Phenomenological theory of collective decision-making

Anna Zafeiris[a], Zsombor Komán[b], Enys Mones[c], Tamás Vicsek[a,b]

[a] *Statistical and Biological Physics Research Group of HAS, Pázmány Péter sétány 1A, H-1117, Budapest, Hungary*
[b] *Department of Biological Physics, Eötvös University, Pázmány Péter sétány 1A, H-1117, Budapest, Hungary*
[c] *Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

Abstract

An essential task of groups is to provide efficient solutions for the complex problems they face. Indeed, considerable efforts have been devoted to the question of collective decision-making related to problems involving a single dominant feature. Here we introduce a quantitative formalism for finding the optimal distribution of the group members competences in the more typical case when the underlying problem is complex, i.e., multidimensional. Thus, we consider teams that are aiming at obtaining the best possible answer to a problem having a number of independent sub-problems. Our approach is based on a generic scheme for the process of evaluating the proposed solutions (i.e., negotiation). We demonstrate that the best performing groups have at least one specialist for each sub-problem but a far less intuitive result is that finding the optimal solution by the interacting group members requires that the specialists also have some insight into the sub-problems beyond their unique field(s). We present empirical results obtained by using a large-scale database of citations being in good agreement with the above theory. The framework we have developed can easily be adapted to a variety of realistic situations since taking into account the weights of the sub-problems, the opinions or the relations of the group is straightforward. Consequently, our method can be used in several contexts, especially when the optimal composition of a group of decision-makers is designed.

Keywords: Collective behavior, Decision-making, Interdisciplinary research, Optimal decisions

Highlights

- Quantitative formalism of complex collective decision-making scenarios is proposed.
- We search for the optimal competence distribution of heterogeneous agents.
- The best groups have at least one specialist for each sub-problem.
- The specialists have some insight into other sub-problems as well.
- Good agreement with empirical results obtained from large-scale citation database.

1. Introduction

Addressing the process of collective decision making has represented a great scientific challenge for a long time [1, 2, 3, 4]. It is a highly relevant aspect of the behavior of social groups, in particular, because as it has been argued, measured and shown analytically: the “wisdom of crowds” can go qualitatively beyond that of the individuals’ [2].

This statement also holds for animal assemblies [5, 6, 7]. A rarely considered, but essential case is when the problem to be solved is complex, i.e., has many facets. Under such conditions the quality of the collective solution is highly influenced by the composition of the group. Obviously, if the members of the group are identical, the group’s performance can hardly go beyond that of any of its member’s. However, if the problem to be solved is complex – i.e., has a number of different aspects or “dimensions” [8] – a group having members specialized in their respective kinds of sub-problems is expected to be much more efficient in providing a high quality answer, than a uniform one. The stress is on the independent nature of the sub-problems, making the problem high-dimensional. In a way our present work can be considered as a quantitative approach to the problem of division of labor [9, 10] in the context of collective decision making (the task/labor is to bring about a decision; the division is made among the specialists of the sub-problems).

In spite of the above almost trivial observation regarding heterogeneous, diverse or “multidimensional” groups, a quantitative demonstration of its validity needs a carefully constructed framework. Prior works involving quantitative analysis have almost exclusively focused on problems that could be regarded as “one-dimensional” [2, 11, 12, 13] from our point of view which considers a problem having several dimensions (being multidimensional) if it can be broken down into sub-problems, each having its own characteristic feature independent of those of the others’. In the case of one-dimensional problems it has been demonstrated – using approaches from theory (see, eg., the pioneering works [11, 13]) through genetic optimization [13] to agent based modeling/simulations [14] and observations [15, 16] – that diverse groups can outperform homogeneous ones.

Intuition suggests that a group of specialists (one competent person for each sub-problem) should be optimal regarding the quality of the solution with the constraint of minimizing costs at the same time. Here we present a generic agent-based approach which – due to its minimal assumptions – quantitatively demonstrates that the breadth of knowledge of its members makes a group more efficient, i.e., being capable of using a smaller amount of resources to produce a more beneficial solution in a wide variety of potential applications. This is what corresponds to the “synergy” resulting in a better decision relative to the one following from a simple “linear” aggregation of the proposed solutions. And what we show in our work is how this synergy can emerge from a negotiation process. Naturally, negotiation is absent (generally) in animal societies. Specialization is the result of age or hormone level etc.

Many opinion formation models exist in the contemporary literature, among which many considers ”heterogeneous” agents as well, often with continuous opinion values (For a review see [17]) However, agents in these studies are usually heterogeneous regarding their (i) confidence thresholds (or bounds of confidence, meaning that interacting agents adjust their opinions towards that of the others, but only if the two opinions are closer to each other than a certain threshold, a phenomenon closely related to the one called homophily), (ii) conviction, or (iii) influencing ability (aka. social influence). In contrast, our agents are homogeneous with respect to the above mentioned characteristics, but they are heterogeneous regarding their abilities, and what is more, their entire spectrum of abilities.

Other fundamental differences between the opinion formation models in contemporary literature and our approach include the followings:

- Most of them consider two-valued opinions (0/1, yes/no, etc.), motivated by the Ising-model. Their popularity is due to their simlicity, despite which they can lead to very deep results [18].

- Most contemporary models assume a *simple* update rule: a (usually randomly selected) agent simply changes opinion "suitably" to its neighbours. For example, if some neighbours of the selected agent share an opinion, the focal agent simply adopts it. In contrast, we detail the mechanism of "convincing": how it happens in iterative rounds with members evaluating the proposals of others and discussing it, all of which is affected by personal abilities.

- Contemporary models usually consider entire societies (often even assuming that $N \to \infty$), with mostly binary interactions. In contrast, we consider a relative small group ($N \approx 10$), but with intense interaction, in which all members participate.

- The aim of the above mention models is usually to gain an insight of the *spread and dynamics* of opinions, with emphasis on occurrent consensus or stalemate situations. In contrast, *we aim to find the optimal composition of a group, regarding the characteristics of the members* – in this case, (multidimensional) abilities.

A paradigmatic example for our approach is that of a board of directors for a large company (however, there are many other possible examples ranging from a group of animals searching for resources up to a government or simply a team carrying out interdisciplinary research). In the case of a board of directors a potential candidate problem is that of finding the best possible placement and product for a new factory. Obviously, the various aspects of this problem are quite diverse, each of them requiring specific knowledge, i.e., the decision involves knowledge of the history of the given country, various features of the labor force (education, etc.), geographical and logistic conditions, potential market in the region, and so on. It is an important feature of the situation that the members of the group cannot get any information about the quality of their propositions from an "outsider" who could know the optimal solution *ab ovo*.

