{BA})) \\ &\quad + (1-F_i(0))\, F_j(0)\, (1-F_k(t_{AB})) \\ &= q_A. \end{aligned}$$

So we have $Q = q_A \equiv q$. Substituting $F_a(0)$, $F_b(0)$ and t_{AB} into 4-7, we get:

$$q(a,b,c) = \frac{1}{32}\left(4(4+a+b) + \frac{4(a-b)^2}{(4-ab)c} + (4-ab)c\right) \tag{4-8}$$

For the case $c \le |1-2\pi_{AB}|$ (including $c = 0$), the tiebreaker always follows the highest ability initial juror. Let $F(s)$ and $G(s)$ be the cumulative distribution functions of the highest ability juror. So the quality, when $c < |1-2\pi_{AB}|$ or $c = 0$, is:

$$\begin{aligned} Q =& \pi_0(1-F(0)) + (1-\pi_0)G(0) \\ =& (1-F(0)) = q_A \\ =& \frac{\max\{a,b\}+2}{4}. \end{aligned}$$

Combining (4-7) and (4-8), the quality is:

$$Q(a,b,c) = \begin{cases} q(a,b,c) & \text{if } t_{AB}(a,b,c) < 1 \\ \frac{\max\{a,b\}+2}{4} & \text{if } t_{AB}(a,b,c) \ge 1 \end{cases} \tag{4-9}$$

With different ability jurors, we want to know what the optimal role assignment is. It turns out the median ability juror should always be the tiebreaker.

Theorem 4.1 (Alpern and Chen)**.** *Given any set of jurors with abilities $a,b,c \in [0,1]$ and $a \le b \le c$, then the quality is maximised if the juror with median ability b is the tie breaker, i.e. $Q(a,c,b) \ge Q(a,b,c)$ and $Q(a,c,b) \ge Q(b,c,a)$.*

Proof. Let us first define

$$\begin{aligned} \Delta_1(a,b,c) &= Q(a,c,b) - Q(a,b,c) \\ \Delta_2(a,b,c) &= Q(a,c,b) - Q(b,c,a). \end{aligned}$$

We want to show that $\Delta_1(a,b,c) \ge 0$ and $\Delta_2(a,b,c) \ge 0$, for any set of abilities in $K = \{(a,b,c) : 0 \le a \le b \le c \le 1\}$.

First, note the following facts, which can be confirmed using 4-5:

$$t_{AB}(b,c,a) = \frac{2(c-b)}{a(4-bc)} \ge 1 \iff P_1(a,b,c) \equiv abc - 4a - 2b - 2c \ge 0 \tag{4-10}$$

$$t_{AB}(a,c,b) = \frac{2(c-a)}{b(4-ac)} \ge 1 \iff P_2(a,b,c) \equiv abc - 2a - 4b + 2c \ge 0 \tag{4-11}$$

$$t_{AB}(a,b,c) = \frac{2(b-a)}{c(4-ab)} \le \frac{2b}{3c} < 1. \tag{4-12}$$

Next we will look at the two possible cases $t_{AB}(a,c,b) \ge 1$ and $t_{AB}(a,c,b) < 1$.

Case 1: $t_{AB}(a,c,b) \ge 1$. According to 4-9 we have that

$$Q(a,c,b) = \frac{c+2}{4}$$

Since $P_1(a,b,c) - P_2(a,b,c) = 2(b-a) \geq 0$, using 4-10 and 4-11, we can see that:

$$t_{AB}(a,c,b) \geq 1 \implies P_2(a,b,c) \geq 0 \implies P_1(a,b,c) \geq 0 \implies t_{AB}(b,c,a) \geq 1.$$

This implies $Q(b,c,a) = (c+2)/4$ and therefore:

$$\Delta_2(a,b,c) = Q(a,c,b) - Q(b,c,a) = 0.$$

The next step is to show that $\Delta_1(a,b,c) \geq 0$. Using 4-12 and 4-9, we know that $Q(a,b,c) = q(a,b,c)$ and therefore

$$\Delta_1(a,b,c) = \frac{c+2}{4} - q(a,b,c) = \frac{d_1(a,b,c)}{32c(4-ab)}, \tag{4-13}$$

where

$$d_1(a,b,c) \equiv -4a^2 + 8ab - 4b^2 - 16ac - 16bc + 4a^2bc + 4ab^2c + 16c^2 - a^2b^2c^2.$$

The denominator of 4-13 is always positive, so we only need to show that $d_1(a,b,c) \geq 0$. Note that according to 4-11, we have that $P_2(0,b,c) = -4b + 2c \geq 0$, so $c \geq 2b$. We are going to prove that $d_1(a,b,c) \geq 0$ with the following two steps:

1. $d_1(a,b,c)$ is non-decreasing in c, for $c \geq 2b$.
2. $d_1(a,b,2b) \geq 0$, for $a \leq b \leq 1/2$.

