COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2012) 1–19

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2011.01223.x

The Wisdom of the Crowd in Combinatorial Problems

Sheng Kung Michael Yi,[a] Mark Steyvers,[a] Michael D. Lee,[a] Matthew J. Dry[b]

[a]*Department of Cognitive Science, University of California, Irvine*
[b]*Discipline of Pharmacology, University of Adelaide*

Received 1 January 2010; received in revised form 13 June 2011; accepted 21 June 2011

Abstract

The ''wisdom of the crowd'' phenomenon refers to the finding that the aggregate of a set of proposed solutions from a group of individuals performs better than the majority of individual solutions. Most often, wisdom of the crowd effects have been investigated for problems that require single numerical estimates. We investigate whether the effect can also be observed for problems where the answer requires the coordination of multiple pieces of information. We focus on combinatorial problems such as the planar Euclidean traveling salesperson problem, minimum spanning tree problem, and a spanning tree memory task. We develop aggregation methods that combine common solution fragments into a global solution and demonstrate that these aggregate solutions outperform the majority of individual solutions. These case studies suggest that the wisdom of the crowd phenomenon might be broadly applicable to problem-solving and decision-making situations that go beyond the estimation of single numbers.

Keywords: Wisdom of the crowd; Problem solving; Traveling salesman problem; Minimum spanning tree problem

1. Introduction

When judgments are made by a group of people, the judgment obtained by aggregating their judgments is often as good as, or might even be better than, the best person in the group. This phenomenon, known as a wisdom of the crowd effect, relies on being able to sift out the noise in individual judgments to get closer to the ground truth (see Surowiecki, 2004, for an overview). The wisdom of the crowd effect has most often been demonstrated for tasks such as making continuous point estimates of physical quantities (e.g., the number

Correspondence should be sent to Mark Steyvers, University of California, Irvine, Department of Cognitive Sciences, 2316 Social & Behavioral Sciences Gateway Building, Irvine, CA 92697-5100. E-mail: mark.steyvers@uci.edu

of jelly beans in a jar) or general knowledge (e.g., the number of people living a country), or providing answers to multiple choice questions (e.g., choosing which of a set of cities is the capital of a country).

However, many practical forms of knowledge cannot be represented with a single continuous or discrete answer. An important challenge for the wisdom of the crowd research, therefore, involves its application to problems in which each answer consists of multiple elements. Recently, for example, Steyvers, Lee, Miller, and Hemmer (2009) found a wisdom of the crowds effect for ordering problems, such as listing chronologically the US Presidents, or ranking cities according to their populations. For these sorts of combinatorially challenging problems, it is not usually possible to take a mean or mode of individual answers to obtain a group answer. Instead, Steyvers et al. (2009) developed an aggregation method that provides an account of how people solve the problem and allows for the possibility of individual differences. In this way, to tackle combinatorially challenging problems, modeling the wisdom of the crowds needs input from the theories and methods of cognitive science.

In this article, we investigate the wisdom of the crowds in multidimensional problem-solving tasks from computer science and operations research known as the minimum spanning tree problem (MSTP) and traveling salesperson problem (TSP). Our goal is to develop aggregation approaches that take individual human solutions to MSTP and TSP problems and combine them into an aggregate solution. Aggregation in this domain is challenging for a number of reasons. MSTP and TSP problems are inherently high-dimensional in nature, and solutions require the coordination between many problem elements. In addition, any suitable aggregation approach that combines individual solutions needs to ensure that the aggregate solution is a valid MSTP or TSP solution obeying the task constraints. One advantage of using MSTPs and TSPs is that they have previously been studied in the experimental psychology literature. This means we already know something about the range of human performance, and the existence and nature of individual differences.

We develop two methods for combining individual human solutions to these problems, and then measure the performance of the aggregate solutions relative to the individual solutions. Our primary focus is on a method that finds the local aspects of solutions that are common across individuals. These common solution fragments are then combined into a valid global solution. We also explore a second aggregation approach that does not decompose the solution into parts. Instead, this method finds the individual solution that is most similar to other individual solutions, analogous to the computational problem of finding prototypes in the category learning literature (e.g., Estes, 1994; Nosofsky, 1992). Because this method is constrained to select the prototypical solution from the individual human solutions, it cannot identify new solutions that were not proposed by any individual. Therefore, in contrast to the first method, this method can never propose aggregate solutions that are better than any individual human solutions.

The MSTP and TSP can both be characterized as classic optimization problems. There is a specific cost function involving the total distance of a solution path that needs to be minimized. For these optimization problems, there are well-known algorithms that will give optimal (or near-optimal) solution paths. Therefore, none of our aggregation approaches can outperform the results of these optimization algorithms. Instead, the goal is to perform as closely as possible to the ideal results. In our final study, we investigate a combinatorial problem that does not have an explicit cost function. This problem involves a short-term memory task where stimuli consist of randomly generated spanning trees. The task for the subject is to reconstruct from memory the studied stimulus at a later time. Because of the absence of an explicit cost function, there is no optimal method that can be applied. However, we find that the same aggregation approaches developed for MSTPs can be used to aggregate the reconstructed memories across individuals. Collectively, therefore, we demonstrate that the wisdom of the crowd effect for combinatorial problems applies both to standard optimization problems, as well as problems in which only human judgment can be used to construct the solution.

2. The wisdom of the crowds in MSTPs and TSPs

In TSPs, a set of cities or nodes must be visited in a closed cycle that visits each node once, with the goal of minimizing the distance covered over the total tour. The TSP serves as a classic example of an NP-complete problem, where computationally scalable solution methods for guaranteed optimal solutions are not known (Applegate, Bixby, Chvátal, & Cook, 2006). As the problem size grows, optimal solution methods quickly require infeasible computational resources. Instead, to get close to optimal performance, various approximation algorithms are employed (e.g., Helsgaun, 2000, 2009). Despite the computational complexity present in TSPs, the evidence from studying human performance is that people are able to create solutions quickly while still maintaining good performance, for at least some versions of the problem. In particular, for planar Euclidean TSPs (i.e., those where the nodes can be represented as points in a two-dimensional space), people are able to complete TSPs in approximately linear time over problem sizes (Dry, Lee, Vickers, & Hughes, 2006; Graham, Joshi, & Pizlo, 2000). This contrasts with computational approaches, which have solution times that tend to scale at least on the order $O(n \ln n)$ with problem size (Applegate et al., 2006).

