Information-Sharing in Three Interacting Minds Solving a Simple Perceptual Task

Michał Denkiewicz (michal.denkiewicz@gmail.com)
Department of Psychology, University of Warsaw
Stawki 5/7, 00-183 Warsaw, Poland

Joanna Rączaszek-Leonardi (raczasze@psych.uw.edu.pl)
Institute of Psychology, Polish Academy of Sciences
Jaracza 1, 00-378 Warsaw, Poland

Piotr Migdał (piotr.migdal@icfo.es)
ICFO–Institut de Ciències Fotòniques
08860 Castelldefels (Barcelona), Spain

Dariusz Plewczynski (darman@icm.edu.pl)
Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw
Pawińskiego 5a, 02-106 Warsaw, Poland

Abstract

We study three-person groups solving a simple, two alternative forced choice task of perceptual nature. The group members provide individual answers and afterwards discuss and reach a joint decision. Different models of information sharing describe the theoretical relationship between group and individual performance. Experimental data shows that average-performing members can benefit from cooperation, but the groups do not outperform their best members. Results point to voting as the best explanation of the behavior of the groups.

Keywords: group decision making; distributed cognition; information sharing

Introduction

Whether "two heads are better than one" is no settled question in psychology. Many studies report groups to be less proficient than their most capable members (Corfman & Kahn, 1995), or that there is no benefit (Heaney, Foster, Gregor, O'Neill, & Wood, 2010). Groups are often regarded as source of negative influence on individual performance, stemming from conformism (Asch, 1951) or social loafing (Allport, 1924).

There are, however, studies that report benefit from cooperation (Kerr & Tindale, 2004; Hastie & Kameda, 2005). In one of such studies Bahrami (B. Bahrami et al., 2010) determined that dyads can outperform their members in a simple two-choice perceptual task, provided that those members have similar individual effectiveness. This group benefit disappeared when communication was forbidden; hence free information sharing was a key factor.

This result is interesting because there is no obvious reason for group benefit to occur in such a simple task. For certain types of tasks, such as concept mastery, concept attainment or learning, group members can pool cognitive resources, or utilize complimentary skill or knowledge. Then a group can provide better solutions in terms of quality, though not necessarily efficiency (Steiner, 1966; Hill 1982). A notable exception is brainstorming, where participating in a group has a negative impact both on quality and quantity of creative solutions (Taylor, Berry, & Block, 1958).

If resources cannot be shared, perhaps solutions can be. For so-called "eureka-tasks" a solution can be demonstrated in objective terms, thus (at least in theory) a single participant, who finds the correct answer, can easily persuade other members, leading to a correct group solution. Hence, the chance of group solving a task grows with its size. Examples of such task are Remote Associates Test (Laughlin & Bitz, 1975), or simply scrambled letters/anagrams.

In Bahrami's study neither of these conditions was met. The participants first performed the perceptual task on their own, without the ability to divide it into parts. Then a group decision was made, based solely on what the participants perceived individually. No reasoning or previous knowledge or was of any use and the only thing, that the participants could communicate was their subjective idea, of what they think the answer was. Still not only did the pairs perform better, than chance (which in this case means the averaged effectiveness of the two participants), but they also outperformed the better of the two members.

Bahrami tested several theoretical models of information sharing in communication, developed in the spirit of signal detection theory, and concluded, that dyad members communicate their own relative confidences, which allows the group benefit to occur.

Later these models were theoretically extended to groups of arbitrary size by Migdał (Migdał, Rączaszek-Leonardi, Denkiewicz & Plewczynski, 2012). "Aggregation of decisions" was also considered, that is a situation when subgroups of a larger group first reach their decisions, and then try to convince one another.

The topic of this paper is an experimental attempt to verify the applicability of these theoretical models to three-person groups, and shed light onto how these groups reach their decisions.

Decision-Making Models

Consider the following task. On each trial a group consisting of three participants views two sets of stimuli, one after another, for a brief period (85 ms). Each of the two sets consists of six Gabor patches. All patches are identical, with the exception of one (target) patch – which has higher contrast. The task is to determine which screen, the first or the second, contains the target patch. The difference in contrasts between the target patch and non-target patches determines the difficulty of the task. We use the convention, that this difference is positive, when the target is in the second set, and negative, if it's on the first.