For step 1, the partial derivative of $d_1(a,b,c)$ with respect to c is:

$$\begin{aligned}\frac{\partial d_1(a,b,c)}{\partial c} &= -16(a+b) + 32c + 4a^2b + 4ab^2 - 2a^2b^2c \\ &\geq 0,\end{aligned}$$

because $0 \leq a \leq b \leq c \leq 1$, implies $32c \geq 16(a+b)$, and $4a^2b \geq 2a^2b^2c$. We have shown that $d_1(a,b,c)$ is non-decreasing in c, for $c \geq 2b$. For step 2, we have that:

$$d_1(a,b,2b) = -4a^2 - 24ab + 28b^2 + 8a^2b^2 + 8ab^3 - 4a^2b^4.$$

The partial derivative of $d_1(a,b,2b)$ with respect to b is:

$$\begin{aligned}\frac{\partial d_1(a,b,2b)}{\partial b} &= -24a + 56b + 16a^2b + 24ab^2 - 16a^2b^3 \\ &\geq 0,\end{aligned}$$

because $56b \geq 24a$ and $16a^2b \geq 16a^2b^3$. Since $d_1(a,b,2b)$ is non-decreasing in b, and $d_1(a,a,2a) = 16a^4 - 4a^6 \geq 0$, we have that $d_1(a,b,2b) \geq 0$.

Case 2: $t_{AB}(a,c,b) < 1$. We have $Q(a,c,b) = q(a,c,b)$, which implies

$$\Delta_1(a,b,c) = q(a,c,b) - q(a,b,c) = \frac{(c-b)d_2(a,b,c)}{8bc(4-ab)(4-ac)},$$

where

$$d_2(a,b,c) = 4a^2 - 8ab + 4b^2 - 8ac + 4bc + 2a^2bc - ab^2c + 4c^2 - abc^2.$$

If we can show $d_2(a,b,c) \geq 0$, then $\Delta_2(a,b,c) \geq 0$. This can be done by taking partial derivatives of $d_2(a,b,c)$. The partial derivative to b is

$$\frac{\partial d_2(a,b,c)}{\partial b} = 8(b-a) + c(4+2a^2-2ab-ac).$$

We can show that $\frac{\partial d_2(a,b,c)}{\partial b} \geq 0$. First of all $8(b-a) \geq 0$ and secondly

$$c(4+2a^2-2ab-ac) \geq c(4+0-2-1) = c \geq 0.$$

With a similar argument, we also have

$$\frac{\partial d_2(a,b,c)}{\partial c} = 8(b-a) + b(4+2a^2-2ac-ab) \geq 0.$$

Since the partial derivatives of $d_2(a,b,c)$ to b and c are positive, $d_2(a,b,c)$ is an increasing function in b and c. So we can conclude $d_2(a,b,c) \geq d_2(a,a,a) = 0$ and consequently $\Delta_2(a,b,c) \geq 0$.

The last thing to show is $\Delta_2(a,b,c) \geq 0$. We consider two separate cases: $t_{AB}(b,c,a) \geq 1$ and $t_{AB}(b,c,a) \leq 1$.

If $t_{AB}(b,c,a) \geq 1$, then $Q(b,c,a) = \frac{c+2}{4}$ and

$$\Delta_2(a,b,c) = q(a,c,b) - \frac{c+2}{4} = \frac{(P_2(a,b,c))^2}{32b(4-ac)} \geq 0.$$

If $t_{AB}(b,c,a) \leq 1$, then

$$\Delta_2(a,b,c) = q(a,c,b) - q(b,c,a) = \frac{(b-a)d_3(a,b,c)}{8ab(4-ac)(4-bc)},$$

where

$$d_3(a,b,c) = -4a^2 - 4ab - 4b^2 + 8ac + 8bc + a^2bc + ab^2c - 4c^2 - 2abc^2.$$

In Appendix A we show that the following optimisation problem:

$$\begin{aligned} \min_{a,b,c} \; & d_3(a,b,c) \\ \text{s.t. } & P_1(a,b,c) \leq 0 \\ & 0 \leq a \leq b \leq c \leq 1, \end{aligned}$$

has the optimal objective function value $d_3 = 0$.

■

4-3 β-Density Functions

With the β-distribution, we use the signal space $\mathscr{S} = [-1,1]$. A juror i has a ability level $a_i \in [1,\infty)$, and his probability density function is:

$$f_i(s_i) = \frac{a_i}{2}\left(\frac{1+s_i}{2}\right)^{a_i-1} \quad \text{if nature is } A$$
$$g_i(s_i) = \frac{a_i}{2}\left(\frac{1-s_i}{2}\right)^{a_i-1} \quad \text{if nature is } B,$$

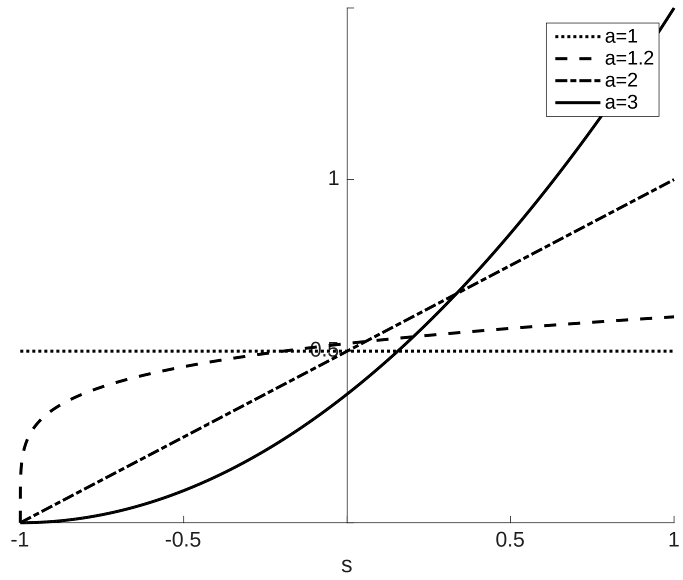

Figure 4-2: Example of the beta distribution functions $f(s)$ with different a.