The solutions generated by people consistently follow some basic heuristics that promote good performance. They tend to connect nodes along the convex hull and avoid making intersections in the path (MacGregor, Chronicle, & Ormerod, 2004; MacGregor & Ormerod, 1996; van Rooij, Stege, & Schactman, 2003). There is also evidence that human solvers are sensitive to proximity between nodes, generally connecting nodes with their nearest neighbors (Vickers, Mayo, Heitmann, Lee, & Hughes, 2004). TSP solutions have even been linked to the automatic perception of minimal structures and aesthetics. When people are asked to evaluate solutions to TSPs in terms of aesthetics, the solutions that are evaluated higher tend to also be those that have shorter lengths (Vickers, Lee, Dry, Hughes, & McMahon, 2006). Earlier research by Vickers, Butavicius, Lee, and Medvedev (2001) also found similarities between solution paths created by people whose given goals were to create aesthetically pleasing circuits and paths created by subjects who performed the standard TSP task.

Despite the evidence for general principles underlying people's solutions, there is also evidence for stable and significant individual differences in human TSP performance. While early results gave conflicting accounts of the level and nature of individual differences (e.g., MacGregor & Ormerod, 1996; Vickers et al., 2001), a recent reconciliation seems to have been reached which argues for the presence of individual differences at least for sufficiently difficult problems (Chronicle, MacGregor, Lee, Ormerod, & Hughes, 2008). The prospect of individual differences in human TSP solutions makes it a potentially fruitful application for the wisdom of the crowd idea. In particular, it raises the question of whether it is possible to combine individual solutions to find a group solution that is closer to optimal than all, or the majority, of the individual solutions.

A similar combinatorial optimization task to the TSP is the MSTP. In a MSTP, a set of nodes must be linked by edges into a network such that it is possible to trace a path between any two pairs of nodes (the graph is connected), with the goal of minimizing the total length of edges placed in the network. Though there has been less empirical work on human performance on MSTPs, findings again suggest the presence of individual differences in performance on the task (Vickers et al., 2004). However, even though the MSTP is similar in description to the TSP, the optimal solution can be found using much simpler methods involving greedy algorithms (Jarník, 1930; Prim, 1957).

To demonstrate the wisdom of crowds idea for MSTPs and TSPs, we use previously collected data in which each individual independently generated a solution to a given MSTP or TSP. We propose aggregation processes that are restricted in two important ways. First, we assume that the cost function to evaluate the quality of a solution is not available until *after* the final aggregate solution is proposed. Therefore, it is not possible to refine the solution iteratively during the aggregation process to optimize the tour distance[1] or total edge length. This restriction is important because, otherwise, it would be possible to ignore the human solutions altogether and just directly optimize the tours or edges using computational means. The goal here is to see what information is collectively contained in the human solution, and the absence of the cost function during aggregation ensures that the human solutions are the only available source of information. Second, we assume that the aggregator does not have access to any spatial information, such as the location of cities or nodes. For TSPs, we assume that the only information available is the order in which the nodes are visited on the tours proposed by a group of individuals. Similarly, for the MSTPs, we assume that the only available information is which nodes are connected in human solutions. This restriction allows us to propose relatively simple aggregation procedures that analyze which nodes tend to be connected by individuals, regardless of their spatial layout.

2.1. Dataset

The data analyzed in this section were collected and reported by Burns, Lee, and Vickers (2006). A brief summary of the experiment follows, and more details can be found in the original article. As part of a larger study looking at correlations between cognitive ability and performance on optimization problems, 101 individuals completed a series of three planar Euclidean MSTPs of 30, 60, and 90 nodes and three planar Euclidean TSPs of 30, 60, and 90 nodes. Each problem was comprised of nodes placed in a square array, with each coordinate location for each node independently drawn from a uniform distribution. Individuals completed problems using a computer interface that allowed them to connect cities in any order, offering great flexibility in the strategies they could use. The optimal path and percentage length the individual's path exceeded that of the optimal were displayed after each problem to try and maintain task motivation. Fig. 1A shows solutions from 20 individuals to the 30 node MSTP.

2.2. MSTP aggregation methods

In traditional wisdom of the crowd research, the mean or mode of respondents' judgments often serves as a proposal solution to the queries being presented. For problems such as spanning trees, however, such straightforward methods are inapplicable because it is, for example, possible that no two individuals propose the same tree solution. Instead, we pursue aggregation approaches that either break down the problem into common pieces (referred to as the *local decomposition method aggregation*) or identify prototypical solutions that are globally most similar to all individual human solutions (referred to as the *global similarity aggregation method*).

The local decomposition method considers how individuals tend to connect nodes locally on their tours. We expect that good local connections between nodes tend to be selected by more individuals than those connections which are part of bad solutions. A solution that includes connections that agree more with individuals, then, should have better performance than a different solution that includes connections that have lower agreement with the group. Therefore, we propose that the spanning tree that maximizes the collective agreement across edges as a good aggregate solution.

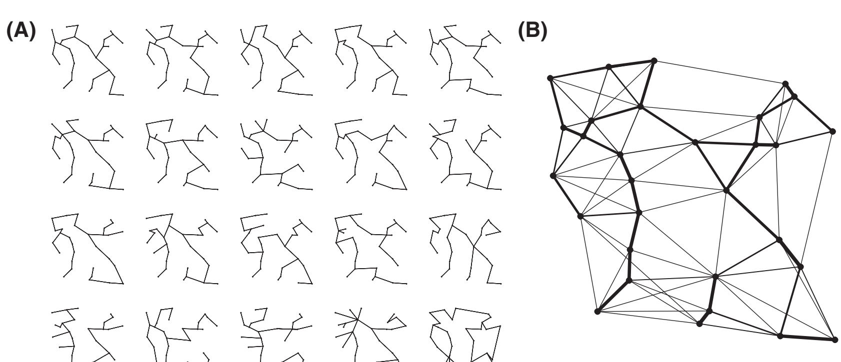

Fig. 1. Illustration of individual solutions and agreement across individuals for the 30-node MSTP. Plot (A) shows 20 of the 101 individuals' solutions ranging from the best subject on the upper left to the worst in the lower right. Plot (B) shows the degree of agreement across all 101 individuals, where each edge selected by at least one individual is drawn in, and edges selected by more individuals are drawn with thicker lines.

