Information-sharing and aggregation models for interacting minds

Piotr Migdal\textsuperscript{a,b}, Michał Denkiewicz\textsuperscript{c}, Joanna Rączaszek-Leonardi\textsuperscript{d}, Dariusz Plewczynski\textsuperscript{e}

\textsuperscript{a}Institute of Theoretical Physics, University of Warsaw, Warsaw, Poland
\textsuperscript{b}ICFO–Institut de Ciències Fotòniques, 08860 Castelldefels (Barcelona), Spain
\textsuperscript{c}Department of Psychology, University of Warsaw, Warsaw, Poland
\textsuperscript{d}Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland
\textsuperscript{e}Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Pawłańska 5a, 02-106 Warsaw, Poland

Abstract

We study mathematical models of the collaborative solving of a two-choice discrimination task. We estimate the difference between the shared performance for a group of $n$ observers over a single person performance. Our paper is a theoretical extension of the recent work of Bahrami et al. (2010) from a dyad (a pair) to a group of $n$ interacting minds. We analyze several models of communication, decision-making and hierarchical information-aggregation.

The maximal slope of psychometric function (closely related to the percentage of right answers vs. easiness of the task) is a convenient parameter characterizing the decisive performance. For every model we investigated, the group performance turns out to be a product of two numbers: a scaling factor depending of the group size and an average performance. The scaling factor is a power function of the group size (with the exponent ranging from 0 to 1), whereas the average also varies: it is arithmetic mean, quadratic mean, or maximum of the individual slopes. We conclude that voting can be almost as efficient as more elaborate communication models, given the participants have similar individual performances.

Keywords: group decision making, two-alternative forced choice, decision aggregation,
group information processing, shared cognition, discriminative judgments, accuracy, discrimination difficulty, bias, information sharing, group size, two-choice decision, distributive cognitive systems, communication models, cognitive process modeling

1. Introduction

Everyone who ever took part in a group decision making or problem solving, probably asked oneself whether it actually made any sense — wouldn’t it be better if simply the most competent person made the choice? In different words, the question is whether a group can outperform its most capable member. There have been many studies that reported group decisions to be less accurate [Corfman and Kahn (1995)]; some, however, concluded that groups using even simple majority voting can make better decisions, than their members alone [Grofman (1978); Kerr and Tindale (2004); Hastie and Kameda (2005)]. We ask a more general question — how does the group performance depend on the individual performances of its participants, and the ways in which they communicate?

This question is put in a new light by recent trends in cognitive psychology, which, after half-century long fascination with isolated cognition in an individual, admit its constant interaction within social environment. It is increasingly realised that joint action and cognition is not limited to the situations of committee/voters’ decisions but pervade our everyday life, requiring constant coordination and integration of cognitive and physical abilities. This trend, called distributed cognition [Hutchins and Lintern (1995), or extended cognition [Clark (2006)], brings the focus of research to the mechanisms of cognitive and physical coordination [Kirsh (2006)] that effectuate this integration, as well as questions about the comparison of the performance of a group to the individual performance. For some tasks that require different types of knowledge and abilities from group participants, the groups are likely to outperform the individuals [Hill (1982)]. For others, such as simple discrimination tasks or estimates, a question arises if indeed, and when, a group may be better than the best of its members.

Group decision making obviously involves members interacting with each other. Casting a vote requires minimum amount of communication — one only needs to inform about his or hers choice. However other group decision situations allow extensive communication and negotiations of a decision. The question is: which forms of communication are most likely to bring an improved outcome, and what actually is being communicated in those groups.
Recent experiments by [Bahrami et al. (2010)] have shown that cooperation can be beneficial, even in case of an extremely simple task, and that this benefit is best explained by the participants communicating their relative confidences. In their study dyads (pairs) performed a perceptual two-choice discrimination task — on every trial the participants were to decide, which of two consecutive stimuli (Gabor patches) had higher contrast. First, decisions were collected from both persons; then the participants were allowed to communicate in order to reach a joint decision. The decision data obtained from a person was used to fit a psychometric function — i.e. probability of that person giving a certain answer, as a function of the contrast value. This function describes the persons skill in the considered task. Similarly, a function describing the skill of the group as a whole can be estimated from the group decisions.

Various assumptions about the nature of the within-group interactions during the decision-making process can be made. From these assumptions we can derive theoretical relationships between the parameters of members' functions and the parameters of the group function — these are the models of decision making. The correctness of each model can then be tested against empirical data.

[Bahrami et al. (2010)] described and evaluated four such models. One was his own, in which group members communicate their relative confidence in their individual choices. Another stems from a signal detection theory approach [Sorkin et al. (2001)] — if the members know each other’s psychometric functions, the group can make a statistically optimal choice. Thus, under certain conditions, we have an upper bound on group performance. The third model stated, that the dyad is as good as its best member. Finally, the last model was random response selection. The study concluded that, when similarly skilled persons meet, they can both benefit from cooperation. The model in which participants communicate their relative confidences best explains this benefit.

We extend the models from [Bahrami et al. (2010)] study to groups of \( n \) participants and compare their predictions. Furthermore, we add a model in which a participant either knows the correct answer, or guesses. In the case of larger groups it may be so, that only small subgroups of participants can communicate simultaneously. We address this issue by considering hierarchical schemes of decision aggregation, in which decisions are first made by subgroups, then some of these subgroups interact with each other and reach a shared decision and so on.

The paper is organized as follows. In Section 2 we introduce the standard model of the discrimination of two stimuli. We use it to assess performances
of individuals and groups. Section 3 contains a series of models of communica-
tion, which express performance of a group of \( n \) persons as a function of their individual performances. In Section 4 we investigate how each model works, assuming several schemes of decision aggregation. Section 5 concludes introduced models and gives insight into further experimental and theoretical work.

2. Model of discrimination

Consider an experiment in which a participant has to make simple dis-
criminatory decisions with varying difficulty. Each trial is assigned a param-
eter \( c \) that describes physical distance between the stimuli to be discriminated (e.g. in the Bahrami et al. experiment it was the difference of the contrast between Gabor patches). Negative \( c \) describes a situation, when the right choice is the first one, whereas positive — the second one. The absolute value of \( c \) reflects the difficulty of a given trial — the lower it is, the more difficult the trial.

Knowing the choices of a certain decision-making agent (in our case either a single participant or a group making the decision together) for a range of contrasts one can construct a mathematical description of its performance on the task. For such an agent we can determine a psychometric function — probability of choosing the second answer as a function of the displayed contrast, \( P(c) \). An ideal responder would be described by the Heaviside step function: \( P(c) = 0 \) for all negative contrasts and \( P(c) = 1 \) for all positive contrasts. Since responders make errors, the actual functions are different. In particular a cumulative of the normal distribution:

\[
P(c) = H \left( \frac{c+b}{\sigma} \right),
\]

where

\[
H(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left( -\frac{t^2}{2} \right) dt,
\]

turns out to be a good fit for the experimental data [Bahrami et al. (2010)]. The parameter \( \sigma \) is the width parameter — it can be seen an expression of the participant’s uncertainty about the decision. The parameter \( b \) is the bias (offset) — it represents the tendency to choose a particular answer. The \( P(c) \) function defined as above can be seen as a convolution of the step function (the correct answer) and the Gaussian distribution (the discriminative error) — see Fig. 1.