His cumulative distribution function is:

$$F_i(s_i) = \left(\frac{1+s_i}{2}\right)^{a_i} \quad \text{if nature is } A$$
$$G_i(s_i) = 1 - \left(\frac{1-s_i}{2}\right)^{a_i} \quad \text{if nature is } B.$$

Examples of the beta distribution functions for different ability levels are shown in Figure (4-2). The advantage of using beta density functions over the linear densities, is that we can model very competent jurors. With a prior $\pi_i = 1/2$, the probability that i gives the correct verdict goes to one as the ability level goes to infinity, while in the model with linear densities, the best juror is only correct with probability $3/4$.

Similar to the previous section, we have three jurors i, j and k, where i and j are the initial jurors, and k is the tiebreaker. The ability levels for i, j and k, are a, b and c respectively. The private signals corresponding to each juror are s_i, s_j and s_k. The initial juror's prior are $\pi_i = \pi_j = \pi_0 = 1/2$. An initial juror, for example i, believes the probability that nature is A, is:

$$\begin{aligned} P(A|s_i) =& \frac{\pi_0 f_i(s_i)}{\pi_0 f_i(s_i) + (1-\pi_0) g_i(s_i)} \\ =& \frac{(1+s_i)^{a-1}}{(1+s_i)^{a-1} + (1-s_i)^{a-1}} \end{aligned} \tag{4-14}$$

For the lowest ability level, $a = 1$, we have that $P(A|s_i) = 1/2$, and in such case, the juror will randomly select between A and B with equal probability.

When the initial jurors vote for the same state, then we are done, and that state is the final verdict. If they vote opposite to each other, the casting juror has to make the final call. The tiebreaker's updated prior, when i votes A, and j votes B, is:

$$\pi_{AB}(a,b,c) = \frac{\pi_0\,(1-F_i(0))\,F_j(0)}{\pi_0\,(1-F_i(0))\,F_j(0)+(1-\pi_0)\,(1-G_i(0))\,G_j(0)}$$
$$= \frac{2^a-1}{2^a+2^b-2}$$

Note the following facts:

$$\pi_{AB}(a,b,c)\begin{cases} = 1/2, & \text{if } a=b,\\ < 1/2, & \text{if } a<b,\\ > 1/2, & \text{if } a>b,\\ \to 0, & \text{as } b\to\infty,\\ \to 1, & \text{as } a\to\infty. \end{cases}$$

Similarly, we can find that the prior is updated to $\pi_{BA} = 1-\pi_{AB}$, when i votes B, and j votes A.

The probability that the tiebreaker believes that nature is A, is:

$$P(A|i \text{ votes } A, j \text{ votes } B, s_k) = \frac{\pi_{AB}\,f_k(s_k)}{\pi_{AB}\,f_k(s_k)+(1-\pi_{AB})\,g_k(s_k)}.$$

When the tiebreaker has the lowest ability, $c=1$, then his posterior is the same as his prior. Which means that he follows the juror with the highest ability.

For $c>1$, there is some threshold t_{AB}, such that

$$P(A|i \text{ votes } A, j \text{ votes } B, s_k) > 1/2,$$

for $s_k > t_{AB}$. This threshold is:

$$t_{AB}(a,b,c) = \frac{\left(\frac{1-\pi_{AB}}{\pi_{AB}}\right)^{\frac{1}{c-1}}-1}{\left(\frac{1-\pi_{AB}}{\pi_{AB}}\right)^{\frac{1}{c-1}}+1}$$
$$= \frac{\left(\frac{2^b-1}{2^a-1}\right)^{\frac{1}{c-1}}-1}{\left(\frac{2^b-1}{2^a-1}\right)^{\frac{1}{c-1}}+1}.$$

4-3-1 Quality of Verdict

The quality Q is the probability that a correct verdict is given. We want to know for jurors with different abilities, which role assignment maximises Q.

Let us denote the cumulative distribution function of the best initial juror as $F(s)$ and $G(s)$. When $c=1$, then that quality is the probability that the best initial juror gives the correct verdict:

$$Q(a,b,c) = \pi_0\,(1-F(0))+(1-\pi_0)G(0)$$
$$= (1-F(0))$$
$$= \left(\frac{1}{2}\right)^{\max(a,b)}.$$

Table 4-1: Quality comparison for different tiebreakers. Abilities are sorted as $a \leq b \leq c$.

Ability levels (a, b, c)	$Q(b, c, a)$	$Q(a, c, b)$	$Q(a, b, c)$
(1.0, 1.1, 2.3)	0.66267	0.79180	0.79180
(3.2, 4.1, 4.6)	0.98770	0.98867	0.98749
(2.1, 2.2, 2.3)	0.88060	0.88115	0.88052
(1.0, 1.6, 3.3)	0.78646	0.89797	0.89735
(1.4, 2.5, 3.9)	0.90663	0.94385	0.93426
(1.9, 3.3, 4.3)	0.95969	0.96985	0.96117
(2.6, 4.4, 5.0)	0.98753	0.98936	0.98647
(2.8, 3.8, 4.3)	0.98047	0.98234	0.98011
(3.9, 4.7, 4.9)	0.99430	0.99443	0.99417
(1.4, 2.1, 2.8)	0.85401	0.87868	0.86611

When $c > 1$, the quality is:

$$
\begin{aligned}
Q(a,b,c) =& q(a,b,c) \\
=& (1 - F_i(0)) \left(1 - F_j(0)\right) \\
&+ (1 - F_i(0)) F_j(0) (1 - F_k(t_{AB})) \\
&+ F_i(0) \left(1 - F_j(0)\right) (1 - F_k(-t_{AB})) \\
=& 2^{-a-b} \left(-\left(2^a - 1\right) \left(\frac{1}{\left(\frac{2^b-1}{2^a-1}\right)^{\frac{1}{1-c}} + 1} \right)^a - \left(2^b - 1\right) \left(\frac{1}{\left(\frac{2^b-1}{2^a-1}\right)^{\frac{1}{c-1}} + 1} \right)^a + 2^{a+b} - 1 \right),
\end{aligned}
$$

where we have used the fact that $q_A = q_B \equiv q$.