Specifically, we first collect all individual solutions into a $n \times n$ agreement matrix, where n is the number of nodes in the problem. Each entry a_{ij} in the matrix records the proportion of individuals that connect nodes i and j. A visualization of the agreement values for the 30-node problem is presented in Fig. 1B. These agreement matrix values are transformed into values for a cost matrix with the function $c_{ij} = 1-a_{ij}$, such that edges with higher agreement are given lower costs. We can then obtain a proposal aggregate solution by solving for the MSTP over the cost matrix, thus obtaining a spanning tree that maximizes the agreement with subject solutions. Fig. 2 illustrates the optimal solutions for three MST problems and the aggregate solution found with the local decomposition method.

The MSTP can be solved optimally in polynomial time through the use of simple greedy algorithms such as Prim's algorithm (Jarník, 1930; Prim, 1957). When edge costs are equal to Euclidean distances between nodes, the algorithm produces a network that minimizes the total length of edges. In the current context, the edge costs upon which Prim's algorithm is applied are set using the cost matrix based on individual agreement above. The algorithm will still produce a network with minimum total cost, but in this case, the network represents the spanning tree that has the highest agreement with the participant solutions. It is this solution that is generated by the aggregation method.

We also develop an alternative aggregation method based on global similarity, where the goal is to find the individual human solution that is globally most similar to the other individual solutions. We calculate similarity by the proportion of solution edges that are coincident with the solution edges placed by all other individuals. We then find the individual solution that has the highest agreement with the other individuals. This individual solution is then selected as the aggregate solution. The global aggregation method is analogous to the Kemeny–Young method used in the aggregation of rank-order data (e.g., Dwork, Kumar, Naor, & Sivakumar, 2001) where the goal is to identify rank-orderings that have the smallest summed distance to all observed rank-orderings. Note that with this aggregation strategy it is not possible to exceed the performance of the best individual for any particular problem.

2.3. Results

Performance of solutions given by individuals and the aggregate was computed in terms of percentage length above the optimal solution (PAO = 100*[empirical length/optimal

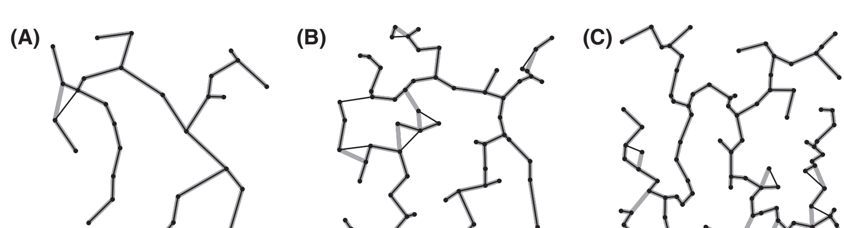

Fig. 2. Solution paths for the local decomposition aggregate method (thin black) and the optimal MST (thick gray) for the (A) 30-node, (B) 60-node, and (C) 90-node problems.

Table 1
Individual and aggregate performance on MST

	Subject Performance		Aggregation by Local Decomposition				Aggregation by Global Similarity			
Problem	Best	Mean	PAO	B	S	W	PAO	B	S	W
30 nodes	0.000%	5.672%	0.059%	1	0	100	0.288%	3	1	97
60 nodes	0.037%	6.010%	1.410%	21	0	80	1.042%	11	1	89
90 nodes	0.235%	6.533%	0.310%	1	0	100	1.029%	7	1	93
Overall	0.644%	6.072%	0.593%	0	0	101	0.786%	2	0	99

Note. MST, minimum spanning tree; PAO, percentage length above the optimal solution.

length−1]). Summary statistics for the performance of individuals and aggregation method are presented in Table 1. For each aggregation method, a count of the number of individuals whose performance is better than, same as, or worse than the aggregate is also provided, indicated by the B, S, and W columns, respectively. For individual problems, the aggregation method based on local decomposition performs much better than the average individual; in the 30- and 90-node problems, it is only outperformed by one individual. When performance is averaged over all problems, the local decomposition method leads to an aggregate solution that is closer to the optimal than any individual, as shown in Fig. 3. For the aggregation method based on global similarity, we calculated the correlation between task performance and solution agreement. There is a very strong correlation between task performance and solution agreement ($r = -.9602$), justifying the intuition behind the method that good solutions are more similar to other individual's solutions. However, performance for this global similarity method was not as good relative to the local decomposition method. Averaged over the three problems, the PAO for this method was 0.786% compared to the local decomposition method at 0.593%.

We next investigated the dependence of the wisdom of the crowd effect on the number of individuals in the aggregate, focusing on the local decomposition method. Fig. 4

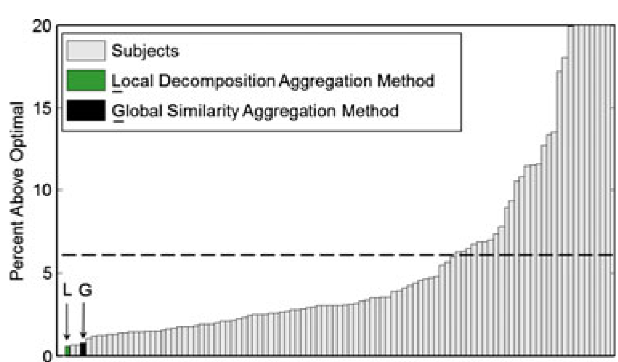

Fig. 3. Ranked task performance of participants and aggregate methods over all MSTPs. Dashed horizontal line indicates mean participant performance.