2. The model

2.1. Formalizing interdisciplinary decision-making

We have aimed at a model that is simple, but is still appropriate for projecting a wide class of realistic situations onto it. In order to do so, we consider groups of N individuals solving a problem P having M sub-problems P_j $(j = 1, 2, ..., M)$ such that for addressing each sub-problem a unique (specific) skill is needed. Referring to the company example introduced in 1, the board of directors counts N members, P is the problem of finding the best place for a new factory, having such sub-problems as: knowledge of: (P_1) the law and taxation systems in the candidate countries, (P_2) logistic conditions, (P_3) local working culture and education, etc.

Thus, we are dealing with a set of $N \times M$ abilities or levels/degrees of skill, A_{ij} $(i = 1, 2, ..., N)$, proportional to the ability/competence (e.g., accuracy) of an individual i to give the best answer for the jth sub-problem. We do not initially specify the A_{ij} parameters: the method we apply (optimization with genetic algorithm) will result in their optimal values.

Without losing the generality of the above setting, we assume that A_{ij}-s take their values from the unit interval $[0, 1]$. The ability matrix A_{ij} is also related to the costs involved in finding a solution (since acquiring a high ability to successfully address a sub-problem involves costs, such as experience, learning, etc.). It is obvious that the cost of obtaining an ability A is typically not a linear function of A, since achieving the capacity of perfect knowledge $(A = 1)$ is much more costly than achieving a partial knowledge (e.g., $A = 0.5$). For the sake of simplicity we assume that the cost C for obtaining ability A, is

$$C = f(A) = Const \cdot A^x \tag{1}$$

where $1 < x$, and $Const$ is a constant corresponding to the relative weights of the costs, when calculating the fitness of a group for given A_{ij}-s. We start with a random distribution of the A_{ij} values and search for their optimal distribution (by letting them evolve). Here optimal distribution means one which provides the best possible solution for the smallest possible – or for a given prefixed – cost.

2.2. The stages of collective decision-making

In our formal model, the process of collective decision-making is divided into four basic stages (See Chart 1).

1. Each group member i suggests a solution for each sub-problem P_j in such a way that the quality of the given proposition Q_{ij} depends only on i's corresponding ability, A_{ij}. This assumption, in the simplest case means that

$$Q_{ij} = A_{ij}. \tag{2}$$

   In other words, we assume that specialists provide high-quality propositions for their own field-of-expertise, while people without the know-how provide inefficient ones. (Adding noise to the above equation did not change our results.)

2. During the "information diffusion" phase, members interact by evaluating each other's proposals (each member evaluates all the propositions). The evaluation made by member i' regarding the quality Q_{ij} is denoted by $E_{ij}^{i'}$ and it is proportional to both Q_{ij} (the quality of that given proposal) and $A_{i'j}$ (the savvy of i' for field j). The accuracy of such an evaluation is distorted by a stochastic factor representing that those members who have small abilities to evaluate a proposal tend to make mistakes in their appreciation with an amplitude involving randomness. These evaluations ($E_{ij}^{i'}$) represent the central ingredient of our approach.

3. These are next (in several rounds of an imaginary "round table discussion") modified by further interactions (communication/evaluation) with other group members, i'', chosen with a probability proportional to their abilities concerning problem P_j, $A_{i''j}$. The total number of additional evaluations in a given decision-making event is equal to $X\%$ of N. This step refers to the stage when somebody (most often, but not always an expert of the given field) tries to convince other members of the group about her/his opinion by sharing her/his ideas. Characteristically, 10, 20 or 30 % of the group can give an evaluation(remark/speech) for each subproblem.

4. The quality of the solution for a given P_j is obtained by accepting the proposal of member i^* receiving the highest average evaluation

$$E_j^{max} = E_{i^*j}, \tag{3}$$

   where

$$i^* = argmax_i E_{ij} = argmax_i avg_{i'} E_{ij}^{i'}, \tag{4}$$

by the other members concerning his/her proposition for the solution of problem P_j i.e.,

$$Q_j^{max} = Q_{i^*j} = A_{i^*j}. \tag{5}$$

The quality Q of the solution given for P – provided by the whole group – is then obtained by aggregating the proposals having the highest evaluations for the P_j-s after the last round.

Note that the concrete problem (P) is not specified (just an example is given). In addition, we have only two arbitrary parameters (the level of stochasticity, $Rand$, during the second evaluation step, plus the proportion of the evaluators $X\%$, see also Chart 1. N and M are simple input parameters depending upon an actual situation. The description of the process may seem lengthy, however, it directly corresponds to our everyday practice during group decisions.

From the algorithmic point of view, the input of the above described process (represented on Chart 1 is an A_{ij} ability matrix, and its output is a fitness value F, by which we measure the "decision making quality" of the group. F is calculated from the quality of the solution (Q) and from the cost of knowledge (C) needed to obtain such an answer (see Eq. 1) as