Knowing the answers of a given a participant, we can determine the probability, that he or she chooses the second answer as a function of the contrast difference. We model this relationship with a cumulative of the normal distribution, as it provides a good fit to the experimental data (Bahrami et al., 2010).

Again, following the convention we describe this psychometric function with two parameters: slope (s) and bias (b). The slope is a measure of participant's performance and is of primary interest. The bias describes the tendency to give one particular answer – the task is constructed in such a way, that this parameter should be close to 0. These parameters are related to the standard parameters of a Gaussian curve (mean μ and standard deviation σ) in the following way:

$$s = \frac{1}{\sqrt{2\pi}\sigma} \qquad (1)$$

$$b = -\mu \qquad (2)$$

In the same way as individual decision are used to construct curves for participants, group curves can be obtained from group decisions, allowing for a comparison of group and individual slope parameters.

By making different assumptions about communication and decision making process within the group, a theoretical dependency between individual slopes s_1,s_2,s_3 and group slope parameter s_g can be established, in the form of a function $s_g(s_1,s_2,s_3)$.

If the behavior of the group actually matches the assumptions, then the empirically obtained group slopes should not differ significantly from theoretical predictions. By an information-sharing model we understand a possible set of such assumptions. We consider six such models:

Random Responder (RR) This model assumes that the communication is actually ineffective, and the final decision is randomly chosen from the individual decisions.

Best Decides (BD) It is plausible that the group will simply entrust the decision to the biggest "expert" in the task. That person's decision becomes the group decision. The model assumes that the group initially possesses the knowledge about who is the best member. This is a somewhat simplistic assumption, yet it is feasible that the best member can be determined in the beginning, based on a small number of trials.

Voting (Vot) The group uses the majority rule to determine the group decision. It requires only the communication of individual decisions.

Weighted Confidence Sharing (WCS) Each member shares his/her own relative confidence. The participants are unable to discern their perceived contrast difference from its reliability (determined by the participant's slope).

Direct Signal Sharing (DSS) Group members communicate both their perceived contrast difference and their confidence separately. This allows for a statistically optimal decision (Sorkin & Hays, 2001).

Truth Wins (TW) In this model each participant either knows the correct answer, or is aware of not knowing it. In other words a person cannot falsely believe that he knows the answer. This basically means, that if a single group member finds out the correct answer, the group also answers correctly.

Table 1 summarizes the mathematical relationships between group and individual performance, according to the formulas for arbitrary group sizes presented in Migdał et al., 2012.

Table 1: Group slope as a function of individual slopes according to different models.

Model	$s_g(s_1, s_2, s_3)$
RR	$\frac{s_1 + s_2 + s_3}{3}$
Vot	$\frac{s_1 + s_2 + s_3}{2}$
BD	$max\{s_1, s_2, s_3\}$
WCS	$\frac{s_1 + s_2 + s_3}{\sqrt{3}}$
DSS	$\sqrt{s_1^2 + s_2^2 + s_3^2}$
TW	$s_1 + s_2 + s_3$

All models, with the exception of the RR model, predict that group performance should exceed average individual performance of that group's members.

In Bahrami's (Bahrami et al. 2010) study of dyads, the WCS model best described the behavior of the participants.

We expected that such behavior should also be seen in three-person groups, that is the groups would outperform their best members, at least in cases of homogenous groups (in terms of performance).

Experiment

Subjects

Participants were recruited from general population of Warsaw, Poland, using snowball method. There were 15 three-person groups, which gives 45 participants: 30 female and 15 male. All participants were adults (mean age 38, s.d. 15.6) with normal or corrected-to-normal vision. Members of each group knew each other. Written informed consent was obtained from each participant prior to the experiment and each person was rewarded with 25 PLN (approximately 6 EUR) for completing the experiment.

Experimental setup

The testing room contained three computer stations, each with a LCD display (24 inch, resolution = 1920 x 1280, refresh rate = 60 Hz) and a keyboard. The stations were connected via a local network and arranged in the form of an equilateral triangle, with the displays facing outwards. Each participant saw only his/her own display, but was able to see the fellow group members' faces. To minimize distractions, during the experiment the room was nearly completely dark – the only sources of light were the monitors. However, the lighting was sufficient for the participants to see each other's faces.