For jurors with different ability levels, we want to know which juror should be the casting juror, if we want to maximise the quality of the final verdict. We perform a numerical experiment where we randomly select 100000 ability sets (a, b, c), in the interval $[1, 20]$. When we compute and compare qualities for different casting juror roles, it turns out that the median ability juror should be the tiebreaker. See Table 4-1 for a couple of examples. Note that the quality is the highest when the median ability juror is the tiebreaker. Based on the numerical results, we give the following conjecture:

Conjecture 3. *For β-density functions, the quality is maximised when the median juror has the casting vote i.e., for $1 \leq a \leq b \leq c$:*

$$
\begin{aligned}
\Delta_1(a,b,c) = Q(a,c,b) - Q(a,b,c) \geq 0 \\
\Delta_2(a,b,c) = Q(a,c,b) - Q(b,c,a) \geq 0.
\end{aligned}
$$

4-4 Gaussian Signals

We will now study a model with Gaussian densities. When nature is A, then the signals are Gaussian distributed with mean 1 and when nature is B, the mean is -1. The juror's ability is represented by the variance, where jurors with higher ability have lower variance.

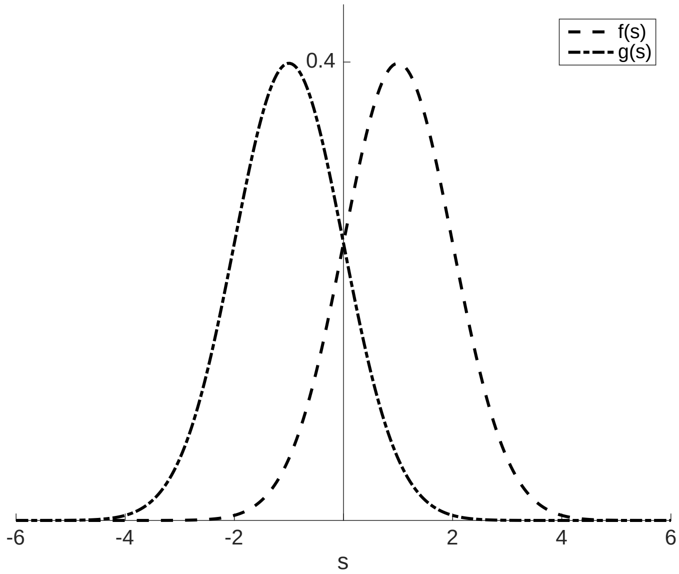

Figure 4-3: Example of probability density functions $f(s)$ and $g(s)$

Let the signal space be $\mathscr{S} = \mathbb{R}$. Juror i has an ability level $a_i \in (0, \infty)$ and his signal is a random variable S_i, with distribution:

$$\begin{aligned} S_i &\sim \mathcal{N}(1, a_i^2), && \text{if nature is } A \\ S_i &\sim \mathcal{N}(-1, a_i^2), && \text{if nature is } B. \end{aligned}$$

The density and cumulative distribution functions, when nature is A, are denoted as $f_i(s_i)$ and $F_i(s_i)$ respectively. When nature is B, these are denoted as $g_i(s_i)$ and $G_i(s_i)$. An example plot is given in Figure 4-3. Note that higher ability jurors have a lower a_i. This is different compared to the models with linear and β density functions, where higher a_i represent better jurors.

Let π_i be i's prior probability that nature is A. The probability that nature is A, when i receives a realisation s_i of his random signal is:

$$P(A|s_i) = \frac{f(s_i)\pi_i}{\pi_i f(s_i) + (1-\pi_i)g(s_i)}.$$

We can see that $P(A|s_i)$ is strictly increasing in s_i, because

$$\frac{g_i(s_i)}{f_i(s_i)} = e^{\frac{-4s_i}{2a_i^2}}$$

is strictly decreasing in s_i. This means that there exists some threshold t_i, such that $P(A|t_i) = 1/2$ and the juror would believe that $P(A|s_i) > 1/2$, if $s_i > t_i$.

Same as before, we have three jurors i, j and k, with abilities a, b and c respectively. Jurors i and j are the initial jurors, and k is the casting juror.

The initial jurors have a prior $\pi_i = \pi_j = \pi_0 = 1/2$, so their thresholds are $t_i = t_j = 0$, and they vote A, if their signals are higher than the threshold and they vote B otherwise.