$$F = Q - C \tag{6}$$

2.3. The flowchart of the model

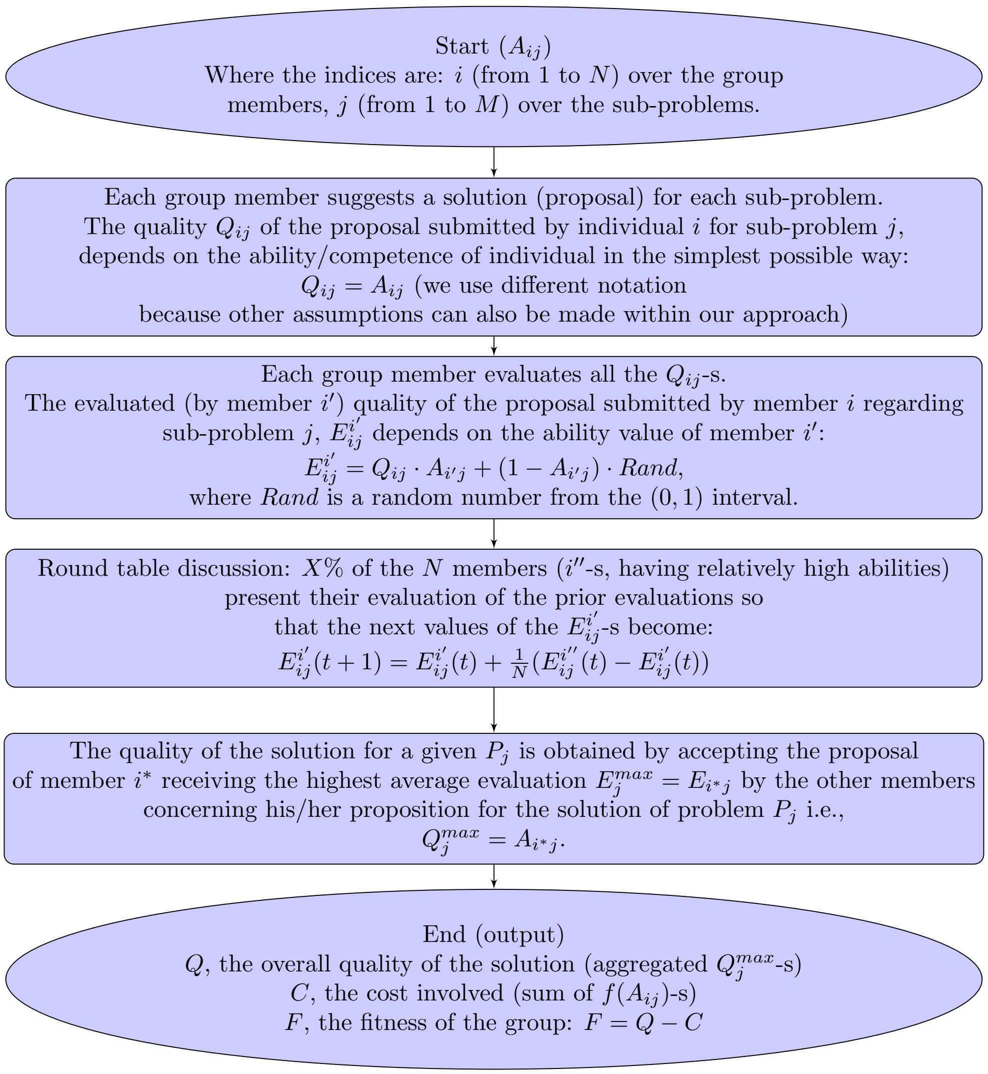

Chart 1: A simplified flow diagram of how the model works (the steps of a decision making process). The flow diagram represents a single step (within each generation) during the genetic optimization. For notations see the text. This chart gives a description of the process through which the corresponding fitness (efficiency) value F is calculated based on the given ability matrix (A_{ij}-s). In the next step of the genetic algorithm these F-s are used as weights based on which the "parents" of the new generation are chosen (after the combination of two parents, random perturbations, "mutations" are applied before finalizing the groups of the "young" generation).

3. Methods

We use a genetic algorithm [19] to find optimal solutions because this approach is known to be effective when extreme values for a function with a relatively large number of variables is being searched. Here relatively large means numbers above $8-10$, i.e., we are looking for optima of a function which is defined in a high-dimensional space. In addition, (just like in the case of fitness landscapes or the free energy landscape for spin glasses and alike) our fitness function is likely to have a huge number of local maxima, and a single (or a class of) configuration(s) (sets of A_{ij}-s with maximal fitness) corresponding to a global maximum. This is why we also integrate into our approach a technique analogous to the one called "simulated annealing" [20] used for finding the minima of the free energy in the case of problems from statistical mechanics. In our case this is realized by temporarily increasing the mutation rate when the actual solution seems to converge (stops changing as a function of the generation number). Even when applying this method, one cannot be sure that in the limit of a large number of generations the absolute optimum can be reached. Thus, usually a further, quite natural, and implicitly widely used approach is taken by assuming that the pseudo-global, optimal solutions possess the same statistical features.

It is important to note that this approach (optimizing *groups* with genetic algorithm) is not related to the question of kin versus group selection in any way. Genetic algorithm in this context is purely an optimization method.

For more details and parameters see the Appendix.

4. Results

4.1. Results from simulations

The core of our results is summarized on Figure 1 C. On this, each column represents a sub-problem (specialty), each row refers to an individual, and the color in their intersection indicates the ability/knowledge of the given individual in the given field (See the corresponding colorbar on Fig. 1 D). As it can be seen, there is exactly one red square in each column, meaning that exactly one expert is needed for each sub-problem. Up to this point, our results pretty much overlap with the general intuition. What is less intuitive is that the rest of the squares are not homogeneously dark blue (corresponding to close-to-zero knowledge), but they are all shades of blue, meaning that in a group, optimal decision can be made if everybody has an idea of some other people's field-of-experts. We assume that this is due to better flow of information.

In order to confront these results with the general intuition (holding that once a group has a specialist for all fields, no more "extra knowledge" is required from other members) we have compared the optimality of the two types of groups: "two-valued", when an ability value can be either 0 or 1, and "continuous" when the ability values can be anything between 0 and 1. Subsection 4.1.3 covers these results.

Of course, the group members and their specialties are commutable in the sense that different runs of the same optimization method result in different layouts (Fig 2). However, as long as these two characteristics hold true (i) one specialist for each sub-field and ii) group members need to have at least some level of know-how in their mates' field of experts) we consider the results as being the same. We discuss this question in more detail in subsection 4.1.2.

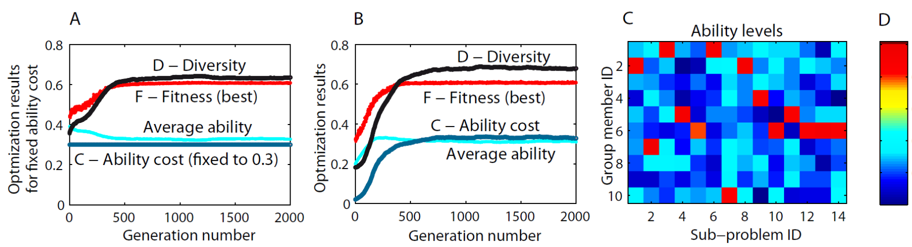

Figure 1: Illustration of both the process (A,B) and the end result (C) of calculating the optimal distribution of abilities/competences, A_{ij}^{max}, using a genetic optimization method. In (A) the generation number (G) dependence of the average fitness values (F) of the groups is plotted (red) for the fixed amount of cost, $C = 0.3$ (dark blue). The averaging is made over a population size of 2000 groups. The corresponding diversity, D, is indicated by the black line. The groups had $N = 10$ members and $M = 14$ sub-problems had to be answered. In (B) we display the evolution of the relevant parameters when the optimization is done with non-fixed ability cost C. (C) Displays the optimal ability matrix visualized with colors – the scale being indicated in (D). These results are for a generic case into which a few plausible assumptions are incorporated: the sub-problems have equal importance (weight) and $X = 30\%$ of the members take role in the round-table discussion. The most relevant message of (C) is that there is one specialist for each sub-problem (not necessarily one person per sub-problem) and, perhaps rather intriguingly, the specialists are found to have a clearly non-negligible competence concerning several of the other sub-problems. If we add some cost for the case when a single person is a specialist of more than one sub-problem, the solution ceases to have multiple specialties per person.