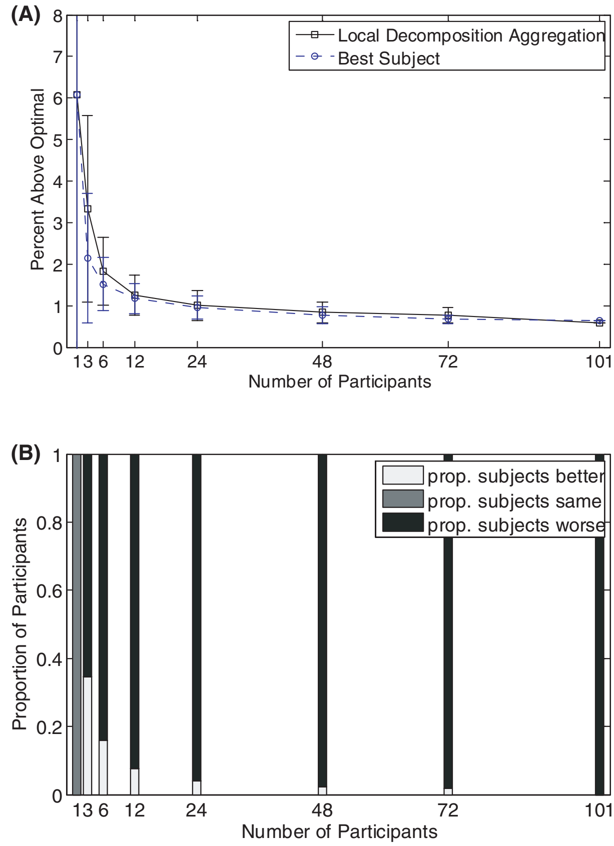

Fig. 4. Performance of the local decomposition aggregation method for MSTPs across selected sample sizes. Plot (A) shows average performance in terms of PAO. Error bars extend one standard deviation in each direction of the mean for each sample size. Dashed line shows the expected performance of the best subject taken from a sample. Plot (B) shows performance as compared to the participants being sampled to create the aggregate solution.

shows aggregate performance averaged over 1,000 random draws at sample subject sizes 1, 3, 6, 12, 24, 48, 72, and 101 (i.e., all subjects). The average performance of the aggregate quickly exceeds that of the average individual even for sample sizes as small as six. At this and larger sample sizes, performance is close to that of the best individual in the sample; on average the aggregation is only beaten by one individual at each sample size.

2.4. TSP aggregation methods

Similar to the methods used for the MSTP, we formed an aggregate proposal solution for the TSP by finding either a tour that maximizes the local agreement with individual solutions (the local decomposition method) or the overall similarity to the individual's solutions (the global similarity method). For the global similarity method, we again calculate the proportion of solution edges that are coincident with the solution edges placed by all other individuals and pick the solution that has the highest agreement.

For the local decomposition method, we made a number of changes to the aggregation method to construct a valid tour. Instead of the simple linear transformation from agreements to costs used in the MSTP, we applied a nonlinear monotonic transformation function on the agreement matrix values to transform agreements into costs for the TSP. The MSTP may be solved with greedy algorithms, so any strictly decreasing transformation function will achieve the same aggregate solution. The same does not apply for the TSP, where the increased restrictions on how tours must be constructed may result in different proposal solutions from the aggregate method depending on the cost function. Some choices of cost function may result in solutions that, when viewed in the original Euclidean problem space, are obviously suboptimal (e.g., containing crossings).

We use the function $c_{ij} = 1 - I^{-1}_{a_{ij}}(b_1, b_2)$, where $I^{-1}_{a_{ij}}(b_1, b_2)$ is the inverse regularized beta function with parameters b_1 and b_2, each taking a value of at least one.[2] A plot of our cost function for selected parameter values is shown in Fig. 5. Costs range from 0 to 1, with higher agreements leading to lower costs. When $b_1 = b_2 = 1$, we have the same linear transformation as used in the MSTP aggregation. As we increase the parameter values, the cost function becomes more nonlinear, which allows us to threshold the agreement values; values above some threshold are mapped to a relatively low cost and values below a

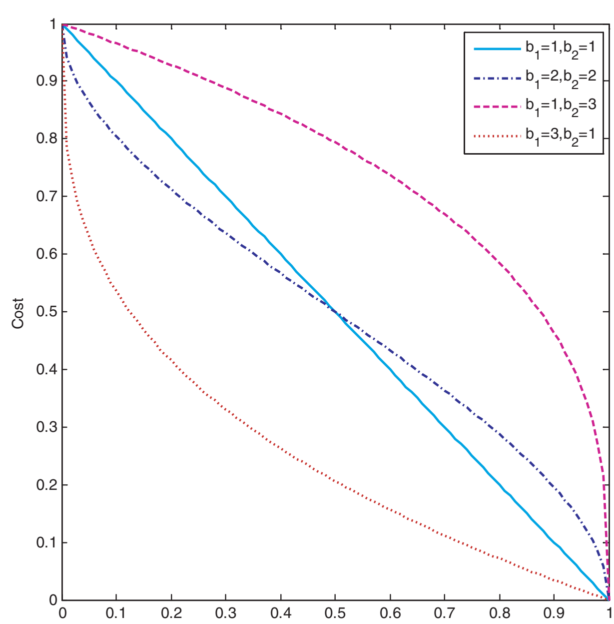

Fig. 5. Transformation functions from agreement matrix to cost matrix from the family $c_{ij} = 1 - I^{-1}_{a_{ij}}(b_1, b_2)$, where $I^{-1}_{a_{ij}}(b_1, b_2)$ is the inverse regularized beta function with parameters b_1 and b_2, for sample values of b_1 and b_2.

threshold are mapped to relatively high costs, with agreement values in between leading to an approximately linear mapping to cost. Ratios that favor b_1 emphasize the avoidance of edges with low agreement while ratios that favor b_2 emphasize the selection of the highest-agreement edges; increased values for both parameters allow both selection effects to be expressed.

