ARTICLE

Check for updates

A numerical study on efficient jury size

Takamitsu Watanabe[1,2,3]✉

For judicial democracy, many societies adopt jury trials, where verdicts are made by a unanimous vote of, conventionally, 12 lay citizens. Here, using the majority-vote model, we show that such jury sizes achieve the best balance between the accuracy of verdicts and the time spent for unanimous decision-making. First, we identify two determinants of the efficient jury size: the opinion homogeneity in a community decreases the optimal jury size by affecting the accuracy of verdicts; the anti-conformity tendency in the community also reduces the efficient jury size by prolonging the time to reach unanimous verdicts. Moreover, we find an inverse correlation between these two determinants, which prevents over-shrinking and excessive expansion of the efficient jury size. Finally, by applying these findings into real-life settings, we narrow down the efficient jury size to 11.8 ± 3.0. Given that such a simple toy model can explain the jury sizes in the actual societies, the number of jurors may have been implicitly optimised for efficient unanimous decision-making throughout human history.

[1] International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan. [2] RIKEN Centre for Brain Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. [3] UCL Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AZ, UK. ✉email: takamitsu.watanabe@ircn.jp

Introduction

Jury trials are regarded as an embodiment of democracy in courtrooms (Ellsworth, 1989; Ellsworth and Getman, 1987), and it has been repeatedly debated whether decisions made by jurors sufficiently represent opinions of the entire community (Ellsworth, 1989; Saks and Marti, 1997). In this sense, ideally, juries may have to consist of all the community members or a sufficiently large number of them; however, in reality, it is impractical to implement such a direct or quasi-direct democracy into every trial. Instead, since the mid-twelfth century at latest, juries have tended to comprise ~12 individuals, especially when they are requested to reach unanimous verdicts (Warren, 1973). In fact, despite some studies suggesting the superiority of much smaller juries (Fay et al., 2000; Nagel and Neef, 1975), the conventional ~12-juror systems have still survived (Forsyth and Macdonnell, 2009).

Why has such a specific jury size been widely chosen? Throughout a line of psychological and sociological studies on various properties of the jury system—for example, its impartiality (Hamilton, 1978; Stephan, 1974; Young et al., 2014), consistency (Davis et al., 1976; Werner et al., 1985) and accuracy (Garrett et al., 2020; Ross et al., 2019)—some researches raised multiple societal and political events that could have shaped the current number of jurors (Ellsworth, 1989; Faust, 1959; Hans, 2008; Maccoun, 1989; Thomas and Fink, 1963). A meta-analysis of 17 experimental and observational studies on group decision-making demonstrated sociological and democratic advantages in 12-juror systems over 6-juror ones (Saks and Marti, 1997).

In the meantime, given such robustness and ubiquitousness of the juries with ~12 lay citizens (Warren, 1973; Forsyth and Macdonnell, 2009), more purely statistical reasonings may underlie this specific jury size.

Here, we searched for such an account by examining a hypothesis that juries consisting of ~12 individuals can achieve the best balance between the accuracy of jury verdicts and the time to reach unanimous decisions. Based on previous literature (Saks and Marti, 1997) and the concept of jury systems (Ellsworth, 1989; Ellsworth and Getman, 1987), the verdict accuracy was defined as how accurately the jury verdicts represented decisions that would be made by the full community members (Fig. 1a). The deliberation time—so-called consensus time (Krapivsky and Redner, 2003; Masuda, 2014)—was defined as how many voting steps were required for the jury to reach a unanimous consensus.

We estimated these two metrics using a majority-vote model with noise (Liggett, 2005; Oliveira, 1992; Tome et al., 1991) on a fully connected complete network (Fig. 1b). In this widely used model (Chen et al., 2017; Costa and de Souza, 2005; Lima, 2010, 2012; Melo et al., 2010; Vilela and Moreira, 2009; Vilela et al., 2012; Vilela and Stanley, 2018), each node represents each individual who has either of two dichotomic opinion status (here, guilty or not guilty) at every time point. Their opinions change over time: each member adopts the majority opinion in its connected neighbours (i.e., all the other members in this study) at the previous time point with a probability $1 - q$ and chooses the minority opinion with a probability q. This noise parameter q is often called a social temperature (Vilela and Stanley, 2018) or an anti-conformity index (Nowak and Sznajd-Weron, 2019) and supposedly represents the degree to which individuals do not obey the majority opinion in their acquaintances.