One of possibilities of describing such response is a signal detection theory [Sorkin et al. (2001)]. According to it, for a stimuli with contrast \( c \) participant
Figure 1: Plot of the psychometric function, with shown slope $s$, a positive bias $b$. The shaded area $W$ is proportional to the error rate of a participant.

perceive contrast $x$, which is a normally distributed random variable centred around $-b$ and with the variance $\sigma$. Two models described in this paper (i.e. Weighted Confidence Sharing and Direct Signal Sharing) use this assumption explicitly.

For our purposes we assume that bias is much smaller than characteristic width parameter, i.e. $|b| \ll \sigma$. This assumption seems to be well justified for this and similar psychological experiments. In the case of other psychological experiments, the bias may be even the main (or the only one) parameter — e.g. the model of averaging biases in a situation when a group guesses a demographical quantity Rauhut and Lorenz (2010). Consequently, $\sigma$ becomes the main determinant of the effectiveness of discrimination. It is convenient to choose the maximal slope of the psychometric function

$$s = \frac{1}{\sqrt{2\pi}\sigma},$$

as the primary measure of responder’s effectiveness.

Now we can proceed to extending [Bahrami et al. 2010] models. We would like to know how performance of a group of $n$ people depends on their individual cognitive performances. Therefore, we need to solve the explicit formulas for the propagation of slopes and biases, when combining several responders within each of the different models of communication:

$$s_{\text{model}}(s_1, b_1, \ldots, s_n, b_n),$$

$$b_{\text{model}}(s_1, b_1, \ldots, s_n, b_n).$$

Each model is described by the shared decision function

$$P_{\text{model}}(c) = f [P_1(c_1), \ldots, P_n(c_n)],$$
where \( f \) is a functional. For all but two models \( P_{\text{model}}(c) = f [P_1(c), \ldots, P_n(c)] \), that is, the dependence is pointwise (i.e. result for a given \( c \) requires only knowing individual \( P_i(c) \) for the same \( c \)).

We can obtain the effective slope \( s \) and bias \( b \) using the straightforward formulas, which involve taking derivative of the psychometric function with respect to the contrast:

\[
s_{\text{model}} = \left. P'_{\text{model}}(c) \right|_{c=b_{\text{model}}} \approx \left. P'_{\text{model}}(c) \right|_{c=0}
\]

\[
b_{\text{model}} = \left[ b \text{ for which } P_{\text{model}}(-b) = \frac{1}{2} \right] \approx \frac{P_{\text{model}}(0) - \frac{1}{2}}{P'_{\text{model}}(c)|_{c=0}},
\]

where approximations are calculated for relatively small biases, i.e. the relative error for both \( s \) and \( b \) is of order \( O(s^2b^2) \) (or equivalently, \( O(b^2\sigma^2) \)). The derivation is in Appendix A. Note that if \( P_{\text{model}}(c) \) is cumulative of Gaussian function (as in (1)) the formulas for slope (3) and (7) are equivalent. The latter, however, is valid in the general case of an arbitrary communication strategy \( P_{\text{model}}(c) \).

A question arises about the relation between the psychometric curve parameters and the expected rate of the errors. To assess the average amount of wrong answers one can expect from a responder, we introduce the following quantity

\[
W(\sigma,b) = \int_{-\infty}^{0} P(c) dc + \int_{0}^{\infty} [1 - P(c)] dc
\]

\[
= \sqrt{\frac{2}{\pi} \sigma} \exp \left( -\frac{b^2}{2\sigma^2} \right) + b \left[ 2H \left( \frac{b}{\sigma} \right) - 1 \right],
\]

where we integrated the error function [Abramowitz and Stegun (1965)]. For uniform distribution of stimuli, and range of stimuli \((-r, r)\) for \( r \gg (\sigma + |b|) \), the rate of the wrong responses is given by \( W(\sigma, b)/(2r) \). The average number of wrong answers is always reduced, when lowering either width or bias, regardless of the other parameter’s value. This fact further justifies the choice of the slope as a proper measure of the effectiveness. When there is no bias, (10) simplifies, i.e. \( W(\sigma, 0) = 2/s \) thus the rate of the wrong responses is \( 1/(rs) \).

3. Information-sharing models

In this section we discuss different models of information-sharing of \( n \) participants. It is important to underline that the models incorporate the
process of perceiving (what the subjects may know), the state of mind (what the subjects know), and the communication and the decision-making process (usually Bayes-optimal). We briefly define assumptions of each model and justify it in psychological terms. We give results in terms of the effective psychometric function $P_{model}(c)$, the effective slope $s_{model}$ and sometimes the effective bias $b_{model}$ (as for a few models the bias is poorly-defined). Whenever calculations of $P_{model}(c)$ are not straightforward, we give some insight into the underlying mathematics.

We investigate the following models:

- 3.1 Random Responder,
- 3.2 Voting,
- 3.3 Best Decides,
- 3.4 Weighted Confidence Sharing,
- 3.5 Direct Signal Sharing,
- 3.6 Truth Wins.

3.1. Random Responder

Model. Each trial decision of a random group member is taken as the group decision.

Motivation. It serves as one of the reference models and it is not expected to be fulfilled in most of realistic settings. Random factors determine the collective decision, i.e. communication is seen as ineffective within framework of this model. Sometimes decision is not based on any support and people may have very misleading impression of their own accuracy. Also, their decision may be depend more on one’s charisma, or persuasive skills that the merits. In the work of Bahrami et al. this model is called ‘Coin flip’.

Results.

$$P_{RR}(c) = \frac{1}{n} \sum_{i=1}^{n} P_i(c)$$

(11)

After the differentiation one obtains the slope (7) and the bias (8):

$$s_{RR} \approx \frac{s_1 + \ldots + s_n}{n}$$

(12)

$$b_{RR} \approx \frac{s_1 b_1 + \ldots + s_n b_n}{s_1 + \ldots + s_n}$$

(13)
The relative error both for $s_{RR}$ and $b_{RR}$ is $O(s_1^2 b_1^2) + \ldots + O(s_n^2 b_n^2)$. Note that $P_{RR}(c)$ is not normal (1).

3.2. Voting

Model. Each participant makes her or his own decision. The majority voting makes the decision of the group. In the case of the equal number of opposite opinions a coin is flipped.

Motivation. People may have no access to their accuracy (or they cannot communicate it reliably), thus a good strategy is to take voting as the final consensus result.