We obtain our aggregate solution of the local decomposition method by solving for the TSP that minimizes the total tour cost. Because the costs can be asymmetric and do not obey the regularities of Euclidian distances, this version of the TSP cannot be solved using many traditional TSP solvers. Instead, we solve for the lowest-cost paths using the LKH program, which solves TSPs using the Lin–Kernighan heuristic (Helsgaun, 2000, 2009). While the heuristic is not guaranteed to produce the optimal solution for extremely large problems, for small problems such as those being observed in this article, the heuristic implementation is able to consistently produce the optimal solution. Examples of solutions chosen by the local decomposition aggregation method with parameter values $b_1 = 2.8$, $b_2 = 3.2$ can be seen in Fig. 6.

2.5. Results

Aggregate and individual solutions were evaluated in the same manner as for MSTPs, focusing on the PAO measure of task performance. Summary statistics of performance are presented in Table 2. Performance of the local decomposition aggregation method solutions is drastically better than most individuals, being only outperformed by two individuals in the 30-node problem and outperforming all individuals in the 60- and 90-node problems, including attainment of the optimal tour in the 90-node problem. The average performance of the local decomposition aggregation method over all three problems is better than all individuals by a large margin, as shown in Fig. 7, corresponding to an average PAO of 0.219%. The global similarity aggregate solution did not perform as well as the local decomposition method, leading to an average PAO of 2.791%, although this is still better than all individuals except one.

Performance of the local decomposition aggregation method can vary significantly depending on the parameter settings, up to 13.839% PAO on the 90-node problem and

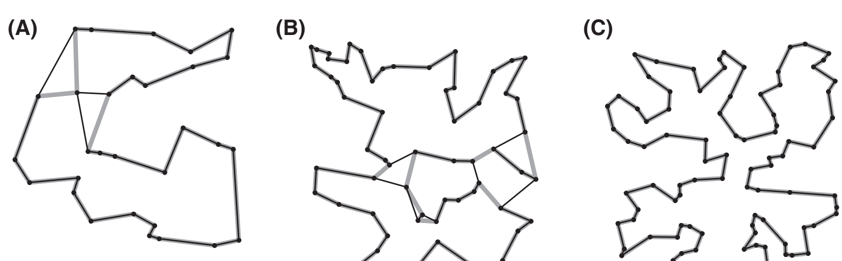

Fig. 6. Solution paths for the best-performing aggregate method parameters (thin black) and the optimal TSP (thick gray) for the (A) 30-node, (B) 60-node, and (C) 90-node problems.

Table 2
Individual and aggregate performance on TSPs

	Subject Performance		Aggregation by Local Decomposition				Aggregation by Global Similarity			
Problem	Best	Mean	PAO	B	S	W	PAO	B	S	W
30 nodes	0.000%	8.116%	0.422%	2	0	99	0.000%	0	2	99
60 nodes	0.859%	10.193%	0.234%	0	0	101	4.137%	10	1	90
90 nodes	1.404%	9.596%	0.000%	0	0	101	4.236%	11	1	89
Overall	2.386%	9.302%	0.219%	0	0	101	2.791%	1	0	100

Note. PAO, percentage length above the optimal solution; TSP, traveling salesperson problem.

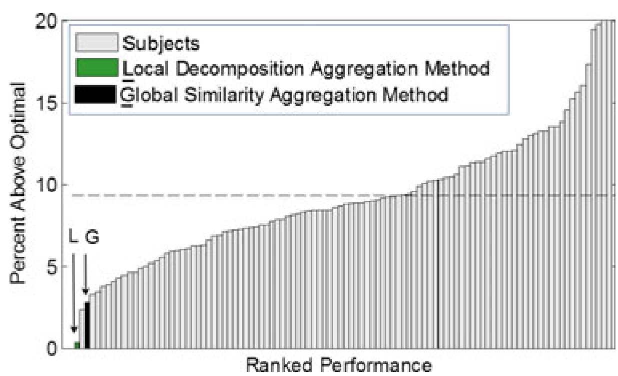

Fig. 7. Ranked task performance of participants and aggregate methods over all TSPs. Dashed horizontal line indicates mean participant performance.

5.590% PAO overall. While there is a general improvement of the aggregate as the parameter values are increased, the major factor dictating performance is the ratio between the two transformation function parameters' values. Performance is poorer when b_1 is much greater than b_2, compared to when b_2 is much greater than b_1, suggesting that it is more important to avoid the selection of edges with low or no agreement. Edges of moderate agreement should be acceptable for proposal solutions if the alternative of taking edges of higher agreement would require the addition of a low- or no-agreement edge as well. There are, however, values slightly favoring b_1 that produce the best performance for the aggregate method, and in general, attention to both high- and low-agreement edges will create good proposal solutions. We found that the best performance of the local decomposition method uses parameters $b_1 = 2.8$, $b_2 = 3.2$. These are the parameter values used for the results reported in Table 2 and Fig. 7.

Fig. 8 shows estimates of the mean performance of the local decomposition aggregation method for parameter values $b_1 = b_2 = 3$ for selected individual sample sizes. For sample sizes as small as 12 individuals, performance of the aggregation method can be expected to rival or exceed that of the best individual in both the sample taken as well as the full dataset.