4.1.1. General properties

In Fig. 1 we show results for A_{ij}^{max}-s using equation 6, i.e., evaluating both the quality and the cost of the obtained A_{ij}-s and considering the average of the entire population at the end of the evolutionary process. In most of the figures – in addition to visualizing the values of A_{ij}^{max} – we also plot how the fitness F, the quality of the solution Q and the cost C changes as a function of the generation number G (as the population of groups evolves). In addition, we also display how the diversity D of the abilities depends on G. In all cases we find that the optimal distribution of the abilities is highly diverse. In all plots we use $N = 10$ and $M = 14$ without loss of generality (the main features of the optimal ability distribution do not differ qualitatively for different N and M pairs).

Figure 1 demonstrates some relevant features of both the process (the progress of the genetic algorithm) and the outcome of optimizing the ability distribution. Random initial conditions correspond to relatively low fitness and high costs. The efficiency/fitness of a group quickly increases at the first stage of the optimization. An important observation is that higher fitness is accompanied by larger diversity values (D), which – after [21] – is calculated as

$$D = \frac{\sum_{i,j}\left(\left(\max_i A_{ij}\right) - A_{ij}\right)}{M \cdot (N-1)} \qquad (7)$$

We have chosen this definition, because it differentiates among the diversity of distributions in a way being both in accord with the intuition and sensitive enough in the range determined by the actual distributions of A_{ij}-s throughout the simulations.

Our results come from simple and realistic assumptions regarding the “negotiation/discussion” process. Although the corresponding rules and calculations are not trivially transparent at all, nevertheless a relatively plausible interpretation for the main result can be provided. Perhaps the most essential step in our algorithm is the one when the group members, one after another, provide an evaluation of the proposals of the other members. If a member has zero ability to evaluate the proposal for a given sub-problem, then the contribution of this member to choosing the otherwise very good proposition becomes totally erratic (see the equation in Chart 1 for $E_{ij}^{i'}$). Conversely, even a relatively small ability to estimate the right value of a proposal results in a decreased level of randomness in the evaluation and, in this way, provides a more accurate estimated proposition quality. When the evaluations are aggregated to choose the best answer, the latter, more consistent contributions become to play an essential role.

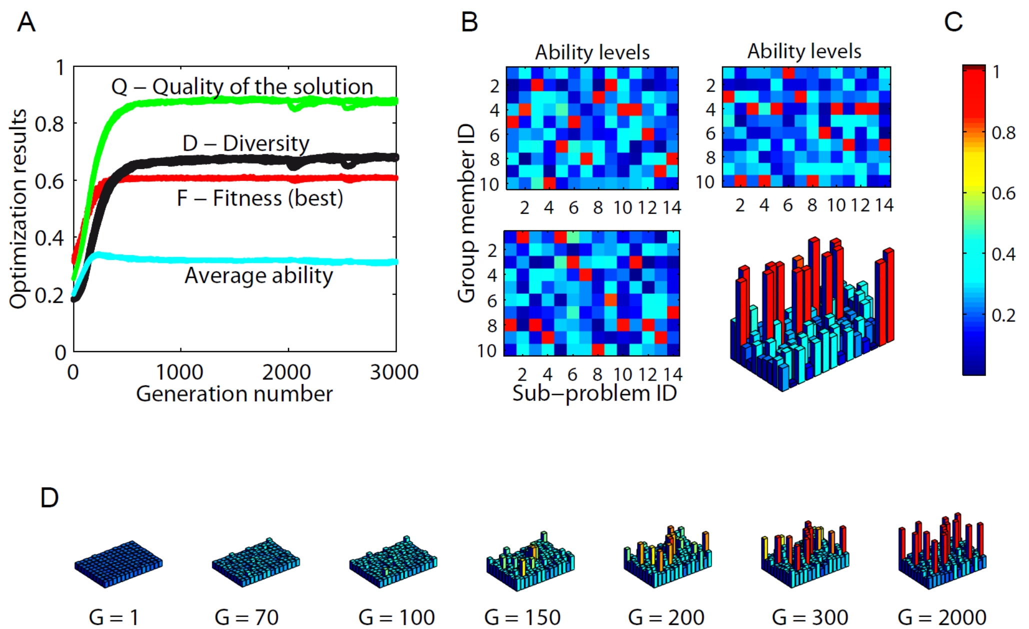

Figure 2: Visualization of the optimization for four different (random) initial conditions and (stochastic) realizations. Although the individual ability distributions are different, they correspond to about the same level of optimality which can be seen from the four curves virtually overlapping in all cases. The wiggles around $G = 2000$ and $G = 2500$ are due to the momentarily increased level of perturbations or "mutations" within the genetic algorithm (in the spirit of simulated annealing, see Materials and methods). (D) displays the development of the ability matrix as the genetic algorithm progresses. Here and in one of the displays in (B) a combination of column heights and colors is used to visualize the values of the ability matrix.

4.1.2. Robustness

Next we investigate the robustness of the new results stemming from our approach by testing the method on a few specific conditions. First, we start the optimization from different initial conditions and check whether the results are consistent with each other (have the same overall features). Figure 2 shows two different directions of the comparison. In Fig. 2A we show how similarly the main quantities (A, D, F and Q) evolve during four (stochastically) independent optimization processes starting from different random initial conditions and lead to rather different final configurations (presented in Fig. 2B). However, the essential features of the solutions are the same and the generation number (G) dependence of the above four quantities is also very similar. Figure 2D shows a number of frames from an imaginary movie visualizing how the ability matrix converges for growing G to its final state for a given set of initial abilities. Related movie files are included in the Appendix.