Results.

\[ P_{Vot}(c) = \sum_{k=1}^{\lfloor n/2 \rfloor} \sum_{i} [1 - P_{i_1}(c)] \ldots [1 - P_{i_k}(c)] P_{i_{k+1}}(c) \ldots P_{i_n}(c) \]

\[ + \left[ \frac{1}{2} \sum_{i} [1 - P_{i_1}(c)] \ldots [1 - P_{i_{n/2}}(c)] P_{i_{n/2+1}}(c) \ldots P_{i_n}(c) \right] \text{ if } n \text{ is even} \]

where sum over $i$ denotes sum over every permutation of participants. We obtain (derivation in Appendix B)

\[ s_{Vot} \approx \frac{s_1 + \ldots + s_n}{n} \times \left\{ \begin{array}{ll} \frac{n}{2^{(n/2)}} & \text{if } n \text{ is even} \\ \frac{n}{2^{(n-1)/2}} & \text{if } n \text{ is odd} \end{array} \right. \]

\[ b_{Vot} \approx \frac{s_1 b_1 + \ldots + s_n b_n}{s_1 + \ldots + s_n} \]

The $P_{Vot}(c)$ is not normal (1). The relative error both for $s_{Vot}$ and $b_{Vot}$ is $O(s_1 b_1) + \ldots + O(s_n b_n)$. Note that the addition of an odd member to a group does not increase its average performance. The formula (16) is an asymptotic expression for large $n$, which makes use of the Wallis formula. For $n = 2$ Random Responder and Voting models give the same results.

3.3. Best Decides

Model. The most accurate member of the group makes decision. This model is called Behavior and Feedback in Bahrami et al. (2010). In this model we will focus on the case with no bias $b = 0$. Nonzero bias would make the result hard to state in explicit form — see (10).
Motivation. In some experimental settings members of the group can determine, who is the best of them (e.g. when the feedback is present). Then they can let him/her make the final decision. Studies by [Henry 1995] suggest that, at least in some types of tasks, participants can identify the most proficient member, so our assumption is plausible. As in the previous models, there is no (effective) communication between the members of the group.

Results.

\[
P_{BD}(c) = P_{\text{member with the highest } s(c)}
\]

\[
s_{BD} = \max(s_1, \ldots, s_n)
\]

In the case when biases are large, the group psychometric function is that of the most effective participant (i.e. one with the lowest \(W(\sigma_i, b_i)\)), \(P_{BD}(c) = P_i(c)\). The strategy is the most beneficial for a group with very diverse individual performances.

3.4. Weighted Confidence Sharing

Model. Group members share their relative confidences \(z_i = x_i/\sigma_i\). The group decision depends on the sign of \(\sum_{i=1}^n z_i\), i.e. for the negative they choose the first option and when positive — the second. This model requires each \(P_i(c)\) to be normal \([1]\).

Motivation. The value \(x_i\) is the contrast perceived by \(i\)-th participant and has the distribution with the density \(P'_i(c)\), as it is in [Sorkin et al. 2001]. The true contrast \(c\) is, of course, common for all participants in a given trial. The relative confidence is equivalent to a \(z\)-score, if the participant is unbiased (i.e. its related to probability that the participant is right). Put differently, participants know their \(z\)-scores on a given trial, but are unaware of their own parameters \(s\) and \(b\). This model was first introduced by [Bahrami et al. 2010]. It is possible that in an experimental trial each participant can estimate and effectively communicate their relative confidence, by saying phrases being a coarse real-world approximation of one’s \(z\)-score, e.g. ‘I lean towards 1st’, or ‘I am almost sure it is the 2nd’). Bahrami’s et al. study suggests that this model most accurately describes dyad performance.

Given relative confidences \(\vec{z} = (z_1, \ldots, z_n)\), the group has to determine, whether to choose the first or the second option. If there are only two participants, in the case of different opinions, the one with the stronger confidence (in this trial) decides. This can be written as follows: the group chooses the first option if \(z_1 + z_2 \leq 0\), the second option otherwise, yielding the optimal
strategy [Bahrami et al. (2010)]. In the general case of $n$ participants we use the Bayes optimal reasoning. We calculate the probability that the contrast is positive (and thus the second answer is correct) given $z$-scores provided by each participant:

$$p(c > 0 | \vec{z}) = \int_{c=0}^{\infty} p(c | \vec{z})dc = \frac{\int_{0}^{\infty} p(\vec{z} | c)p(c)dc}{\int_{-\infty}^{\infty} p(\vec{z} | c)p(c)dc},$$

(20)

where $p(c)$ is probability of a discrimination task with $c$. Probability of observing $z_i$-score given contrast $c$ is $P_i'(c - \sigma_i z_i)$, thus

$$p(\vec{z} | c) = P_1'(c - x_1) \cdot \ldots \cdot P_n'(c - x_n).$$

(21)

Let’s assume that displayed contrast has uniform distribution, i.e. $p(c)$ is constant (not going into mathematical nuances). In order to define decision function we need to know when $p(c > 0 | \vec{z}) \geq 1/2$ or, in other words, when the probability that the second answer is correct is greater than 1/2. As (20) is a Gaussian function of $c$, finding its maximum leads to the condition

$$\frac{x_1}{\sigma_1^2} + \ldots + \frac{x_n}{\sigma_n^2} \geq 0,$$

(22)

or equivalently, using the slope parameter,

$$s_1 z_1 + \ldots + s_n z_n \geq 0.$$

(23)

Thus when the condition holds, choosing the second option is the Bayes optimal choice. Unfortunately, in the considered model we only have access to values of $\vec{z}$, not individual performances. In order to get the precise answer of the optimal choice we need to know the whole distribution of $\sigma_i$ (or $s_i$). Instead, we can use the approximate condition for the choice of the second option

$$z_1 + \ldots + z_n \geq 0,$$

(24)

to obtain a lower bound of the performance. The condition is exact for participants with equal performances (and should be close to the optimal if the values of $\sigma_i$ do not vary much). This equation can be seen as a kind of a weighted voting, where weights depend on subjective confidences, but not on individual performances. Members don’t know theirs, or their peers' parameters, so there is no justification for assigning more or less weight to a particular member for the whole experiment. The only thing that matters is confidence in the present trial.
Results. To calculate $P_{WCS}(c)$ we need to count, given contrast $c$, probability of obtaining set $\vec{z}$ with a positive sum (24), thus

$$P_{WCS}(c) = \int_{x_1/\sigma_1+\ldots+x_n/\sigma_n \geq 0} \exp \left[ -\frac{(c+b_1-x_1)^2}{2\sigma_1^2} + \ldots - \frac{(c+b_n-x_n)^2}{2\sigma_n^2} \right] \frac{dx_1 \cdots dx_n}{(2\pi)^n/\sigma_1 \cdots \sigma_n}$$

$$= H \left[ \sqrt{2\pi}s_{WCS}(c + b_{WCS}) \right], \quad (26)$$

where the integration bases on the fact that a sum of Gaussian random variables $z_i$ is a Gaussian random variable [Piau (2011)]. The resulting parameters are:

$$s_{WCS} = \sqrt{n} \times \frac{s_1 + \ldots + s_n}{n}, \quad (27)$$

$$b_{WCS} = \frac{s_1 b_1 + \ldots + s_n b_n}{s_1 + \ldots + s_n}. \quad (28)$$

Again, note that the above result for $s_{WCS}$ is the low boundary value for the optimal Bayesian reasoning, exact only for $n = 2$ (due to symmetry) and the group of participant with the same performances. Knowing the exact distribution of individual performances one can get a better (or at least the same) group performance. Then instead of the summation of individual $z$-scores (24) one will get a more complicated formula for the decision.