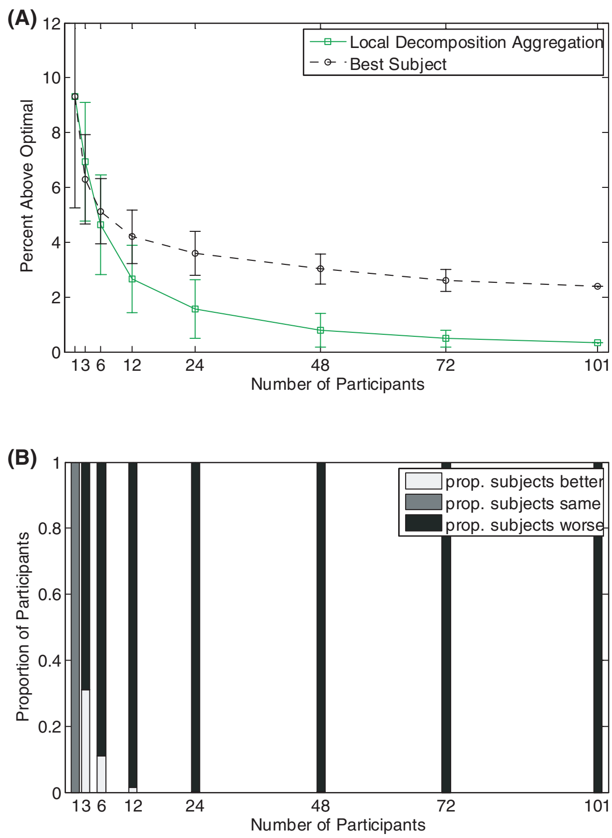

Fig. 8. Performance of the local decomposition aggregation method for TSPs with parameters $b_1 = b_2 = 3$ across selected sample sizes. Plot (A) shows average performance in terms of PAO. Error bars extend one standard deviation in each direction of the mean for each sample size. Dashed line shows the expected performance of the best subject taken from a sample. Plot (B) shows performance as compared to the participants being sampled to create the aggregate solution.

2.6. Discussion

Our results show that the aggregation methods we have developed and applied for MSTPs and TSPs are able to demonstrate a strong wisdom of the crowd effect. Solution paths proposed by the aggregation methods are created solely based on the combined node connections selected by individuals and are independent of spatial information regarding node locations. Despite the limited information available, solutions selected by the aggregation methods perform at a level that is among the best individuals on individual problems, and either exceeds the performance of the best individual when averaged over all problems (the local decomposition method) or exceeds the performance of the vast majority of individuals (the global similarity method). The finding that the local decomposition method outperforms the global similarity method suggests that it is better to identify the local aspects of problem solutions where individuals agree and then combine these into a global solution than to identify entire solutions that are most similar to other solutions.

3. Combinatorial wisdom of crowds in the absence of a cost function

Thus far, we have investigated the problem of aggregating solutions to two well-known optimization problems where there is an explicit cost function that can be evaluated. In a real-world application, however, where information about the node positions is available, one does not have to rely on an aggregation method involving human judgment to achieve good performance. Optimal (or near-optimal) solutions in these cases can be obtained by standard optimization algorithms, like those we have used to evaluate our aggregation methods. Aggregation methods and the wisdom of the crowd effect will be more useful in situations where an optimal method is unavailable and reliance on human solutions is necessary for success.

Given our demonstration of wisdom of the crowds effect in human performance on the MSTP and TSP, it seems possible or likely that there may be other combinatorially challenging problems in which an aggregation approach may be viable to find good solutions. These situations involve problems that are difficult to solve by computational means but nonetheless can be solved reasonably well, with some inherent variability in performance, by people.

As one example of such a task, we investigate the use of aggregation in a spanning tree memory task. In this task, participants are required to perform short-term recall with non-minimum spanning tree stimuli. Without knowledge of the original stimuli, obtaining good performance from an aggregation method will need to rely on the solutions given by subjects for information, as knowledge of the positions of the nodes does not necessarily provide any insight into the nature of the stimulus. If different people are able to recall different parts of the stimuli accurately, then there is potential for a more accurate picture to be created through aggregation than is achieved by any person alone.

3.1. Method

3.1.1. Participants

Thirty volunteers from the UC Irvine Social Sciences Research Participation Pool completed the spanning tree memory task and were compensated with either course credit (16 participants) or $10 for their participation.

3.1.2. Stimuli

Stimuli in the task were comprised of randomly generated 25-node spanning trees. In similar fashion to Vickers et al. (2006), constraints were placed on the node locations. Nodes were randomly generated in the unit square, constrained by number of hull nodes (8–9), mean distance between node pairs (0.50–0.55), standard deviation of distance between node pairs (0.23–0.26), and minimum distance between node pairs (0.03). To create spanning trees with properties amenable to at least partial memorization, constraints were placed on the set of edges that could be used to generate trees. A high rate of coincidence of solutions from the Burns et al. (2006) dataset with Delaunay triangulation edges (0.9916), along with previous results showing that people may perceive structure in this fashion (Dry, 2008; Dry, Navarro, Preiss, & Lee, 2009), suggests that people will be able to quickly memorize random spanning tree stimuli that are subsets of the Delaunay triangulation. While the minimum spanning tree is a subset of the Delaunay triangulation, edges in the generated stimuli are only partially coincident with that of the MSTs (mean 0.7899, range 0.625–0.917). Constraints were also placed on the path length of the generated trees. Eight problems in each of PAO constraints of 0%–5%, 5%–10%, and 10%–15% were generated for the main experimental task, with an additional two problems in each level generated as practice problems to acclimate the participant to the task. Stimuli were presented in random order in each phase of the task.

3.1.3. Procedure

The spanning tree memory task was run using a computer interface programmed with MATLAB. In each trial, a blank square axis was first presented for 2 s, followed by the presentation of the nodes for 2 s. Afterward, the spanning tree was presented for 10 s for study. The blank axis was presented for 10 s after study before the participant was given the nodes again, with the goal of recalling the edges to the best of their memory. Participants added edges to their answers by sequentially clicking between two nodes and could remove edges in the same fashion. There was no time limit on the completion of each problem, and participants were not allowed to submit an answer unless it was a complete spanning tree. After submitting each solution, participants were given feedback noting the number of edges their solution matched the actual tree, including visual feedback showing the original stimulus.

3.2. Results

Because problem solutions were in the form of spanning trees, we can apply the same aggregation methods as used in the MSTP aggregation task. The information received by the aggregation methods remains the same as well, restricted to knowledge of the edges completed by each participant on each problem.