4.1.3. Continuous vs. two-valued

In Fig. 3 we display results obtained from an optimization of the ability matrix where the A_{ij} values are, in the first case, arbitrary (continuous between 0 and 1), while in the alternative case, either 1 (full competence) or 0 (zero competence). In the two-valued case we expect that the trivial optimal solution is a group having 1 specialist for each sub-problem (the same member can be a specialist for more than 1 sub-problem, but we expect a single specialist per sub-problem). Such a solution would indeed be optimal if full knowledge was not too expensive and no discussion/evaluation took place. In reality this is not the case though since both the independent evaluation of an expert and the cost of hiring him/her are very high. This aspect of the problem can be accounted for by our cost function $f(A)$.

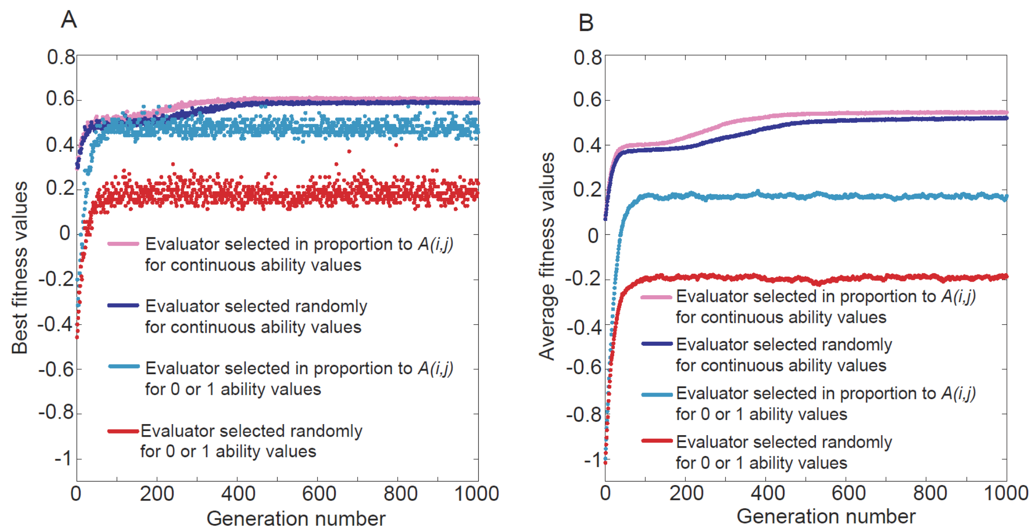

Figure 3: (A) Best and (B) average performance (fitness) values as a function of G obtained for the two fundamental variants for the possible A_{ij}-s: one allowing any values between 0 and one, while the second having only two possible (0 or 1) values. This is an important test to demonstrate that the intuitive, trivial choice for the abilities of the members, i.e., 1 corresponding to perfect competence while 0 corresponding to zero competence (with regard to a given sub-problem) results in less efficient groups. The reason is increased cost for $A_{ij} = 1$ and the inefficient discussion phase (due to the presence of the totally incompetent members). The average fitness of the 2000 groups in a generation is significantly lower than that of the best performing ones in the binary case. The continuous distribution is much less sensitive to random perturbations than the two-valued one, the best and average performances are also very similar.

Indeed, we find for realistic situations (full special knowledge is expensive and discussion can improve finding the optimal solution) that our approach results in a multiple-valued ability distribution performing better than the one constructed only from the trivial 1 and 0 abilities. In short, our formalism can be used to find the appropriate strategy (choosing between hiring top specialists or implementing longer discussions). Through adding some cost for the length of the discussion phase, even the optimal discussion time can be determined.

4.2. Results based on big data analysis

The above results are also exemplified by a number of studies on collaboration, especially on the creative groups formed by scientists working on solving increasingly complex problems. At a very recent meeting [22] on interdisciplinary science it was concluded that productive interdisciplinary researchers have a deep knowledge of at least one field but also a working awareness of others. Or, in other words, during broad collaborations individuals' breadth is as important as depth of knowledge in collective decision-making. In fact, Uzzi and collaborators have shown using huge bibliographic data sets (see [22, 23]) that papers of high impact tend to be produced by larger collaborations involving a broader wealth of knowledge.

It is highly non-trivial to test our theory against observations since the quantities we use are very rarely available. Still, an analysis based on a huge database (Web of Science - WoS [24]) provides "experimental" evidence supporting our main theoretical result. Our method to find evidence supporting the prediction(s) of our approach was based on a very motivating remark by P. Ball [25].

To measure the effect of the heterogeneous ability distribution in solving a task by a group of individuals, we calculated the level of interdisciplinarity of scientific publications using the WoS database, where subject classes are assigned to each article, which in our view correspond to the different types of sub-tasks. We define the level of interdisciplinarity, $I_{\mathcal{P}}$, of a published paper by the Shannon entropy over the subject class distribution of the publications in its reference list[26]. More precisely, we collect all subject classes from the papers appearing among the references of the article ($\mathcal{S}_{\text{ref}}(\mathcal{P})$) and consider the distribution obtained, thus:

$$I_{\mathcal{P}} = -\sum_{s\in\mathcal{S}_{\text{ref}}(\mathcal{P})} p_s \ln p_s, \tag{8}$$

where p_s denotes the probability of subject class s in the set of subject classes based on the papers in the reference $\mathcal{S}_{\text{ref}}(\mathcal{P})$.

Analogously, an author's interdisciplinarity is related to the average entropy of the publications this author has:

$$I_a = \langle I_{\mathcal{P}} \rangle_{\mathcal{P}\in\mathbb{P}(a)}, \tag{9}$$

i.e., the higher entropy corresponds to a higher level of interdisciplinarity of an author. Here $\mathbb{P}(a)$ denotes the papers of author a. We use the publication entropy instead of the subject class of the author's papers, since there can be authors who publish in a small number of different journals but can be still interdisciplinary. In other words, in calculating the heterogeneity of the authors' abilities, the entropy of their publications has a higher resolution and thus it provides a more accurate description of their interdisciplinarity. Finally, each paper is considered as a task, and the level of heterogeneity in the distribution of the authors' ability is defined by the average interdisciplinarity of the authors. Here we measure the success of solving the task by the number of citations the paper receives.