3.5. Direct Signal Sharing

Model. Group members share both their perceived contrasts $x_i$ and their $\sigma_i$. The group decision depends on the sign of $\sum_{i=1}^n x_i/\sigma_i^2$. This model requires each $P_i(c)$ to be normal [1].

Motivation. As for the WCS, we assume that the value $x_i$ is the contrast perceived by $i$-th participant and has the distribution with the density $P_i'(c)$, as it is in [Sorkin et al. (2001)]. The group possesses complete knowledge about the characteristics of its members and their perception, so its effectiveness is hindered only by the skill of the participants, not by communication. This model constitutes the upper bound for group performance, provided that the stimuli are fully defined by their contrast values (and perceived according to the discussed model). In the case of more complex, non-perceptive task it is possible for a group to exceed this bound [Hill (1982)] — for example when participant’s skills complement each other. People know the strength of the stimuli, but also their own sensitivity. If the feedback is provided, one can plot $x$ versus $c$ to get $\sigma$. 

11
Results. The final group decision follows the standard derivation of \( n \) classifiers collecting independent results with normal distribution (eg. Sorkin et al. (2001) and Bahrami et al. (2010)):

\[
P_{DSS}(c) = \frac{1}{\text{normalization}} \int_{-\infty}^{c} P_1'(x) \cdots P_n'(x)dx
\]

\[
s_{DSS} = \sqrt{s_1^2 + \cdots + s_n^2} = \sqrt{n} \times \frac{\sqrt{s_1^2 + \cdots + s_n^2}}{n}
\]

\[
b_{DSS} = \frac{s_1^2 b_1 + \cdots + s_n^2 b_n}{s_1^2 + \cdots + s_n^2}
\]

Note that regardless of the distribution of the individual performances, the group performance outscores both Best Decides and Weighted Confidence Sharing.

3.6. Truth Wins

Model. We assume that on each trial each member is in one of the two states: either knowing the right answer or being aware of his/her ignorance. In the latter case a random guess is made. So it is sufficient to have a single group member to perceive the stimuli correctly in order to get the correct group answer. We assume no bias as there is no possible way to treat it consistently, as it introduces false convictions.

Motivation. For so called eureka-type problems the signal-theoretic limit can be exceeded (Hill 1982). The key is that the answer to such a problem has the property of demonstrability — it allows a single member, who figured out the answer, to easily convince the rest of the group about its correctness (Laughlin et al. 1975). People may know if they see the contrast stimuli (and all errors are only due to guessing, not to false observations). The model has received much attention in the group decision theory, e.g. in Davis (1973). It is appropriate for situations, when the correctness of a solution can be demonstrated. However, we don’t expect this model to be realized tasks similar to that of Bahrami et al. (2010). It serves as a reference and an explicit example of a result beyond one provided by the Direct Signal Sharing model, we included it with the aim of generalization of the models to different decision situations.

Results. The chance the responder knows with certainty the right answer is

\[
R(c) = |2P(c) - 1|
\]
That is, it is a reversed formula saying that when one knows answer with probability $R(c)$ then effectively answers correctly with probability $R(c) + (1 - R(c))/2$ (as there is chance to answer correctly by a random guess). The probability that at least one person knows the correct answer is

$$R_{TW}(c) = 1 - [1 - R_1(c)] \cdot \ldots \cdot [1 - R_n(c)].$$

Consequently,

$$P_{TW}(c) = \text{sign}(c) R_{TW}(c) + 1 \over 2$$

$$s_{TW} = n \times \frac{s_1 + \ldots + s_n}{n},$$

where the slope is a result of the straightforward differentiation \[7\].

The model yields much better result than other models; note however that the absence of false observations is the strong requirement. Other models have to operate without this assumption. Note that the $P_{TW}(c)$ is not normal.

4. Aggregation of information in hierarchical schemes

So far we have assumed that information from all participants is simultaneously collected and used in the group decision. One may argue that this is unrealistic model of human communication for groups of more than a few people. We therefore propose hierarchical models (schemes) in which only small subgroups can communicate at a time. Each of these subgroup reaches its own decision, in a way described by one of the models introduced in the previous section. Hence, it can be regarded as a decision-making agent, described by slope and bias. The subgroup can then communicate with other subgroups or individual members, which results in larger groups being created, until all information is gathered and the final decision is made.

The results of employing a multi-level decision system can significantly deviate from what simultaneous information collection predicts. For instance in a two-level voting system, which has been widely studied in the context of election results (e.g. Davis (1973); Laughlin et al. (1975)), the final outcome depends heavily on the distribution of votes in subgroups, sometimes allowing minority groups to overcome the majority, sometimes exaggerating the power of the majority. It is thus interesting to study the possible effects of such hierarchical systems.

Let’s propose the following model for communication of $n$ participants:
1. At the beginning there are $n$ agents.
2. Each turn only $g$ (for our purpose: 2 or 3) agents (groups or individuals) share their information according to a chosen model. Then they are merged into one agent (defined by $s_{model}(s_1, \ldots, s_g)$).

In other words, a group of people, who shared information, is treated as a single agent in the next turn. There are two free parameters:

- Model used to combine members’ parameters into group parameters.
- Structure in which groups are formed, i.e. the way to determine, which agents should interact in given turn.

Let’s consider following ways of the group forming (see Fig. 2 for the diagram of the two first schemes):

- **4.1 Shallow hierarchy:** Each turn $g$ agents from the groups with the least number of participants interact.
- **4.2 Deep hierarchy:** Each turn $g - 1$ agents join to the group with largest number of participants.
- **4.3 Random hierarchy:** Each turn $g$ random agents interact.

![Shallow Scheme](image1.png) ![Deep Scheme](image2.png)

Figure 2: Diagram of the interaction ordering for aggregation schemes for $g = 2$: Shallow Scheme—each turn two agents from the least numerous groups interact, Deep Scheme—each turn a single participant joins the previously formed group.

For some models the way in which groups are formed is irrelevant for obvious reasons. It is the case for Random Responder, Best Decides, Direct Signal Sharing and Truth Wins. The result is always the same and equivalent to the simplest situation without any hierarchy. Models, which are affected to a certain degree, are: Weighted Confidence Sharing and Voting.