Each computer was assigned a color: blue, orange or yellow, for the purpose of identification. The experiment was controlled by custom software based on the PsychoPy framework ("http://www.psychopy.org/"; Peirce, 2008).

Stimuli

Each of the two stimuli sets consisted of six Gabor patches evenly distributed around the center of the screen, at a distance of 8 degrees. All patches were vertically oriented and had the following parameters: standard deviation of the Gaussian envelope = 0.45 deg. spatial frequency = 1.5 cycles/deg. The contrast parameter equaled 10% for the non-target patch and 11.5% 13.5%, 17% or 25% for the target patch (hence, the contrast difference value was 1.5% 3.5%, 7% and 15%, respectively). The position of the target patch within its set was chosen randomly each trial.

The background was uniform and gray at all times.

Task and Procedure

The experiment consisted of a practice block followed by three experimental blocks. After each experimental block the participants changed their places, moving one seat to the left, so that each participant spent about the same time using each computer. The experimenter was present in the testing room during the entire experiment, to assure that the procedure was followed.

There were 288 experimental trials - three blocks of 96 trials. The practice sessions consisted of 8 trials. The number of trials with each combination of difficulty level and correct answer combination was equal within blocks. Each trial started with a black fixation cross, placed in the center of the screen, displayed for 500 ms. The two sets of stimuli followed, each visible for 85 ms, separated with a blank screen presented for 1000 ms. Finally, a white question mark appeared, indicating that the participants are to make their individual decisions. The decisions were made by pressing an appropriate key - left or right arrow indicating that the target was in the first or in the second set respectively. The keys were labeled "1" and "2". After a button had been pressed the question mark was replaced with a message "Wait for other participants' decisions". It stayed on the screen until all individual decisions were made. So far the participants were not allowed to communicate with each other.

Next, the group decision phase followed. Individual decisions were displayed for 1.5 s, one above the other, each in the color of the respective computer. Then a message appeared asking the participants to discuss the group decision. After the group decision has been agreed upon, a single person was required to input the decision, in the same way as the individual decisions were made. This was always the person to the left of the previous decision maker.

During the group decision phase the participants [] could communicate freely, but were not allowed to leave their seats. The method they could use to arrive at a group decision was not constrained in any way, there was no predefined decision-making scheme (e.g. voting).

After the group decision had been made, feedback was displayed, containing information about the correctness of the group decision and of each individual decision. Feedback was visible for 1.5 s and after it disappeared, the next trial followed immediately.

If the individual decisions were unanimous, the group decision was automatically assumed to be the same. In such case the group decision phase was skipped and feedback was displayed.

The duration of the experimental session depended solely on the pace at which the participants completed the trials, and this was on average 44 minutes.

Results

For each participant and each group the slope, *s*, and bias, *b*, parameters were estimated by fitting a probit regression model to that individuals (or that groups) decisions. Individual slopes were used to compute the theoretical values of group slope, predicted by each model, according to appropriate formula from Table 1 (s_{model}). These values were compared to the values obtained empirically (s_{group}). If the ratio of the empirical and theoretical value, s_{group} / s_{model}, was greater than one, it means that the groups outperformed the model, if it was less than one – the groups did not reach predicted accuracy. A one sample Student's t-test was used to determine the significance of the deviations from theoretical predictions; the quotient was compared with the value of 1. The results are summarized in Table 2.

Table 2: Experimental results.

model	*M*	*SD*	*t(14)*	*p*	
RR	1.42	0.32	5.16	<.001	***
BD	0.93	0.24	-1.11	.29	
Vot	0.95	0.21	-0.94	.36	
WCS	0.82	0.18	-3.77	.002	**
DSS	0.71	0.15	-7.19	<.001	***
TW	0.47	0.11	-19.23	<.001	***

Note: * $p < .05$. ** $p < .01$. *** $p < .001$.

The ratio s_{group} / s_{model} was significantly greater than 1 for the RR model and significantly less than 1 in the case of WCS, DSS and TW models. In other words the actual slope parameters of the groups were significantly greater than those predicted by RR model, and significantly lower, than the predictions of the WCS, DSS, TW models.