With this model, we simulated opinion dynamics between jurors and found that the most efficient jury size was largely affected by the opinion homogeneity and anti-conformity tendency in the original community. In addition, these two factors were inversely correlated, which was considered to prevent over-shrinking/expansion of the efficient jury size. Finally, we applied these findings into real-life networks and narrowed down the most efficient jury size to 11.8 ± 3.0.

Model

Majority-vote model. Using the majority-vote model with noise q ($0 \leq q \leq 0.5$; Fig. 1b) (Liggett, 2005; Oliveira, 1992; Tome et al., 1991), we numerically analysed opinion dynamics in which N jurors ($N \geq 2$) continued discussions until they reached a unanimous verdict. Given actual jury deliberation, we assumed that all the jurors interact with all the other members and implemented the majority-vote model on a fully connected complete graph.

At a time point t ($t = 0, 1, 2, \ldots$), each juror had a dichotomic opinion $\sigma_i(t) \cdot (\sigma_i(t) = \pm 1, \quad i = 0,1,3,\ldots N)$. The initial opinion $\sigma_i(t = 0)$ was defined by random sampling from a community with 10^5 individuals whose opinions were divided into 1 or –1 at the ratio of F_{Major} to $1 - F_{\text{Major}}$ ($0.5 < F_{\text{Major}} \leq 1$). At the following time point $t + 1$, the jurors choose the majority at the previous time point with a probability $1 - q$ or the opposite opinion with a probability q. Therefore, individual decision-making at a given time point can be partially affected by its own opinion at the previous time point.

When the opinions were equally divided at the time point t, $\sigma_i(t + 1)$ was determined at random.

We repeated this opinion update until all the opinions became the same. The time point when such unanimity was achieved was defined as the deliberation time T, and the unanimous verdict was also recorded as a jury verdict.

We conducted these calculations 10^6 times (Fig. 1a). That is, for one set of N, F_{Major} and q, we simulated 10^6 different jury trials, and consequently obtained 10^6 pairs of T and verdict accuracy score.

Deliberation time. The deliberation time T was summarised by the median of T, $\langle T \rangle_N$ because it showed a skewed distribution (Fig. 1c).

Verdict accuracy. Regarding the jury verdict, we first evaluated its accuracy by comparing it with that made by all the community members from which the jurors were selected. In other words, the verdict accuracy was defined as how accurately the jury verdicts represented the community decisions.

The community verdicts were estimated in essentially the same manner as the jury verdicts were, except for the following two points. First, we assumed that the verdicts of the community, which hypothetically consisted of 10^5 individuals, were made through full interactions between all the 10^5 individuals, not randomly selected N members. Second, the community verdicts were made based on a majority-vote rule, not on unanimity, because the opinions in the community were not likely to converge to unanimous ones. The community verdicts were defined by the majority opinions at $t = 100$.

When the jury verdict was the same as the community verdict, the verdict accuracy score was set at 1. Otherwise, it was set at 0.

Next, we calculated the average of the accuracy score $\langle \text{Accuracy} \rangle_N$ across the 10^6 jury trials. To infer effects of the jury size on the verdict accuracy, we then compared this $\langle \text{Accuracy} \rangle_N$ with the accuracy of the most basic group decision-making (i.e., verdict accuracy of 2-juror juries $\langle \text{Accuracy} \rangle_2$). That is, the jury effect on the verdict accuracy $\Delta \langle \text{Accuracy} \rangle_N$ was defined as $\langle \text{Accuracy} \rangle_N - \langle \text{Accuracy} \rangle_2$.