Note that agents in principle do not know their slopes, so the order of interactions cannot depend on their individual (or group) $s_i$. However, as both $s_{WCS}$ and $s_{Vol}$ depend linearly on $s_i$, the averaging over every permutation of participants yield in the result, which is proportional to the
arithmetical mean of $s_i$, or $\langle s \rangle$. Consequently, to investigate the influence of the hierarchical information-aggregation on the result, it suffices to treat each participant as if its performance equals to $\langle s \rangle$.

For our convenience we consider a more general model with the parameter $a_g$ (the amplification multiplier) depending on $g$ (the group size):

$$s_{a_g}(s_1, \ldots, s_g) = a_g \frac{s_1 + \cdots + s_g}{g}.$$  \hfill (36)

It covers both WCS ($a_2 = \sqrt{2}$, $a_3 = \sqrt{3}$, ...) and Voting ($a_3 = 3/2$, ...) models and allows us to give results in the elegant general form.

### 4.1. Shallow hierarchy

The justification of the Shallow hierarchy is the following: people may locally find their partners and then make a collective decision. Then iteratively groups of the same (or similar) size make the collective decision.

The analysis is simple for the number of participants being a power of $g$, i.e. $n = g^k$ where $k$ is a natural number. Then, each a few elementary steps the number of agents is reduced by the factor of $g$, and agents’ slopes are multiplied by the factor $a_g$. In the end we get

$$s_{a_g,\text{Shallow},g} = (a_g)^k \langle s \rangle = n^{\log_g(a_g)} \langle s \rangle.$$ \hfill (37)

In particular for the Weighted Confidence Sharing (i.e. $a_g = \sqrt{g}$) we reach the saturation

$$s_{\text{WCS},\text{Shallow},g} = \sqrt{n} \langle s \rangle$$ \hfill (38)

and thus the aggregation process does not introduce even the slightest decrease in the group performance, comparing with collecting all information at once. The formula (38) holds only for $n$ that is a power of $k$. However, for different $n$s the formula works as a very good approximation — see Fig. 3 for the numerical results. The relation (i.e. that for groups of size $n = g^k$ we reach the efficiency of model without aggregation or $s_{a_g,\text{Shallow},g} = s_{a_g,\text{Shallow}}$) is true for every model described by (36) with $a_g = g^{\alpha}$ for any $\alpha$.

In the Voting model we need to consider the aggregation in the group of at least three (i.e. $g = 3$ and $a_3 = 3/2$) — otherwise it is the Random Responder model. For $n$ being the power of three we get

$$s_{\text{Vot},\text{Shallow},g=3} = n^{\log_3 3/2} \langle s \rangle \approx n^{0.37} \langle s \rangle,$$ \hfill (39)

which works as a good approximation also for the general odd $n$. For every even $n$ there is at least one process with two parties, significantly decreasing the total performance (as voting for two participants reduces to a coin flip).
4.2. Adding one or two at a time

In this case, there is a single group to which single agents join one after another. The resulting slope is for the Weighted Confidence Sharing model

\[ s_{\text{WCS,Deep},g=2} = 2^{-(n-1)/2} \langle s \rangle + \sum_{i=1}^{n-1} 2^{-i/2} \langle s \rangle = \left( 1 + \sqrt{2} - 2^{1-n/2} \right) \langle s \rangle \]  

(40)

and for the Voting model for an odd \( n \) and aggregation of three

\[ s_{\text{Vot,Deep},g=3} = 2^{(n-1)/2} \langle s \rangle + \sum_{i=1}^{(n-1)/2} 2^{-i} \langle s \rangle = \left( 2 - 2^{-(n-1)/2} \right) \langle s \rangle \]  

(41)

We see that the Deep hierarchy is very inefficient — the multiplier of \( \langle s \rangle \) converges to a constant. This leads to a conclusion, that the simultaneous aggregation (Shallow hierarchy) is not only more natural, but also much more efficient.

To obtain the asymptotic value of \( s_{a_g,\text{Deep},g} \) one can consider an equilibrium situation when \( g - 1 \) individuals join the group, which already reached the limit

\[ s_{a_g,\text{Deep},g} = a_g \left( \frac{g - 1}{g} \langle s \rangle + \frac{1}{g} s_{a_g,\text{Deep},g} \right) \]  

(42)

leading to

\[ s_{a_g,\text{Deep},g} = \frac{g - 1}{g/a_g - 1} \langle s \rangle. \]  

(43)

4.3. Random hierarchy

But what happens between the Shallow hierarchy and the Deep hierarchy? If the groups merge at random, is the final \( s \) closer to the most efficient aggregation scheme, or non-scaling as in adding a few at a time? The answer, not surprisingly, lies between.

We parameterize time with \( t \) starting from 0. Each turn \( g \) agents merge into one of the slopes \( [36] \). The current number of agents is described by \( n_t = n_0 - (g-1)t \). We investigate how the density function of slopes \( \rho_t(s) \) evolves with time, which reads

\[ \rho_{t+1}(s) - \rho_t(s) = \\
- g \frac{\rho(s)}{n_t} + \int \frac{\rho(s_1)}{n_t} \cdots \frac{\rho(s_g)}{n_t} \delta \left( s_{\text{model}}(s_1, \ldots, s_g) - s \right) ds_1 \cdots ds_g, \]  

(44)
where $\delta$ is the Dirac delta. The difference in distributions $\rho_{t+1}(s) - \rho_t(s)$ involves two processes. The first expression means that we take $g$ random agents, so they interact and removed from the distribution. The second — for every possible group of $g$ agents (with slopes $s_1, \ldots, s_n$) there is created one with the slope $s_{\text{model}}(s_1, \ldots, s_g)$.

Note that we use integrals, but sum over a finite set will give the same result. The parameter we care the most is the mean slope, that is

$$\langle s \rangle_t = \frac{1}{n_t} \int s \rho_t(s) ds.$$  \hspace{1cm} (45)

Let’s multiply (44) by $s$ and integrate $\int \cdot ds$. For our case (36) it gives a relatively simple results

$$n_{t+1} \langle s \rangle_{t+1} = n_t \langle s \rangle_t - g \langle s \rangle_t + a_g \langle s \rangle_t$$

or

$$\langle s \rangle_t = \frac{n_0 - (g - 1)t + a_g - 1}{n_0 - (g - 1)t} \langle s \rangle_{t-1}.$$  \hspace{1cm} (46)