In case of some models, namely BD, Vot and TW, a group decision on a given trial can be determined from the individual decisions. This allows for a per-trial comparison of the theoretical decisions and the decisions actually made. A group decision agrees with the BD models if it is the same, as the decision of the best group member (i.e. the one with the highest slope parameter). In case of the Vot model it is the decision made by the majority of participants. Finally, in the TW model it is the correct answer, unless all members were wrong. For each group three values were calculated, indicating the number of group decisions that were consistent with each of these tree models. A paired t-test was performed for each pair of these models. The differences in consistency between the Vot model (M = 254.7, SD = 21.18) and both the BD model (M = 218.7, SD = 27.6, p < .001), and TW model, (M = 213.5, SD = 19.6, p < .001) were significant and positive. The difference in consistency between BD model and TW model was not significant (p = .77).

Discussion

Analysis of individual and group slopes allows us to reject the RR, WCS, DSS and TW models, leaving Vot and BD as plausible explanations of group performance. A trial-by-trial analysis points to Vot model (majority voting) as the best explanation.

The rejection of some models does not mean that the types of behavior they describe did not occur or that they are impossible. It merely shows that they were not dominant in the course of the experiment. Indeed in a small, but not negligible number of trials (about 12%), the group decisions corresponded neither to BD nor Vot model predictions.

Conclusions

Results of the experiment indicate that three-person groups prefer voting as a method of reaching a joint decision, and more advanced communication is rarely employed. Groups outperformed, so to say, their average members, but best members generally made better decisions than the group. This group benefit can be attributed solely to the use of voting as a decision-making scheme.

The failure to outperform the best group members, as it was in the case of pairs in the study by Bahrami et al. (2010), can be explained in many ways. First, the requirement for cognitive resources used for communication and integration of information increase as more group members are added, leading to deterioration of information processing performance. Secondly, in groups of size three, as opposed to dyads, voting becomes possible. Since group members are not directly rewarded for accuracy, and the experimental task is somewhat tedious, employing a simple, relatively good, and socially acceptable method of reaching a group decision seems tempting.

Conformism, or caring for group's coherence, can also play a role, as it can shy away a single correct group member, from confronting the majority decision (and prolonging the decision phase). The social acceptance of voting can afterwards serve as a justification for this group member, if he or she were blamed for not insisting on the correct answer. On the other hand the responsibility for being wrong diminishes, if one is the member of a majority.

It is feasible that the impossibility of automatically resolving a tie in dyads fosters communication and, in turn, increases performance. Adding a third member provides an opportunity to use a simpler and less effective decision making system and, paradoxically, diminish performance. This shows how seemingly simple task and situations can produce non-trivial dependencies.

Acknowledgements

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

References

Allport, F. H. (1924). Response to Social Stimulation in the Group. *Social Psychology* (pp. 260–291). Boston: Houghton Mifflin Company.

Asch, S. E. (1951). Effects of group pressure on the modification and distortion of judgements. In H. Guetzkow (Ed.), *Groups, leadership and men.* (pp. 177–190). Pittsburgh:PA: Carnegie Press.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally Interacting Minds. *Science*, *329*(5995), 1081–1085.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, *10*(8), 6–10.

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

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological review*, *112*(2), 494–508.

Heaney, R., Foster, F. D., Gregor, S., O'Neill, T., & Wood, R. (2010). Are two heads better than one? An experiment with novice share traders. *Australian Journal of Management*, *35*(2), 119.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual review of psychology*, *55*, 623–55.

Laughlin, P. R., & Bitz, D. S. (1975). Individual versus dyadic performance on a disjunctive task as a function of initial ability level. *Journal of Personality and Social Psychology*, 31(3), 487–496.

Migdał, P., Rączaszek-Leonardi, J., Denkiewicz, M. & Plewczyński, D. (2012). Information-sharing and aggregation models for interacting minds, *Journal of Mathematical Psychology*, 56, 417–426.

Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in neuroinformatics*, *2* (January), 10.

Sorkin, R., Hays, C., (2001). Signal-detection analysis of group decision making. *Psychological Review - New York* - 108 (I), 183–203.

Steiner, I. D. (1966). Models for inferring relationships between group size and potential group productivity. *Behavioral Science*, 11, 273-283.

Taylor, D.W., Berry, P.C., Block, C.H. (1958). Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administrative Science Quarterly*, 3, 23-47.