Jury efficiency and efficient jury size. Finally, the jury efficiency was calculated as $\Delta \langle \text{Accuracy} \rangle_N / \langle T \rangle_N$. We repeated this calculation for different N ($2 \leq N \leq 50$) and searched for the most
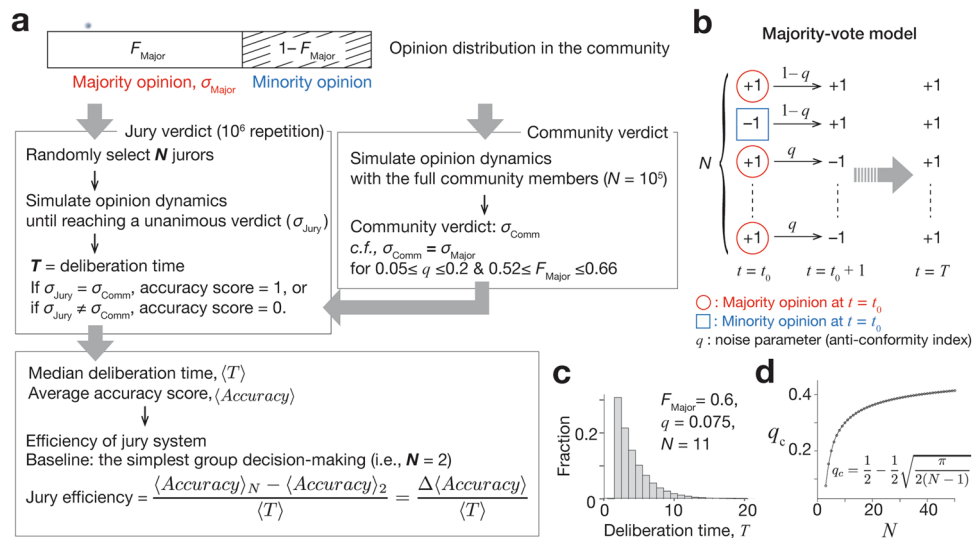
Fig. 1 Research design. a–c We evaluated the efficiency of juries with N jurors by comparing the verdict accuracy with the deliberation time (**a**) using the majority-vote model with noise on a complete graph (**b**). The verdict accuracy ⟨Accuracy⟩ was defined as whether the jury verdicts σ_{jury} was the same as the verdicts hypothetically made by the full community members σ_{Comm}. We then quantified the beneficial effect of N jurors on the verdict accuracy Δ ⟨Accuracy⟩$_N$ by calculating how better the verdict accuracy in the N-juror system ⟨Accuracy⟩$_N$ was compared to ⟨Accuracy⟩$_2$, the accuracy of the most basic collective decision-making system. The deliberation time T was summarised by its median value ⟨T⟩ because T showed a skewed distribution (**c**). Finally, we estimated the jury efficiency by calculating the ratio of Δ⟨Accuracy⟩ to ⟨T⟩. In the majority-vote model (**b**), the jurors change their opinions to the majority opinion at a time point with a probability $1-q$ or to the minor one with a probability q. If the opinion in the jury is equally divided, the jurors change their opinions randomly. The q is a noise parameter, which is also stated as social temperature (Vilela and Stanley, 2018) and anti-conformity index (Nowak and Sznajd-Weron, 2019). The F_{Major} represents the proportion of the major opinion in the entire community. **d**. The graph shows an association between the number of jurors N and the critical noise q_c in the majority-vote model on complete graphs (Chen et al., 2015). In the range of q was set to [0.05, 0.2] so that q is less than q_c even when the number of the jurors is relatively small (e.g., $N \sim 10$).

Table 1 Basic properties of real-life networks.

	Number of nodes	Mean degree	Mean clustering coefficient	Diameter (longest shortest path)
Collaboration network	23133	8.08	0.633	14
Email network	1005	50.89	0.399	7
Facebook network	4039	43.69	0.606	8

efficient jury size $N_{efficient}$. Technically, we differentiated the fitted polynomial curve and defined $N_{efficient}$ as N for a local maximum of the jury efficiency.

Brute-force search and ranges of parameters. Using this procedure, we conducted a brute-force search for the most efficient jury size in wide ranges of F_{Major} and q.

Regarding F_{Major}, we set its range at $0.52 \le F_{Major} \le 0.66$ because the jury efficiency is likely to matter particularly when the community opinions on trials are not so homogeneous but somewhat divided. After all, if supermajority exists in the community (i.e., $F_{Major} > 2/3$), the jury can reach a verdict representing such a dominant community opinion in a relatively short deliberation time.

Regarding q, we set its range to [0.05, 0.2] so as to increase the possibility that the jury can reach a unanimous verdict. Previous analytical studies (Chen et al., 2015; Fronczak and Fronczak, 2017) demonstrated that, in the majority-vote model on a complete network with N nodes, the critical noise value q_c can be described as $\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\pi}{2(N-1)}}$ (Fig. 1d). Given this, to enhance the tendency for the network to reach the ordered state (i.e., ferromagnetic state), we made q less than q_c even when the number of the nodes in the network is relatively small (e.g., $N \sim 10$).

Two community factors and efficient jury size. To examine effects of F_{Major} and q on $N_{efficient}$, we first conducted regression analyses. Based on the regression equations, we explored a simplified expression of $N_{efficient}$ and found that $N_{efficient}$ can be described as $\frac{\beta_{Acc}}{\beta_{Time}}$, where β_{Acc} and β_{Time} were regression coefficients. Next, we calculated Pearson's correlation coefficients r between the two components of $N_{efficient}$—β_{Acc} and β_{Time}— and the two community parameters—F_{Major} and q. The correlation coefficients were statistically compared in tests of the significance of the difference of the correlation coefficient.