In case the initial jurors vote opposite to each other, then k has to break the tie. When i votes A, and j votes B, then the tiebreakers updated prior is:

$$\begin{aligned}\pi_{AB}(a,b) =& \frac{\pi_0(1-F_i(0))F_j(0)}{\pi_0(1-F_i(0))F_j(0)+(1-\pi_0)(1-G_i(0))G_j(0)}.\\ =& \frac{\left(\operatorname{erfc}\left(\frac{1}{\sqrt{2}a}\right)-2\right)\operatorname{erfc}\left(\frac{1}{\sqrt{2}b}\right)}{\left(\operatorname{erfc}\left(\frac{1}{\sqrt{2}a}\right)-2\right)\operatorname{erfc}\left(\frac{1}{\sqrt{2}b}\right)+\left(\operatorname{erfc}\left(-\frac{1}{\sqrt{2}a}\right)-2\right)\operatorname{erfc}\left(-\frac{1}{\sqrt{2}b}\right)}\end{aligned}$$

Note the following facts:

$$\pi_{AB}\begin{cases}\to 1-F_i(0) & \text{as } b\to\infty\\ \to F_j(0) & \text{as } a\to\infty\\ \to 0 & \text{as } b\to 0\\ \to 1 & \text{as } a\to 0\\ =1/2 & \text{if } a=b\\ >1/2 & \text{if } a<b\\ <1/2 & \text{if } a>b.\end{cases} \tag{4-15}$$

The casting voter's threshold is:

$$t_{AB}(a,b,c)=\frac{c^2}{2}\ln\left(\frac{1-\pi_{AB}(a,b)}{\pi_{AB}(a,b)}\right),$$

where the following facts are true:

$$t_i\begin{cases}=0, & \text{if } \pi_{AB}=\frac{1}{2}\\ \to -\infty, & \text{as } \pi_{AB}\to 1\\ \to \infty, & \text{as } \pi_{AB}\to 0\\ \to 0, & \text{as } c\to 0.\end{cases}$$

Similarly, we can show that $\pi_{BA}=1-\pi_{AB}$, and $t_{AB}=-t_{BA}$. The probability a correct final verdict is given, is:

$$\begin{aligned}Q(a,b,c) =& q_A\\ =& (1-F_i(0))(1-F_j(0))\\ &+(1-F_i(0))F_j(0)(1-F_k(t_{AB}))\\ &+F_i(0)(1-F_j(0))(1-F_k(-t_{AB})),\end{aligned}$$

where we have used the fact that $q_A=q_B$.

When we have 3 jurors with different abilities, we want to know in which order they should vote to maximise the probability of having the correct verdict. The order of the first two jurors does not matter, so we only want to know who should be the tiebreaker: the worst-, median-, or best juror.

With Matlab, we generate random abilities that are uniformly selected in $[0,100]$. We simulate 1000000 random ability sets (a,b,c), and we compare the qualities with different tiebreakers. From the numerical results, it turns out that the median ability juror should have the casting vote. Based on the numerical results, we propose the following conjecture:

Table 4-2: Quality comparison. Example with 10 different set of jurors. The abilities (a,b,c) are ordered so that $a \le b \le c$.

(a,b,c)	$Q(a,b,c)$	$Q(a,c,b)$	$Q(c,b,a)$
(23.5, 73.0, 94.5)	0.51699	0.51704	0.51334
(2.9, 38.9, 75.4)	0.63540	0.63540	0.57549
(17.2, 17.2, 63.8)	0.52632	0.52823	0.52823
(15.9, 21.1, 98.3)	0.52541	0.52755	0.52539
(3.7, 5.3, 95.1)	0.60582	0.61217	0.59945
(1.0, 77.9, 97.0)	0.83790	0.83790	0.67355
(41.0, 51.5, 61.8)	0.51206	0.51219	0.51199
(8.7, 79.8, 89.7)	0.54580	0.54580	0.52763
(6.4, 23.9, 85.6)	0.56219	0.56220	0.54212
(64.7, 78.0, 79.2)	0.50819	0.50820	0.50816

Conjecture 4. *For jurors with abilities $a \le b \le c$, where lower values represent more reliable jurors, the quality Q is maximised when the median ability juror has the casting vote i.e.,*

$$\Delta_1(a,b,c) = Q(a,c,b) - Q(b,c,a) \ge 0$$
$$\Delta_2(a,b,c) = Q(a,c,b) - Q(a,b,c) \ge 0.$$

4-4-1 Looking at the limit case.

We can confirm that Conjectures 3 and 4, are true in the limit case. We will look at the case with Gaussian densities. The case with β-densities is similar.

Regarding $\Delta_1(a,b,c)$, the qualities have to following limits:

$$\lim_{a\to 0} Q(a,c,b) = 1 \tag{4-16}$$

$$\begin{aligned}\lim_{a\to 0} Q(b,c,a) =& (1-F_j(0))(1-F_k(0)) \\ &+ (1-F_j(0))F_k(0) \\ &+ F_j(0)(1-F_k(0)) \\ =& P(\text{ atleast one initial juror votes } A \mid A).\end{aligned} \tag{4-17}$$

The limit $a \to 0$, means that juror i is very good. In Equation (4-16), when i is an initial juror, then the probability that A will be selected, goes to one. Even if the other initial juror votes B, the casting juror will follow juror i. In Equation (4-17), when i has the casting vote, there remains the possibility that both initial jurors vote B, and the very good juror is not used at all. That is why (4-17) is never better than (4-16).