A selection of experiment problems with the local decomposition aggregate solutions plotted against the original solutions can be found in Fig. 9. Table 3 collects summary statistics for individual and aggregate performance on the memory task. Task performance is calculated in terms of the proportion of edges placed that matched the actual stimuli. Due to the smaller number of individuals and the lower variability of their solutions, there are five problems for which the local decomposition aggregate has multiple possible solutions with the same net agreement, and evaluation measures are presented as means over these possibilities. Fig. 10 shows a ranking of performance for individuals and the aggregates over all problems, including a measure of how a participant would perform if he or she ignored or did not know the stimuli presented and instead replied with the minimum spanning tree as

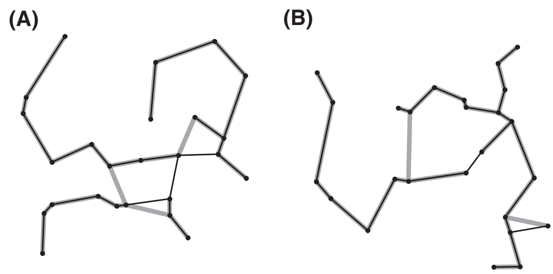

Fig. 9. Solution paths for the aggregate method (thin black) and the original spanning tree (thick gray) in the spanning tree memory task for sample problems in the (A) 0%–5% PAO, (B) 5%–10% PAO, and (C) 10%–15% PAO problem types.

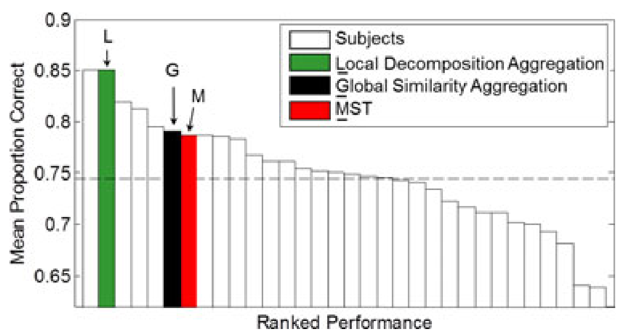

Fig. 10. Ranked task performance of participants and aggregate methods over all spanning tree memory trials. Dashed horizontal line indicates mean participant performance.

Table 3
Individual and aggregate performance on network reconstruction task

Problem	Subject Performance		Aggregation by Local Decomposition				Aggregation by Global Similarity			
	Best	Mean	Prop. Matched	B	S	W	Prop. Matched	B	S	W
0%–5% PAO	0.912	0.785	0.896	1	0	29	0.805	7	0	23
5%–10% PAO	0.828	0.726	0.841	0	0	30	0.797	2	0	28
10%–15% PAO	0.844	0.723	0.815	1	0.5	28.5	0.776	4	3	23
Overall	0.851	0.745	0.851	0.34	0.31	29.3	0.793	4	0	26

Note. PAO, percentage length above the optimal solution.

his or her solution. Despite the fact that most individual solutions are less accurate overall than an uninformed MST, aggregating over solutions is still able to provide a substantive advantage over the majority of individuals. On average, the local decomposition aggregate is more accurate than the global similarity aggregate. In addition, the local decomposition aggregate is more accurate than most individuals on each set of problem types. Compared to the best participant, the local decomposition aggregate is better on the harder (5%–10% PAO, 10%–15% PAO) problems but worse on the easier (0%–5% PAO) problems, with performance on par with the best individual averaged over all problems.

Performance of the local decomposition aggregate with smaller samples continues the trend observed with the MSTP and TSP, as shown in Fig. 11. At sample size 6, the local decomposition aggregate is able to propose solutions that significantly improve upon the average individual. As with the full dataset, only the best individual in the full group is able to outperform the aggregate consistently.

3.3. Discussion

A wisdom of the crowd effect similar to that found for the MSTP and TSP datasets was observed for the spanning tree memory experiment. Using simple aggregation methods,

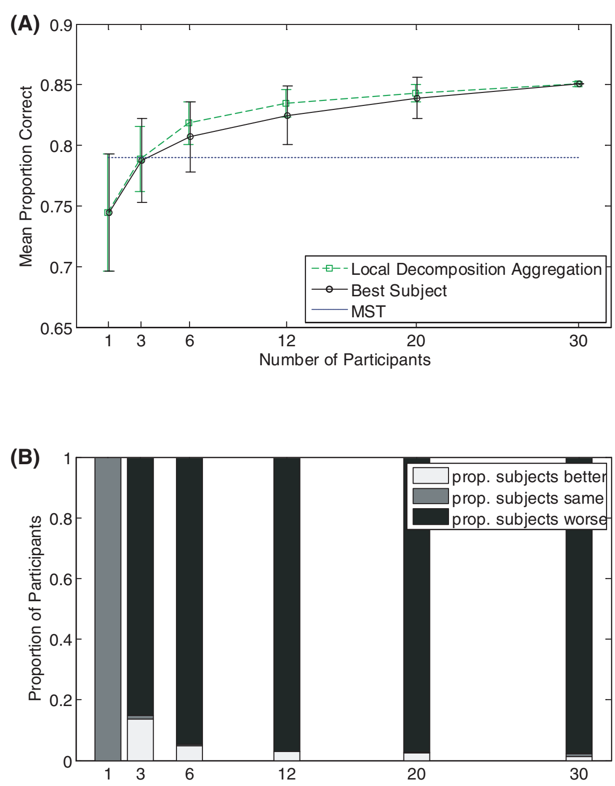

Fig. 11. Performance of the local decomposition aggregation method for spanning tree memory trials across selected sample sizes. Plot (A) shows average performance in terms of PAO. Error bars extend one standard deviation in each direction of the mean for each sample size. Dashed line shows the expected performance of the best subject taken from a sample. Plot (B) shows performance as compared to the participants being sampled to create the aggregate solution.

proposed solutions were obtained in a scenario where there is no optimal solution method. Instead, people's solutions were used to reconstruct the original stimulus. Both aggregation methods, but the local decomposition method in particular, are able to produce solutions that are significantly more accurate than the average person, and performing at approximately the same level as the best person.

4. General discussion

Most previous research in the wisdom of the crowds has focused on the situation in which responses take the form of single numeric estimates or multiple choice selections. In this article, we have demonstrated a wisdom of crowds effect for two combinatorial optimization problems and a short-term memory task with multidimensional stimuli. We have developed aggregation methods that either combine the common parts of individuals' solutions into a global solution or identify the solution that is globally most similar to other individual solutions. The first of these aggregation methods, in particular, based on local decomposition, is able to create solutions that are as good or better as those generated by people. Even for small numbers of available solutions, the local decomposition approach is able to break down the task in both problem types in a way that leads to good aggregation, despite their initial complexities.