First, we selected articles published in the years 1997-1999 separately (therefore we have three sets of papers) and calculated the entropy for each article. Author's entropy was restricted to the papers published by them in the years considered (1997-1999), and only authors with at most 50 papers were considered to account for valuable contributions. Then we plot the citations of the papers as the function of the average entropy of their authors, limiting the results to papers having at least 3 and at most 50 authors. We expect highly interdisciplinary publications to show the effect of receiving high attention (and citations) only after some delay (around 10 years) to a higher extent than the less interdisciplinary ones. Therefore, for each of the four years, we considered citations from a single year with a delay of 9, 10 and 11 years. Thus, we obtained nine data sets in total, describing papers being published from 3 different years and their citations calculated with 3 different time delays. We then binned papers by their entropy in bins with 0.1 width. Results are shown in Figure 4 (we ignored bins that had less than 10 papers). As the lower and upper quartiles illustrate, for papers having high average author entropy, the citation distribution prefers higher values as well, which is also supported by the inset, where all 9 curves are shown. There is a clear trend towards more successful papers as the average author entropy increases.

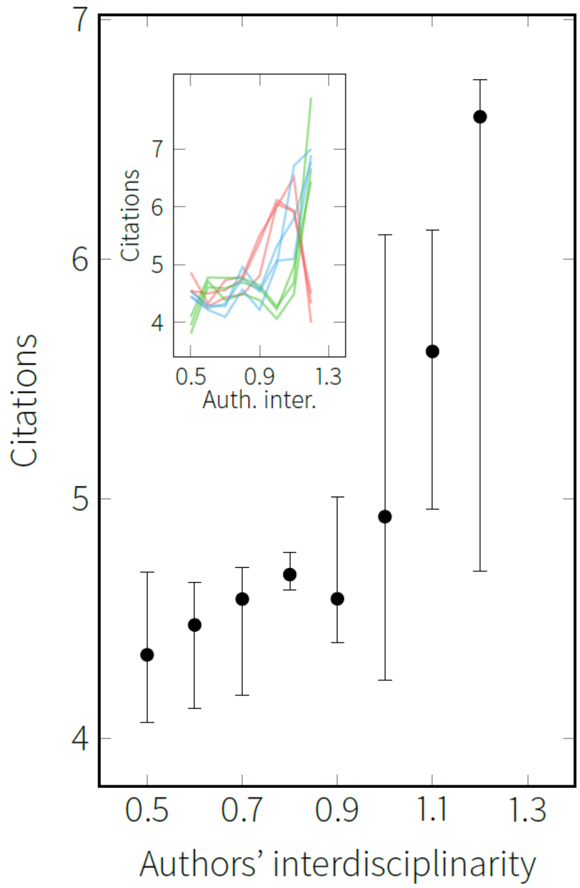

Figure 4: Relative success of papers written by collaborating interdisciplinary scientists. Median citation number of the publications as a function of the average interdisciplinarity of the corresponding authors (measured by the average entropy of each author's publications – averaged over the authors), error bars denote lower and upper quartiles. Data shows the median of nine trends (papers published between 1997 and 1999; for each year, annual citation count is calculated 9, 10 and 11 years post-publication). Single trends include only bins with at least 10 papers. Inset shows the different trends obtained by the three publication years and three citation years. Only papers with number of authors between 3 and 50 and authors with less than 50 publications have been considered.

5. Conclusions

Our formalism allows its application to more specific cases corresponding to various actual situations. It is, in some sense, the equivalent of the "division of labor" concept translated to the field of decision-making. It can be easily generalized to cases with various relative weights/influences assigned to the group members (depending, e.g., on their social status in an organization) when their assessment is considered. Additional future research could address further interesting questions such as, e.g., the optimal size of a group for a given number of sub-problems, the most reasonable time interval spent on discussions, the effect of "overlapping" problems, etc. Furthermore, the bilateral relations among the members of the group (which may be interpreted as an underlying network) can play an important role in finding the best solution. However, the main goal of our present study, instead of demonstrating particular applications, has been to provide a general framework for further quantitative estimations of essential parameters during collective decision making concerning complex problems to be solved by multidimensional groups (as far as concerning the abilities of their members).

[1] D. R. Forsyth, Effective group meetings and decision making, Working for peace: A handbook of practical psychology (2006) 88–97.

[2] J. Surowiecki, The wisdom of crowds, Anchor, 2005.

[3] S. H. Clearwater, B. A. Huberman, T. Hogg, Cooperative solution of constraint satisfaction problems, Science 254 (1991) 1181–1183.

[4] I. Planas-Sitja, J.-L. Deneubourg, C. Gibon, G. Sempo, Group personality during collective decision-making: a multi-level approach, Proceedings of the Royal Society of London B: Biological Sciences 282 (2015) 20142515.

[5] L. Conradt, C. List, Group decisions in humans and animals: a survey, Philosophical Transactions of the Royal Society of London B: Biological Sciences 364 (2009) 719–742.

[6] M. Nagy, Z. Ákos, D. Biro, T. Vicsek, Hierarchical group dynamics in pigeon flocks, Nature 464 (2010) 890–893.

[7] I. D. Couzin, C. C. Ioannou, G. Demirel, T. Gross, C. J. Torney, A. Hartnett, L. Conradt, S. A. Levin, N. E. Leonard, Uninformed individuals promote democratic consensus in animal groups, science 334 (2011) 1578–1580.

[8] T. Vicsek, Complexity: The bigger picture, Nature 418 (2002) 131–131.

[9] A. Smith, M. Garnier, An Inquiry into the Nature and Causes of the Wealth of Nations, T. Nelson, 1838.

[10] E. Durkheim, The division of labor in society, Simon and Schuster, 2014.

[11] S. E. Page, Diversity and complexity, Princeton University Press, 2010.

[12] L. Hong, S. E. Page, Problem solving by heterogeneous agents, Journal of economic theory 97 (2001) 123–163.

[13] A. Zafeiris, T. Vicsek, Group performance is maximized by hierarchical competence distribution, Nature communications 4 (2013).

[14] V. Guttal, I. D. Couzin, Social interactions, information use, and the evolution of collective migration, Proceedings of the national academy of sciences 107 (2010) 16172–16177.

[15] B. H. Hamilton, J. A. Nickerson, H. Owan, Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation, Journal of political Economy 111 (2003) 465–497.

[16] M. N. Ruderman, M. W. Hughes-James, S. E. Jackson, Selected research on work team diversity., American Psychological Association, 1996.

[17] P. Sen, B. K. Chakrabarti, Sociphysics: An Introduction, Oxford Univ. Press, 2013.

[18] S. Galam, Sociphysics: A Physicist's Modeling of Psycho-political Phenomena (Understanding Complex Systems), Springer, 2012.

[19] D. E. Goldberg, Genetic algorithms in search, optimization and machine learning addison-wesley, 1989, Reading, MA (1989).

[20] S. Kirkpatrick, M. P. Vecchi, et al., Optimization by simmulated annealing, Science 220 (1983) 671–680.