Regarding $\Delta_2(a,b,c)$, we have to following limits:

$$\lim_{c\to\infty} Q(a,c,b) = \frac{1}{2}(1-F_i(0)) + \frac{1}{2}\left[(1-F_i(0))(1-F_j(t_{AB})) + F_i(0)(1-F_j(-t_{AB}))\right] \tag{4-18}$$

$$\begin{aligned}\lim_{c\to\infty} Q(a,b,c) =& (1-F_i(0)) \\ =& P(\text{ juror i votes } A \mid A).\end{aligned} \tag{4-19}$$

Here juror k, with ability c, is very bad. If a very bad juror is a the casting voter, then he will always follow the best juror. We can see in (4-19), that the quality is just the probability that the best juror is correct. In (4-18), the last term is one half times the quality in the advisor-decider mechanism. An advisor-decider mechanism is always at least as good as a single juror mechanism; A single juror mechanism is the same as an advisor-decider mechanism, with a very bad decider, who always follows the advisor, or a very bad advisor, who is no better than a coin flip. This implies that (4-18) is at least as good as (4-19).

4-5 Conclusion

In this chapter we studied the casting vote scheme problem, with 3 jurors. We wanted to find out, who should have the casting vote, when jurors have different abilities. We gave the results of Alpern and Chen [1] for linear density functions, where the median ability juror should be the tiebreaker. We wanted to know if these results holds true for a more general class of density functions, defined in Definition 4.1. Based on numerical results, the median ability juror should also have the casting vote for β- and Gaussian density functions. We have shown that the result is true in the limit case, and it is quite intuitive to understand. We conjecture that this is true for any density function satisfying 4.1.

Chapter 5

Conclusions and Future Research

We have started off by covering the mechanism design problem on preference voting. The problem in many elections is that voters are incentivised to vote strategically, instead of voting their true preference. Arrow's theorem and the Gibbard-Satterthwaite theorem, tells us that it is impossible the find such mechanisms. The only voting mechanisms that are Pareto efficient, independent of irrelevant alternatives and incentive compatible are dictatorships.

We have introduced the three juror casting vote model of [1], and covered the paper's result that for signals with linear densities, the median ability juror should have the casting vote. We wanted to find out, if the results can be generalised for other probability density functions. we hypothesised that this is true for densities satisfying 4.1, and tested it on β- and Gaussian density functions. For both densities, numerical results suggest that the median juror should have the casting vote. We have shown that this is true in the limit case, when non-median jurors are very bad or very good. We conjecture that the median ability juror should be the casting juror for any density, satisfying Definition 4.1.

Based on the model in [1], we have created an advisor-decider problem. We want to know what the optimal role assignment is, when jurors have different ability levels. We have proved that the best juror should be the decider for random signals with linear density functions. Numerical computations suggest that is also true for Gaussian- and β-density functions. We conjecture that this is true for any density function that satisfy Definition 3.1.

The conjectures are still an open problem for future research. Another thing we could look at is larger juror sizes, but results in [1], suggest that it is probably difficult to find general results. They looked at larger juror sizes, where the abilities are evenly spaced on the ability interval e.g., for five jurors the abilities are: $(a_1, a_2, a_3, a_4, a_5) = (0.1, 0.3, 0.5, 0.7, 0.9)$. For five jurors, the second worst juror should have the casting vote, and for higher number of jurors, numerical computations suggest that the worst should have the casting vote. Intuitively, this makes sense, since for higher number of jurors, there is a higher probability that the majority decision is already reached, before the casting juror can vote. It would be interesting to see if this is also true for other ability distributions.

Alpern and Chen also looked into three juror roll call voting in [2], with discrete signals and abilities. It would be interesting to do an analysis on that with continuous signals and abilities and other probability density functions.

Appendix A

Proof $d_3(a,b,c) \geq 0$

Let

$$\begin{aligned} d_3(a,b,c) =& -4a^2 - 4ab - 4b^2 + 8ac + 8bc + a^2bc + ab^2c - 4c^2 - 2abc^2 \\ P_1(a,b,c) =& abc - 4a - 2b - 2c \end{aligned}$$

We want to find the solution of:

$$\begin{aligned} \min_{a,b,c} \; & d_3(a,b,c) \\ \text{s.t.} \; & P_1(a,b,c) \leq 0 \\ & 0 \leq a \leq b \leq c \leq 1, \end{aligned}$$

and show that the global optimal value for $d_3(a,b,c)$ is at least zero.

A local minima needs to satisfy the Karush–Kuhn–Tucker (KKT) conditions:

$$f_1(a,b,c,\lambda_1,\lambda_2,\lambda_3) = 2abc - 8a + b^2c - 2bc^2 + \lambda_1(bc-4) - 4b + 8c - \lambda_2 + \lambda_3 = 0 \quad \text{(A-1)}$$

$$f_2(a,b,c,\lambda_1,\lambda_3,\lambda_4) = a^2c + 2abc - 2ac^2 + \lambda_1(ac-2) - 4a - 8b + 8c - \lambda_3 + \lambda_4 = 0 \quad \text{(A-2)}$$

$$f_3(a,b,c,\lambda_1,\lambda_4,\lambda_5) = a^2b + ab^2 - 4abc + \lambda_1(ab+2) + 8a + 8b - 8c - \lambda_4 + \lambda_5 = 0 \quad \text{(A-3)}$$

$$P_1(a,b,c)\lambda_1 = (abc - 4a - 2b + 2c)\lambda_1 = 0 \quad \text{(A-4)}$$

$$a\lambda_2 = 0 \quad \text{(A-5)}$$

$$(b-a)\lambda_3 = 0 \quad \text{(A-6)}$$

$$(c-b)\lambda_4 = 0 \quad \text{(A-7)}$$

$$(1-c)\lambda_5 = 0 \quad \text{(A-8)}$$

$$\lambda_1, \ldots, \lambda_5 \geq 0 \quad \text{(A-9)}$$

$$0 \leq a \leq b \leq c \leq 1 \quad \text{(A-10)}$$

$$P_1(a,b,c) = abc - 4a - 2b - 2c \leq 0. \quad \text{(A-11)}$$

For more information on non-linear optimisation, see [11]. We are going to cover the following cases:

1. $b = c$ and $a = 0$
2. $b = c$ and $0 < a = b$
3. $b = c$ and $0 < a < b$
4. $b < c$ and $a = b$
5. $b < c$ and $a < b$

Case: $b = c$ and $a = 0$. All KKT conditions are satisfied for $0 \leq b \leq 1$, $\lambda_1 = \lambda_3 = \lambda_4 = \lambda_5 = 0$ and $\lambda_2 = 4b - b^3 \geq 0$.

Case: $b = c$ and $0 < a = b$. First of all, $a > 0$, implies $\lambda_2 = 0$. For $0 < a = b = c$, we have that $P_1(a,b,c) = a^3 - 4a < 0$, so $\lambda_1 = 0$. The KKT conditions are then satisfied, if:

$$\begin{aligned} 0 <& a = b = c \leq 1 \\ \lambda_3 =& a\left(4 - a^2\right) \\ \lambda_4 =& 2a\left(4 - a^2\right) \\ \lambda_5 =& 0. \end{aligned}$$

Case: $b = c$ and $0 < a < b$. First of all, $a > 0$, implies $\lambda_2 = 0$, and $a < b$, implies $\lambda_3 = 0$. We have that $P_1(a,b,c) = -4a + ab^2 < 0$, so $\lambda_1 = 0$. Substituting λ_1, λ_2, and λ_3 into (A-1), we get $2ab^2 - 8a - b^3 + 4b = \left(b^2 - 4\right)(2a - b) = 0$, and the only feasible solution is:

$$a = b/2. \tag{A-12}$$

Substituting λ_1 and λ_3 into (A-2), we get:

$$\lambda_4 = 4a - a^2 b. \tag{A-13}$$

Substituting (A-12) and (A-12), into (A-3), we get:

$$\lambda_5 = b^3 - 2b < 0,$$

which contradicts (A-9). A KKT point cannot satisfy $0 < a < b = c$.

Case: $b < c$ and $a = b$. Since $b < c$, we have that $\lambda_4 = 0$. If $a = 0$, then $P_1(0,0,c) = 2c > b = 0$. This excludes the possibility that $a = 0$, which implies $\lambda_2 = 0$. Substituting $a = b$ and $\lambda_2 = \lambda_4 = 0$ into (A-1) and (A-2), we get:

$$\begin{aligned} f_1(a,a,c,\lambda_1,0,\lambda_3) =& 3a^2c - 2ac^2 - 12a + 8c + \lambda_1(ac - 4) + \lambda_3 = 0 \\ f_2(a,a,c,\lambda_1,\lambda_3,0) =& 3a^2c - 2ac^2 - 12a + 8c + \lambda_1(ac - 2) - \lambda_3 = 0. \end{aligned}$$

The above equations imply $\lambda_1 = \lambda_3$.

In case $P_1(a,a,c) = 0$, then $c = \frac{6a}{a^2+2}$. It is straightforward to check that the following point is a KKT point:

$$\begin{aligned}(a,b,c) =& \left(a, a, \frac{6a}{a^2+2}\right)\\ \lambda_1 =& \lambda_3 = \frac{8a-2a^3}{a^2+2}\\ \lambda_2 =& \lambda_4 = \lambda_5 = 0,\end{aligned}$$

where $0 < a < 3-\sqrt{7}$.

In case $P_1(a,a,c) < 0$, then $c < \frac{6a}{a^2+2}$ and $\lambda_1 = 0$. Substituting $a = b$, $\lambda_1 = \lambda_2 = \lambda_3 = 0$ into (A-1), we get:

$$f_1(a,a,c,0,0,0) = -12a + 8c + 3a^2c - 2ac^2 = 0.$$

The solutions for c are $c = 4/a$ and $c = 3a/2$, where only $c = 3a/2$ is feasible. Substituting $a = b$, $c = 3a/2$, $\lambda_1 = \lambda_4 = 0$, into (A-3), we get:

$$f_3(a,a,3a/2,0,0,\lambda_5) = -4a^3 + 4a + \lambda_5 > 0.$$

A KKT point cannot satisfy $P_1(a,a,c) < 0$.

Case: $b < c$ **and** $a < b$**.** Since $b < c$ and $a < b$, we have that $\lambda_3 = \lambda_4 = 0$. We cannot have $a = 0$, since $P_1(0,b,c) = 2c - 2b > 0$, so this implies $\lambda_2 = 0$. Substituting $\lambda_2 = \lambda_3 = \lambda_4 = 0$ into (A-1) and (A-2), we get:

$$\begin{aligned}f_1(a,b,c,\lambda_1,0,0) =& 2abc - 8a + b^2c - 2bc^2 + \lambda_1(bc-4) - 4b + 8c = 0\\ f_2(a,b,c,\lambda_1,0,0) =& a^2c + 2abc - 2ac^2 + \lambda_1(ac-2) - 4a - 8b + 8c = 0\end{aligned}$$

It is required that $f_1(a,b,c,\lambda_1,0,0) = f_2(a,b,c,\lambda_1,0,0) = 0$, and that is only if:

$$\left(4-2c^2\right)(b-a) + c\left(b^2-a^2\right) + \lambda_1\left(c(b-a)-2\right) = 0 \tag{A-14}$$

For the case $P_1(a,b,c) < 0$, we have $\lambda_1 = 0$. Equation (A-14) is then:

$$\left(4-2c^2\right)(b-a) + c\left(b^2-a^2\right) > 0,$$

since $b > a$, so there is no KKT point for $P_1(a,b,c) < 0$.