To obtain the final result we need to calculate $\langle s \rangle_{t_{\text{max}}}$ for such time that the only one agent remains. We consider $t_{\text{max}} = (n_0 - 1)/(g - 1)$ to be an integer (e.g. for $g = 3$ it means that we need to consider an odd number of participants, for $g = 2$ there are no restrictions). Then, remembering that $\langle s \rangle_0 = \langle s \rangle$ and $n_0 = n$ we get

$$s_{a_g, \text{Random}, g} = \prod_{t=1}^{t_{\text{max}}} \left( \frac{n_0 - (g - 1)t + a_g - 1}{n_0 - (g - 1)t} \right) \langle s \rangle$$

$$= \frac{\Gamma \left( \frac{1}{g-1} \right) \Gamma \left( \frac{n_0}{g-1} + \frac{a_g - 1}{g-1} \right)}{\Gamma \left( \frac{n_0}{g-1} \right) \Gamma \left( \frac{a_g}{g-1} \right)} \langle s \rangle$$

$$\approx \frac{\Gamma \left( \frac{1}{g-1} \right)}{\Gamma \left( \frac{n_0}{g-1} \right)(g - 1)^{(a_g - 1)/(g-1)} \times n^{(a_g - 1)(g-1)} \times \langle s \rangle}$$  \hspace{1cm} (47)

where $\Gamma(x)$ is the Euler gamma function, and we applied the Stirling approximation. For $g = 2$ we obtain the neat result

$$s_{a_g, \text{Random}, g=2} \approx \frac{1}{\Gamma(a_2)} n^{a_2 - 1} \langle s \rangle,$$  \hspace{1cm} (50)

in particular for the Weighted Confidence Sharing model ($a_2 = \sqrt{2}$) we get

$$s_{\text{WCS,Random}, g=2} \approx 1.13n^{0.41} \langle s \rangle,$$  \hspace{1cm} (51)
whereas for the Voting model for $g = 3$ (and odd number of participants) we get

$$s_{Vot,Random,g=3} \approx 1.22 n^{0.25} \langle s \rangle.$$ (52)

The numerical results, along with their approximations, are plotted in Fig. 3.

Figure 3: Plot of numerically obtained multipliers of $\langle s \rangle$ for models with aggregation of information. Weighted Confidence Sharing with $g = 2$ for aggregation hierarchies: Shallow (circles), Deep (diamonds) and Random (squares). Voting with $g = 3$, and only for odd number of participants, for aggregation hierarchies: Shallow (circles), Deep (diamonds) and Random (squares). The lines are the respective analytical results from Sec. 4.

5. Conclusion and further remarks

In the paper we examined mathematical models of solving a two-choice discriminative task by a group of participants. We were interested how the group performance depends on the performance of individuals, ways of communication and modes of decision aggregation. As a marker of the performance we used the slope of the psychometric function (3), which says how the performance changes with the difficulty of the task. The higher slope $s$, the better performance of an individual (or a group).

We analyzed a number of models, also modifying them by allowing interaction of only a few people at once. Some of the models can be always consider a strategy for the group decision-making: the Random Responder and the Voting. For the Best Decides one need to assume that the group posses information indicating who performs better (e.g. from the feedback). Other models (i.e. the Weighted Confidence Sharing, the Direct Signal Sharing and the Truth Wins) have direct assumptions on the problem structure.
or information that can be shared. Consequently, only in a subset of two-choice discriminative tasks they can be adopted. The list of the models is by no means exhaustive.

For each investigated model we arrived at the formula for the slope of a group as a function of individual slopes:

\[ s_{\text{model}}(s_1, \ldots, s_n) = \text{multiplier}_{\text{model}}(n) \times \text{mean}_{\text{model}}(s_1, \ldots, s_n), \]  

(53)

where the explicit results are placed in the Tab. 1 and Fig. 4. Note that the formula has two parts as factors — the part related to how the group size affects the performance, and the mean of the individual slopes (if the better-performing contribute more to the outcome). For equally skilled participants only the multiplier matters, whereas for a group of people with the high variance of performances, the type of mean is crucial.

We not only solved the problem for a particular list of models, but we constructed a general framework for the collaborative solving of a two-choice task, i.e. the group performance can be written down as

\[ s_{\text{model}}(s_1, \ldots, s_n) = d \times n^\alpha \times \left( \frac{s_1^p + \ldots + s_n^p}{n} \right)^{1/p}, \]  

(54)

where parameters \( d, \alpha \) and \( p \) can be fitted for any experimental data, even not covered by models we investigated. Note that for \( p = 1 \) we arrive at the arithmetic mean, for \( p = 2 \) — the quadratic mean, and \( p \to \infty \) — the maximum. For the models we investigated (54) is either an exact solution (RR, WCS, BD, DSS, TH) or a good approximation (Voting, information aggregation schemes). If the result is exact, then \( d = 1 \) (to be consistent with the case of \( n = 1 \)).

For a given list of slopes \((s_1, \ldots, s_n)\) it is possible to write relations with the performances (slopes) for different models, which reads

\[ s_{\text{RR}} \leq s_{\text{Vot}} < s_{\text{WCS}} \leq s_{\text{DSS}} \leq s_{\text{TW}}. \]  

(55)

An average-performing participant is expected to benefit from participating in a joint task solving, unless the responder is chosen at random, in which case there is neither gain nor loss. It is somewhat more difficult to relate the Best Decides model, as it highly depends on the distribution of the participants’ skills. We can write

\[ s_{\text{RR}} < s_{\text{BD}} < s_{\text{DSS}} \leq s_{\text{TW}}, \]  

(56)

but how the Best Decides model relate to the Voting and the Weighted Confidence Sharing? The answer lies in the comparison of the most skilled
participant to the average performance, i.e. $\max(s)/\langle s \rangle$. If it is greater that $\approx 0.8\sqrt{n}$, the Best Decides model outperforms the Voting. If the ratio is greater that $\sqrt{n}$ — it outperforms the WCS as well. For example, when there is one expert (with $s_{\text{exp}} > 1$ among $s_{\text{non-exp}} = 1$) among the total number of $n$ participants, then only when $s_{\text{exp}} > \sqrt{n} + 1$ its better for a group to use the Best Decides strategy.

| Model | $s(s_{1},s_{2})$ | $s(s_{1},s_{2},s_{3})$ | Mean | Multiplier |
|-------|------------------|------------------------|------|------------|
| RR    | $\frac{s_{1}+s_{2}}{2}$ | $\frac{s_{1}+s_{2}+s_{3}}{3}$ | arithmetic | 1 |
| Vot   | $\frac{s_{1}+s_{2}}{2}$ | $\frac{s_{1}+s_{2}+s_{3}}{3}$ | arithmetic | $\approx 0.8\sqrt{n}$ |
| BD    | $\max(s_{1},s_{2})$ | $\max(s_{1},s_{2},s_{3})$ | maximum | 1 |
| WCS   | $\frac{s_{1}+s_{2}}{\sqrt{2}}$ | $\frac{s_{1}+s_{2}+s_{3}}{\sqrt{3}}$ | arithmetic | $\sqrt{n}$ |
| DSS   | $\sqrt{s_{1}^{2}+s_{2}^{2}}$ | $\sqrt{s_{1}^{2}+s_{2}^{2}+s_{3}^{2}}$ | quadratic | $\sqrt{n}$ |
| TW    | $s_{1} + s_{2}$ | $s_{1} + s_{2} + s_{3}$ | arithmetic | $n$ |

Table 1: Models summary for the six considered models. For each model there is given explicit formula for two and three members. In each model the $s_{\text{model}}$ has the general form multiplier $\times$ mean.