[21] L. C. Freeman, Centrality in social networks conceptual clarification, Social networks 1 (1978) 215–239.

[22] S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge, Science 316 (2007) 1036–1039.

[23] B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact, Science 342 (2013) 468–472.

[24] Isi web of knowledge, `http://scientific.thomson.com/isi/`, 2012. Accessed: 2012-01-01.

[25] P. Ball, private communication, 2015.

[26] J. Wang, B. Thijs, W. Glänzel, Interdisciplinarity and impact: distinct effects of variety, balance, and disparity, PloS one 10 (2015) e0127298.

Appendix

General presentation of the model

On a technical level our approach can be described as optimum searching on a high-dimensional highly rugged surface. Thus it has two main components: the searching mechanism and the function defining the surface.

For the optimum-seeking process we use a genetic algorithm enhanced with some simulated annealing-like features, which induce perturbations in the mutation rate to ensure the stability of the obtained result. An interesting feature of this approach is that it obtains results which are reachable and maintainable.

The function defining the surface includes the modeling of the given real-life problem-class and the estimation of the goodness of the actual evaluated parameters based on the constructed group-dynamical mechanism. This is called the fitness function.

Genetic Algorithm

An altered version of the generic evolutionary algorithm is used to find the optimum of the fitness function. The "individuals" of this evolutionary process are the groups modeled through the ability matrices (a 2d array consisting of a 1d array for each member of the group, resulting in an $N \times M$ matrix, where N is the number of members and M the number of sub-problems). Thus the population on which the evolution acts is a collection of groups (in our case the typical population size is $K = 2000$).

The twist in our approach appears when the random point mutations are applied. Usually a predefined number of abilities from the whole population are selected, and the mutation consists in randomly increasing or decreasing their values (within a mutation amplitude range). If the mutation probability is p_m, the number of mutations

$$n_m = K \cdot N \cdot M \cdot p_m. \tag{10}$$

But in our case the pm value is not entirely fixed, it can change from generation to generation. It has a predefined normal value p_{mn}, which is its starting value as well. But if the difference between the averages of the fitness values in two adjacent 100 (or 500) generations is smaller than 0.1% (or 1%), then the normal mutation rate (p_{mn}) is increased, and then annealed back to the original one (during 50 or 100 generations). This solution helps the algorithm to avoid being stuck in small local optima, and also ensures that the results acquired have high stability, and good resistance to small perturbations. Additionally the actual value used at each generation is defined by the following equation:

$$p_m = p_{mn} \cdot (1 - F), \tag{11}$$

where F is the population average of the fitness value defined in the next section.

After this step is ready, only the normalization is ahead (when it is not applied, the values of the abilities appear as cost in the fitness function): here the A matrices of each group from the emerging generation are normalized such that the

$$avg_{i,j}\left(c \cdot A^{e}_{ij}\right) = avg, \tag{12}$$

where avg is the predefined average value of the abilities.

Fitness function

A short description about a concrete realization of the fitness function is also included in the article, but the aim of this section is to present the most general form of it, underlining the generic mechanism of our approach, and also showing its relation to the concrete case analyzed in the simulations.

Flowchart of the fitness function

This function is in fact where the model of the problem-solving and solution-selection process is encoded in the whole process. The input values are the ability matrix of a given group and the return value is a real

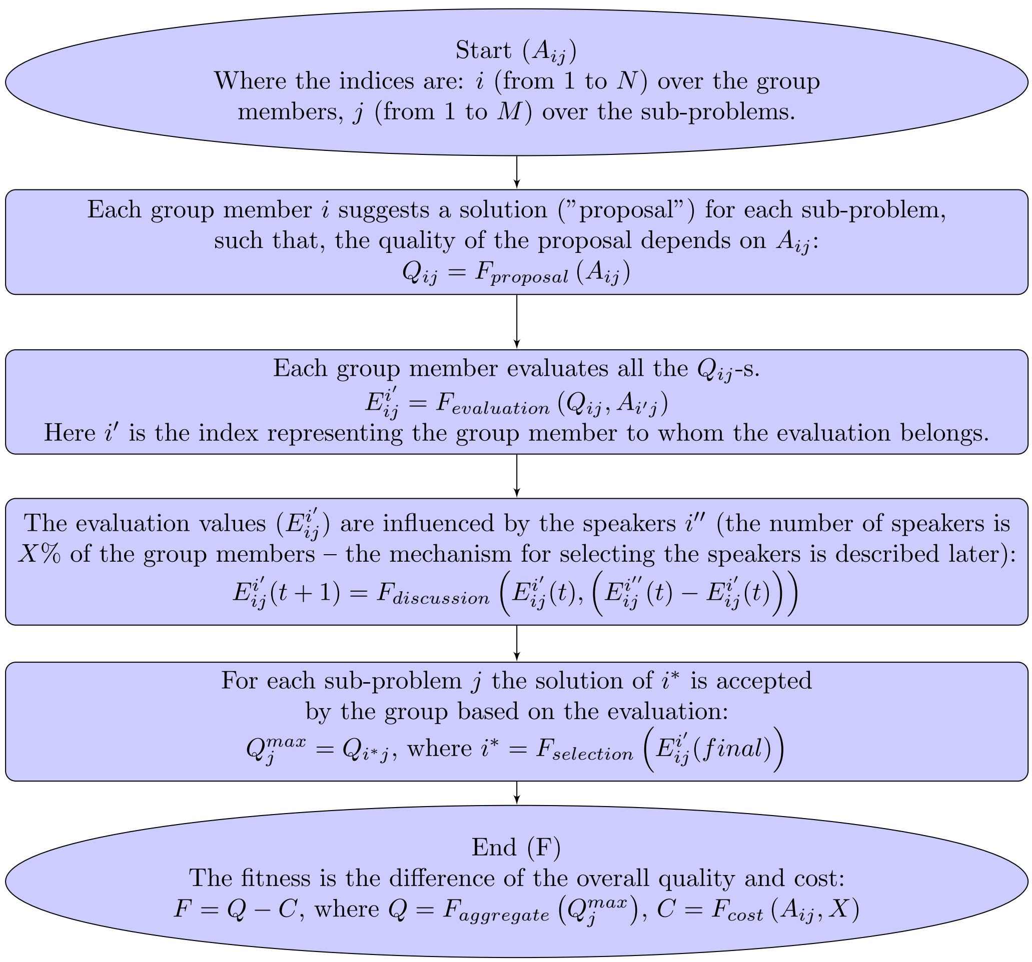

Chart 2: Flow diagram representing the generic fitness function.

number representing the "fitness" of this instance. As it can be seen in Chart 2, the function is totally described by giving the exact forms of the functions $F_{proposal}$, $F_{evaluation}$, $F_{discussion}$, $F_{selection}$, $F_{aggregate}$ and F_{cost} (all of them may include stochasticity as well).