For the case $P_1(a,b,c) = 0$, we have that:

$$c_{P_1=0}(a,b) = \frac{4a+2b}{2+ab}. \tag{A-15}$$

Equation (A-14) implies:

$$\lambda_1(a,b,c) = -\frac{c\left(b^2-a^2\right) + \left(4-2c^2\right)(b-a)}{c(b-a)-2} \tag{A-16}$$

Substituting (A-16) into (A-1) and (A-2), we get

$$f_1(a,b,c,\lambda_1,0,0) = f_2(a,b,c,\lambda_1,0,0) = \frac{(bc-4)\left(\left(a^2-4\right)c + b(6-ac)\right)}{ac-bc+2}$$

We have that $f_1(a,b,c,\lambda_1,0,0) = f_2(a,b,c,\lambda_1,0,0) = 0$, if:

$$c_{f_1=0}(a,b) = \frac{6b}{-a^2+ab+4}. \tag{A-17}$$

We have $c_{P_1=0} = c_{f_1=0}$, when:

$$b = \hat{b}(a) = \frac{a^2+\sqrt{-15a^4+60a^2+4}-2}{4a}. \tag{A-18}$$

Substituting (A-18), into (A-17), we get:

$$\hat{c}(a) \equiv c_{f_1=0}(a,b(a)) = \frac{6\left(a^2+\sqrt{-15a^4+60a^2+4}-2\right)}{a\left(-3a^2+\sqrt{-15a^4+60a^2+4}+14\right)} \tag{A-19}$$

Since $\hat{c}(a) \leq 1$, we need to have that:

$$0 < a \leq \frac{\sqrt[3]{2\left(\sqrt{3666}-54\right)}}{3^{2/3}} - \frac{5\ 2^{2/3}}{\sqrt[3]{3\left(\sqrt{3666}-54\right)}} + 2. \tag{A-20}$$

Substituting (A-18) and (A-19) into (A-16), we get:

$$\hat{\lambda}_1(a) \equiv \lambda_1\left(a,\hat{b}(a),\hat{c}(a)\right). \tag{A-21}$$

Substituting (A-21), (A-18) and (A-19) into (A-3), we get:

$$\hat{f}_3(a) \equiv f_3\left(a,\hat{b}(a),\hat{c}(a),\hat{\lambda}_1(a),0,\lambda_5\right).$$

It can be shown that $\hat{f}_3(a) > 0$, for $\lambda_5 \geq 0$ and a satisfying (A-20). There is no KKT point for $P_1(a,b,c) = 0$, and there is no KKT point for the overall case $b < c$ and $a < b$.

Conclusion All local minimum solutions (a^*, b^*, c^*) are in the set:

$$S_{\min} = \{(0,b,b) : 0 \leq b \leq 1\} \cup \{(a,a,a) : 0 < a \leq 1\} \cup \left\{\left(a,a,6a/\left(a^2+2\right)\right) : 0 \leq a \leq 3-\sqrt{7}\right\}.$$

For every $(a^*, b^*, c^*) \in S_{\min}$, we have that $d_3(a^*, b^*, c^*) = 0$.

Bibliography

[1] S. Alpern and B. Chen, "Who should cast the casting vote? Using sequential voting to amalgamate information," *Theory and Decision*, vol. 83, no. 284, pp. 1–24, 2017.

[2] S. Alpern and B. Chen, "The importance of voting order for jury decisions by sequential majority voting," *European Journal of Operational Research*, vol. 258, no. 3, pp. 1072–1081, 2017.

[3] E. Dekel and M. Piccione, "Sequential voting procedures in symmetric binary elections," *Journal of political Economy*, vol. 108, no. 1, pp. 34–55, 2000.

[4] M. Ottaviani and P. Sørensen, "Information aggregation in debate: who should speak first?," *Journal of Public Economics*, vol. 81, no. 3, pp. 393–421, 2001.

[5] N. Nisan, *Introduction to Mechanism Design (for Computer Scientists)*, pp. 209–242. Cambridge University Press, 2007.

[6] T. Nipkow, "Social choice theory in HOL," *Journal of Automated Reasoning*, vol. 43, no. 3, pp. 289–304, 2009.

[7] K. J. Arrow, "A difficulty in the concept of social welfare," *The Journal of Political Economy*, vol. 58, no. 4, pp. 328–346, 1950.

[8] A. Gibbard, "Manipulation of voting schemes: a general result," *Econometrica: journal of the Econometric Society*, pp. 587–601, 1973.

[9] M. A. Satterthwaite, "Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions," *Journal of economic theory*, vol. 10, no. 2, pp. 187–217, 1975.

[10] J. O. Berger, *Statistical decision theory and Bayesian analysis.* Springer Science & Business Media, 2013.

[11] U. Faigle, W. Kern, and G. Still, *Algorithmic principles of mathematical programming*, vol. 24. Springer Science & Business Media, 2002.