Figure 4: Plot summarizing multipliers for different models.

For schemes of aggregation (Tab. 2) we obtained two interesting results. First, most of models we investigated are completely not affected by gradual aggregation of information. Second, for models that are affected, the optimal solution is also the one with the least effort — one need to group information in the smallest possible groups, i.e. in $g = 2$ for Weighted Confidence Sharing and $g = 3$ for Voting.
Table 2: Summary of information-aggregation results (see section 4) in groups of g agents for affected models, i.e. Voting and Weighted Confidence Sharing. For each model there are provided asymptotic multipliers for three different information-aggregation hierarchies. In each model the \( s_{model} \) has the form multiplier times arithmetic mean. Note that for Voting grouping in \( g = 4 \) is very ineffective (as, in fact, it effectively uses opinions of three out of four participants). Also note that asymptotically the most effective approach (i.e. the best for very large groups) for the Shallow and Deep aggregation schemes is to gather information in the smallest possible groups of agents (i.e. in \( g = 3 \) for Voting and \( g = 2 \) for WCS).

\[
P_{eff}(c) = \sum_{models} w_{model} P_{model}(c),
\]

\[
s_{eff} = \sum_{models} w_{model} s_{model}.
\]

It is possible that the participants’ strategy varies from trial to trial. In such situations the outcome would be a mixture of strategies (with weights \( w_{model} \)), that is

\[
P_{eff}(c) = \sum_{models} w_{model} P_{model}(c),
\]

\[
s_{eff} = \sum_{models} w_{model} s_{model}.
\]

In order to distinguish between models the sole analysis of performance might be not enough, as (psychologically) different models of problem-solving can yield in the same performance. One can test modified schemes that put additional constraints on participant interaction in order to investigate the communicational aspect directly. For example, contact with other members could be limited to voice or text chat communication, or there may be no feedback provided. In addition, the participants might be asked to express their confidence explicitly on a Likert scale. However, further experimental work should be carried out to clarify if the confidence is subjectively accessible and communicated explicitly, or rather read from participant’s behavior. The amount of feedback could be ranging from full information about the stimulus, through simple information about the correctness, to no feedback at all. As a reference there may serve Social Decision Scheme Theory [Davis (1973)], where the group decision is considered as a function of individual

| Model | \( g \) | Shallow hierarchy | Random hierarchy | Deep hierarchy |
|-------|--------|------------------|-----------------|---------------|
| Vot   | 3      | \( n^{0.37} \)   | 1.22\( n^{0.25} \) | 2.00          |
| Vot   | 4      | \( n^{0.16} \)   | 1.15\( n^{0.08} \) | 1.36          |
| Vot   | 5      | \( n^{0.35} \)   | 1.38\( n^{0.19} \) | 2.15          |
| WCS   | 2      | \( n^{0.3} \)    | 1.13\( n^{1.41} \) | 2.41          |
| WCS   | 3      | \( n^{0.3} \)    | 1.25\( n^{1.37} \) | 2.73          |
| WCS   | 4      | \( n^{0.3} \)    | 1.37\( n^{1.34} \) | 3.00          |
| WCS   | 5      | \( n^{0.5} \)    | 1.48\( n^{0.34} \) | 3.23          |
choices, regardless of their skill, confidence or difficulty of the task.

In all models interaction is beneficial for the performance, except for the Random Responder model (where the performance is the same as the averaged performance of each individual). It may be possible as well [Grofman (1978)] that beyond a certain critical size the groups start performing worse instead of better. Models we consider do not predict such collapse, as they are based on information-sharing and does not incorporate phenomena related to motivation and social or technical ability to work in a group.

One needs to be aware of the fact, that the presented models are valid only for a specific situation of the collaborative solving of a two-choice perceptive task, where the difficulty can be smoothly adjusted. Some other tasks may be analyzed within the same paradigm, like integrating information in one mind, i.e. several exposures to the same stimuli by one person, perhaps with different senses or with different noise levels, a similar experiment is described in [Ernst and Banks (2002)]. Perhaps, collaborative solving of other two-choice task (e.g. verbal or mathematical) can be treated in a similar way. However, for many other settings more advanced models are needed, e.g. the ones taking into account more choices or the dynamical interaction between solving a problem in one’s mind and communication with the other participants. Nevertheless, the authors believe that the first step should be to experimentally verify the predicted results of this paper (with the emphasis of the scaling of the performance), before proceeding to more advanced theoretical models.

Acknowledgements

The work was supported by EC EuroUnderstanding grant DRUST and: Spanish MINCIN project FIS2008-00784 (TOQATA) and ICFO PhD scholarship to PM, Marie Curie Reintegration Grant PERG06-GA-2009-256587 to JRL, and the Polish Ministry of Education and Science (grants: N301 159735, N518 409238) to DP.

References

Abramowitz, M., Stegun, I. A., 1965. Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. Dover Publications

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., Frith, C. D., Aug. 2010. Optimally Interacting Minds. Science 329 (5995), 1081–1085. doi:10.1126/science.1185718
Clark, A., Aug. 2006. Language, embodiment, and the cognitive niche. Trends in cognitive sciences 10 (8), 370–4. doi:10.1016/j.tics.2006.06.012

Corfman, K. P., Kahn, B. E., Jan. 1995. The influence of member heterogeneity on dyad judgment: Are two heads better than one? Marketing Letters 6 (1), 23–32. doi:10.1007/bf00994037

Davis, J. H., 1973. Group decision and social interaction: A theory of social decision schemes. Psychological Review 80 (2), 97–125. doi:10.1037/h0033951

Ernst, M. O., Banks, M. S., Jan. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415 (6870), 429–33. doi:10.1038/415429a

Grofman, B., 1978. Judgmental competence of individuals and groups in a dichotomous choice situation: Is a majority of heads better than one? doi:10.1080/0022250x.1978.9989880

Hastie, R., Kameda, T., Apr. 2005. The robust beauty of majority rules in group decisions. Psychological review 112 (2), 494–508. doi:10.1037/0033-295x.112.2.494

Henry, R., May 1995. Improving Group Judgment Accuracy: Information Sharing and Determining the Best Member. Organizational Behavior and Human Decision Processes 62 (2), 190–197. doi:10.1006/obhd.1995.1042

Hill, G. W., 1982. Group versus individual performance: Are N+1 heads better than one? Psychological Bulletin 91 (3), 517–539. doi:10.1037/0033-2909.91.3.517

Hutchins, E., Lintern, G., 1995. Cognition in the Wild. Vol. 262082314. MIT press Cambridge, MA.