First each member of the group proposes a solution for each sub problem; these values are proportional to the members' abilities regarding the given task. In the article we considered the simplest case, where the quality of the proposed solution (for problem j by member i, being Q_{ij}) is equal to the respective ability:

$$F_{proposal}\left(A_{ij}\right) = A_{ij}. \qquad (13)$$

The equality ensures that the small ability values (those close to zero) do not originate from the possible noise introduced at this level.

The next step follows: each member evaluates all the solutions which were given to each sub problem, in the article we use the equation described there:

$$F_{evaluation}\left(Q_{ij}, A_{i'j}\right) = Q_{ij} \cdot A_{i'j} + (1 - A_{i'j}) \cdot Rand, \qquad (14)$$

where Rand is a uniform random number from the interval $(0, 1)$.

In step c, the discussion phase, $X\%$ of the group members (i'') selected with probability proportional to their ability in the respective field share their evaluations with the others (i'), who change their own such values regarding the proposals of everybody (i) for each sub problem (j) based on this information:

$$F_{discussion}\left(E_{ij}^{i'}(t), \left(E_{ij}^{i''}(t) - E_{ij}^{i'}(t)\right)\right) = E_{ij}^{i'}(t) + \frac{1}{N}\left(E_{ij}^{i''}(t) - E_{ij}^{i'}(t)\right). \qquad (15)$$

So here the model supposes that everybody can be influenced in the same way by the current talker, and their opinions are changed so that the difference between their and the talkers opinion is reduced. (Note that the selection of the evaluators – speakers – happens proportionally to their ability values. This passage creates a situation in which the speakers are usually "experts" regarding the given sub-problem.)

Then the evaluations of the members are aggregated. In our case it simply means that for each sub problem the proposition which received the highest average evaluation is accepted as the solution of the group (here no hierarchy coefficient is included):

$$F_{selection}\left(E_{ij}^{i'}(final), H_{ij}\right) = max_i\left(sum_{i'} E_{ij}^{i'}(final)\right). \qquad (16)$$

For calculating the final return value of the fitness function in the article the most simple and intuitive aggregation function is used:

$$F_{aggregate}\left(Q_j^{max}\right) = avg_j\left(Q_j^{max}\right). \tag{17}$$

And in the simplest case (if the average ability cost does not have a pre-defined value contrarily this is just a constant change in the function values) C is simply a monotonous function of the ability values, but it could also include the time of decision making (which we assumed to be proportional to the number of talkers in the discussion phase) or other relevant parameters. The typical case of our approach uses

$$F_{cost}\left(A_{ij}, X\right) = avg_{i,j}\left(c \cdot A_{ij}^{e}\right) \tag{18}$$

(with typical values $c = 4$, $e = 4$).

Animation about the evolution of the ability matrix

The animations reachable through the links present the evolution process of the ability matrix in two different but very similar realizations of the simulation using the parameter set used in the core article as well (in the first case the ability values are represented with colors, in the second case, with colors and bars). It can be nicely seen how the specialist for each sub-problem emerges from the rest of the group, and an optimal distribution wins.

See S1 movie and S2 movie.

Comparing the results for different ability cost coefficients and exponents

In the case, when equation 18 holds, there are two independent parameters, namely the c and e constants. We present in this part the effect of modifying these values.

Firstly, we observed that the ratio of specialists in a group remained $\frac{1}{N}$ in every considered case, meaning that each sub-problem will have one specialist:

$$\frac{M \cdot N}{N} = M. \tag{19}$$

(This result was stable within 1% of error range, where the small error could appear when the two groups – specialists and the rest – could not be separated perfectly.)

Secondly, we considered the average distance between the ability of the specialists (regarding the sub-field in which they are specialists), and the knowledge of the rest of the group. (This measure is in fact a synonym of the diversity presented in the method section of the main text). The results of this inquiry are presented in figure 5.

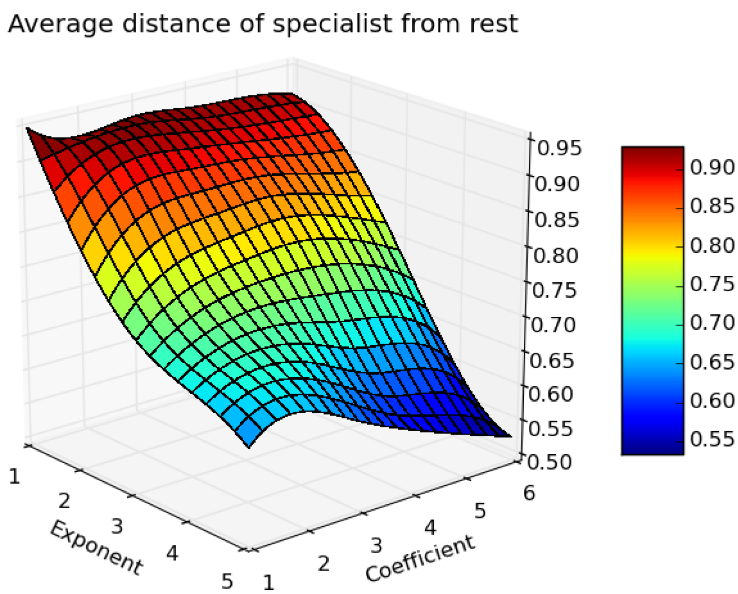

Figure 5: Average distance of a specialist's ability from the rest of the members'. For the explanation of the parameters see the text.

This surface plot makes it clear, that as the cost of outstanding knowledge increases (as the e and c values get higher, the difference between the cost of 0.5 and 1.0 ability values gets emphasized), the optimal difference between the specialist and the other members gets smaller.

Clearly this is just an example from the huge range of possible uses of the model, and unforeseeable range of its applications.

Another outcome of this approach (of changing the two parameters of the ability cost) shows the stability of the outcome, as the results in all cases are very similar, and basically the difference between them is just the mean and standard deviation of the two peaks in the ability values histogram representing the specialists and the rest.

Acknowledgments

The authors are grateful for the helpful comments by P. Ball and I. Scheuring.