These results can potentially be extended in a number of directions. One possibility is to identify the better performed individuals and increase their contribution to the aggregate solutions. As better individuals tend to have higher agreements with the solutions of others, identification of ''experts'' can continue to be done without explicit feedback from a cost function. The challenge is to infer and share this information about expertise across all the problems, in some sort of hierarchical model.

A second possibility is to consider combinatorial problems in the context of within-individual wisdom of the crowds research, also known as ''the crowd within'' (Vul & Pashler, 2008). The basic idea is to consider multiple solutions from the same person on the same optimization problem and test whether the aggregation of these repeated solutions leads to better performance. One nice methodological feature of this problem is that, unlike general knowledge questions, it is relatively easy to test a person on multiple versions of the same problem by applying distance-preserving transformations to the visual problem representation.

Most generally, we think that our demonstration of wisdom of the crowd effects for combinatorial problems shows a generality beyond single numerical estimates. The problems we investigate are inherently high-dimensional in which solutions require the coordination of many elements into a globally acceptable answer. We think that many or most real-world problems have these characteristics, and our results show that the wisdom of the crowds could have a role to play in understanding and improving group decision-making for these problems.

Notes

1. Human solvers of TSPs typically do not have access to the cost function either during problem-solving. The quality of the solution becomes known only after the individual submits the final solution.
2. The inverse regularized beta function is the inverse of the cumulative distribution function for the beta probability distribution. Both the cdf of the beta probability distribution and its inverse map the range [0,1] to [0,1] in a monotonic increasing function; the inverse is chosen for the cost function for its shape properties near the edges of the range. The qualitative properties over the range of parameter values make the function useful for investigation; there may be other similarly shaped functions that could also provide similar properties.

References

Applegate, D. L., Bixby, R. E., Chvátal, V., & Cook, W. J. (2006). *The traveling salesman problem: A computational study*. Princeton NJ: Princeton University Press.

Burns, N. R., Lee, N. D., & Vickers, D. (2006). Are individual differences in performance on perceptual and cognitive optimization problems determined by general intelligence? *Journal of Problem Solving*, *1*(1), 5–19.

Chronicle, E. P., MacGregor, J. N., Lee, M. D., Ormerod, T. C., & Hughes, P. (2008). Individual differences in optimization problem solving: Reconciling conflicting results. *Journal of Problem Solving*, *2*(1), 41–49.

Dry, M. J. (2008). Using relational structure to detect symmetry: A Voronoi tessellation based model of symmetry perception. *Acta Psychologica*, *128*, 75–90.

Dry, M., Lee, M. D., Vickers, D., & Hughes, P. (2006). Human performance on visually presented traveling salesperson problems with varying numbers of nodes. *Journal of Problem Solving*, *1*(1), 20–32.

Dry, M. J., Navarro, D. J., Preiss, K., & Lee, M. D. (2009). The perceptual organization of point constellations. In N. Taatgen, H. van Rijn, J. Nerbonne & L. Shonmaker (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1151–1156). Austin, TX: Cognitive Science Society.

Dwork, S., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods for the web. In V. Y. Shen & N. Saito (Eds.), *Proceedings of the 10th International Conference on World Wide Web (WWW '01)* (pp. 613–622). New York: ACM.

Estes, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

Graham, S. M., Joshi, A., & Pizlo, Z. (2000). The traveling salesman problem: A hierarchical model. *Memory & Cognition*, *28*(7), 1191–1204.

Haxhimusa, Y., Kropatsch, W. G., Pizlo, Z., & Ion, A. (2009). Approximative graph pyramid solution of the E-TSP. *Image and Vision Computing*, *27*(7), 887–896.

Helsgaun, K. (2000). An effective implementation of the Lin-Kernighan traveling salesman heuristic. *European Journal of Operational Research*, *126*, 103–130.

Helsgaun, K. (2009). General *k*-opt submoves for the Lin-Kernighan TSP heuristic. *Mathematical Programming Computation*, *1*, 119–163.

Jarník, V. (1930). O jistém problému minimálním. *Práce Moravské Přírodovědecké Společnosti*, *6*, 57–63.

MacGregor, J. N., Chronicle, E. P., & Ormerod, T. C. (2004). Convex hull or crossing avoidance? Solution heuristics in the traveling salesman problem. *Memory & Cognition*, *32*, 260–270.

MacGregor, J. N., & Ormerod, T. (1996). Human performance on the traveling salesman problem. *Perception & Psychophysics*, *58*(4), 527–539.

Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. Healy, S. Kosslyn & R. Shiffrin (Eds.), *From learning theory to connectionist heory: Essays in honor of William K. Estes* (pp. 149–167). Hillsdale, NJ: Lawrence Erlbaum.

Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Technical Journal*, *36*, 1389–1401.

van Rooij, I., Stege, U., & Schactman, A. (2003). Convex hull and tour crossings in the Euclidean traveling salesman problem: Implications for human performance studies. *Memory & Cognition*, *31*(2), 215–220.

Steyvers, M., Lee, M. D., Miller, B., & Hemmer, P. (2009). The wisdom of crowds in the recollection of order information. In J. Lafferty & C. Williams (Eds.), *Advances in neural information processing systems, 23* (pp. 1785–1793). Cambridge, MA: MIT Press.

Surowiecki, J. (2004). *The wisdom of crowds*. New York: W. W. Norton & Company, Inc.

Vickers, D., Butavicius, M., Lee, M., & Medvedev, A. (2001). Human performance on visually presented traveling salesman problems. *Psychological Research*, *65*, 34–45.

Vickers, D., Lee, M. D., Dry, M., Hughes, P., & McMahon, J. A. (2006). The aesthetic appeal of minimal structures: Judging the attractiveness of solutions to traveling salesperson problems. *Perception & Psychophysics*, *68*(1), 32–42.

Vickers, D., Mayo, T., Heitmann, M., Lee, M. D., & Hughes, P. (2004). Intelligence and individual differences in performance on three types of visually presented optimization problems. *Personality and Individual Differences*, *36*, 1059–1071.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*(7), 645–647.