Kerr, N. L., Tindale, R. S., Jan. 2004. Group performance and decision making. Annual review of psychology 55, 623–55. doi:10.1146/annurev.psych.55.090902.142009

Kirsh, D., Jan. 2006. Distributed cognition: A methodological note. Pragmatics & Cognition 14 (2), 249–262. doi:10.1075/pc.14.2.06kir
Laughlin, P. R., Kerr, N. L., Davis, J. H., Halff, H. M., Marciniak, K. A., 1975. Group size, member ability, and social decision schemes on an intellective task. Journal of Personality and Social Psychology 31 (3), 522–535. doi:10.1037/h0076474.

Piau, D., 2011. A simpler solution of the integral \( \int_{a \geq x_1+...+x_n} \exp \left[ -\pi \left( x_1^2 + \ldots + x_n^2 \right) \right] dx_1 \cdots dx_n \). Mathematics - Stack Exchange. URL http://math.stackexchange.com/q/61215 (version: 2011-09-01)

Rauhut, H., Lorenz, J., Nov. 2010. The wisdom of crowds in one mind: How individuals can simulate the knowledge of diverse societies to reach better decisions. Journal of Mathematical Psychology. doi:10.1016/j.jmp.2010.10.002

Sorkin, R. D., Hays, C. J., West, R., 2001. Signal-detection analysis of group decision making. Psychological Review 108 (1), 183–203. doi:10.1037/0033-295x.108.1.183

Appendix A. Approximations

\( P(c) \) can be expanded in Taylor series of \( c \) around \( c = -b \).

\[
P(c) = P[-b + (c + b)] = P(-b) + (c + b)P'(-b) + \frac{(c+b)^2}{2} P''(-b) + \frac{(c+b)^3}{6} P'''(-b) + \ldots
\]

(A.2)

where \( P^{(i)}(-b) \) can be found explicitly using (1),

\[
P^{(i)}(c) = \frac{1}{\sigma^i} H^{(i)} \left( \frac{c+b}{\sigma} \right).
\]

(A.3)

In particular \( H(0) = 1/2, H'(0) = 1/\sqrt{2\pi}, H''(0) = 0, H'''(0) = -2/\sqrt{2\pi}. \)

In general, making use of Hermite polynomials,

\[
H^{(i+1)} = (-1)^{i/2} \frac{i!}{\sqrt{2\pi}(i/2)!}
\]

(A.4)

for odd \( i > 0 \) and \( H^{(i)} = 0 \) for even \( i > 0 \).

Consequently,

\[
P(c) = \frac{1}{2} + \frac{(c+b)}{\sqrt{2\pi}\sigma} + O \left[ (\frac{c+b}{\sigma})^3 \right],
\]

(A.5)
that it, the approximation error of taking the linear approximation is of the order $(c + b)^3/\sigma^3$ as the quadratic term vanishes. Plugging $c = 0$ we obtain

$$P(0) = \frac{1}{2} + \frac{b}{\sqrt{2\pi}\sigma} + O\left(\frac{b^3}{\sigma^3}\right)$$  \tag{A.6}$$

$$= \frac{1}{2} + sb + O((sb)^3)$$  \tag{A.7}$$

and similarly, the derivative of (A.5) in 0 is

$$P'(c)|_{c=0} = \frac{1}{\sqrt{2\pi}\sigma} + \frac{1}{\sqrt{2\pi}\sigma} O\left(\frac{b^2}{\sigma}\right)$$  \tag{A.8}$$

$$= s \left[1 + O(s^2b^2)\right].$$  \tag{A.9}$$

The last equation gives the approximate equation for slope (7). Another expression

$$P(0) - 1/2 = \frac{b + bO((sb)^2)}{1 + O((sb)^2)} = b \left[1 + O(s^2b^2)\right]$$  \tag{A.10}$$

yields in the approximate equation for bias (7).

**Appendix B. Voting**

$$P_{Vot}(c) = \sum_{k=1}^{n} \sum_i \left[1 - P_{i1}(c)\right] \cdots \left[1 - P_{ik}(c)\right] P_{i_{k+1}}(c) \cdots P_{in}(c)$$  \tag{B.1}$$

$$+ \left[\frac{1}{2} \sum_i \left[1 - P_{i1}(c)\right] \cdots \left[1 - P_{n/2}(c)\right] P_{n/2+1}(c) \cdots P_{in}(c)\right],$$

if $n$ is even.

After plugging the linearization (A.5) in the above, and using $\mu_i = s_i(b_i + c)$, each part has the form of

$$\left[\frac{1}{2} - \mu_{i1} + O(\mu_{i1}^3)\right] \cdots \left[\frac{1}{2} - \mu_{ik} + O(\mu_{ik}^3)\right]$$  \tag{B.2}$$

$$\times \left[\frac{1}{2} + \mu_{ik+1} + O(\mu_{ik+1}^3)\right] \cdots \left[\frac{1}{2} + \mu_{in} + O(\mu_{in}^3)\right]$$  \tag{B.3}$$

$$= \frac{1}{2^n} - \frac{1}{2^{n-1}} (\mu_{i1} + \ldots + \mu_{ik}) + \frac{1}{2^{n-1}} (\mu_{ik+1} + \ldots + \mu_{in})$$  \tag{B.4}$$

$$+ O(\mu_{i1}^2) + \ldots + O(\mu_{in}^2)$$  \tag{B.5}$$
After applying permutations to the main part (i.e. without the error estimation) we get

\[
\frac{1}{2^n} \binom{n}{k} + \frac{1}{2^{n-1}} \binom{n}{k} \left[ -k + (n-k) \right] \frac{\mu_1 + \ldots + \mu_n}{n} \tag{B.6}
\]

\[
= \frac{1}{2^n} \binom{n}{k} + \frac{n}{2^{n-1}} \left[ - \binom{n-1}{k-1} + \binom{n-1}{k} \right] \frac{\mu_1 + \ldots + \mu_n}{n}, \tag{B.7}
\]

which is easy to be summed. The first component sums to 1/2. In the second, binomial coefficients cancel pairwise, except for \( \binom{n-1}{0-1} = 0 \) and \( \binom{n-1}{\lfloor (n-1)/2 \rfloor} \) as some of the elements cancel, leaving only \( \binom{n-1}{k} \) for \( k = \lfloor (n-1)/2 \rfloor \). Consequently, when \( n \) is odd, one gets

\[
P_{\text{Vot, odd}}(c) = \frac{1}{2} + \frac{n}{2^{n-1}} \binom{n-1}{(n-1)/2} \frac{\mu_1 + \ldots + \mu_n}{n} + O(\mu_1^2) + \ldots + O(\mu_n^2) \tag{B.8}
\]

and for even \( n \)

\[
P_{\text{Vot, even}}(c) = \frac{1}{2} + \frac{n}{2^{n-1}} \binom{n-1}{(n-2)/2} \frac{\mu_1 + \ldots + \mu_n}{n} + O(\mu_1^2) + \ldots + O(\mu_n^2). \tag{B.9}
\]

After the differentiation one obtains the slope [7] and the bias [8].