Two community factors in real-life data. Finally, we examined an association between F_{Major} and q in real-life networks. The following three real-life network structures were obtained from Stanford Large Network Data set Collection (https://snap.stanford.edu/data/) (Leskovec and Sosic, 2016): collaboration network (ca-CondMat), email network (email-Eu-core) and Facebook network (ego-Facebook) (Table 1). All the network structures were used as undirected unweighted graphs.

We simulated opinion dynamics on these three networks using the same majority-vote model. First, we set q ($0.05 \le q \le 0.15$) and the initial opinion homogeneity $F_{Major\_initial}$ ($0.52 \le F_{Major\_initial} \le 0.66$). An initial opinion of each individual was randomly assigned so that the opinion bias in the entire network met the

$F_{Major\_initial}$. Then, we updated the opinions until the opinion homogeneity in the community reached a plateau $F_{Major\_converge}$. The plateau was defined as a period where the fluctuation in F_{Major} was within 0.1 over the 10 opinion updates. We repeated this procedure for different $F_{Major\_initial}$ and found that, when the q was constant, the community opinion homogeneity tended to converge to the same $F_{Major\_converge}$ regardless of $F_{Major\_initial}$. We then calculated this $F_{Major\_converge}$ for different q and examined correlations between $F_{Major\_converge}$ and q.

By applying this F_{Major}–q association to the results of the brute-force analysis, we narrowed down the efficient jury size.

Results

Brute-force search for efficient jury size. As an example, let us consider a jury trial employing 12 jurors who were randomly selected from a community with a 60/40 opinion split on the verdict for the trial ($F_{Major} = 0.6$ in Fig. 1a). Also, we set the noise parameter q in the majority-vote model at 0.075 (Fig. 1b), which is smaller than the critical noise q_c (Fig. 1d) and should increase the possibility that the jury reaches a unanimous verdict. With this setting, we simulated 10^6 different opinion dynamics and obtained 10^6 different deliberation time lengths T and unanimous jury verdicts.

The deliberation time to reach a unanimous verdict showed a skewed distribution (Fig. 1c); thus, we adopted the median of the time, 3.76, as a representative deliberation time length $\langle T \rangle_{12}$.

To estimate the verdict accuracy, we first calculated how many verdicts of jury were the same as those that would be made by the full community members (here, hypothetically 10^5 individuals). We then divided the number of the accurate verdicts by the total number of the simulated opinion dynamics (i.e., 10^6) and obtained the verdict accuracy $\langle Accuracy \rangle_{12} = 0.75$.

Next, we measured the beneficial effect of the jury systems on the verdict accuracy, $\Delta\langle Accuracy \rangle_{12}$, by comparing $\langle Accuracy \rangle_{12}$ with the accuracy of the most basic collective decision-making system (i.e., $\langle Accuracy \rangle_2 = 0.60$).

Finally, we calculated the ratio of $\Delta\langle Accuracy \rangle_{12}$ to $\langle T \rangle_{12}$ to determine the jury efficiency (here, $\Delta\langle Accuracy \rangle_{12}/\langle T \rangle_{12} = (\langle Accuracy \rangle_{12} - \langle Accuracy \rangle_2)/\langle T \rangle_{12} = 0.040$).

Based on this definition, we then searched for the most efficient jury size. In this setting of F_{Major} and q, we repeated the calculation of the jury efficiency for different jury sizes ($2 \leq N \leq 50$). Both the $\langle T \rangle$ and $\Delta\langle Accuracy \rangle$ increased along with the number of the jurors (Fig. 2a, b), and the jury efficiency showed a peak at $N_{efficient} = 12.1$ (Fig. 2c).

In the same manner, we conducted a brute-force search for the most efficient jury size by independently changing the two parameters in broader ranges ($0.52 \leq F_{Major} \leq 0.66$ and $0.05 \leq q \leq 0.2$). $N_{efficient}$ was found between 6.70 ($F_{Major} = 0.66$ and $q = 0.2$) and 18.40 ($F_{Major} = 0.52$ and $q = 0.05$) (Fig. 2d).

Simple expression of efficient jury size. Next, to infer determinants of the efficient jury size, we explored a simple expression of $N_{efficient}$.

In the above example, $\langle T \rangle$ showed an exponential increase when N increased ($R^{2*} = 0.96$; Fig. 2a), whereas $\Delta\langle Accuracy \rangle$ increased more slowly in a manner that was well fitted to a log–log linear regression model ($R^{2*} = 0.95$; Fig. 2b).

Such associations were seen in a broader parameter space ($0.52 \leq F_{Major} \leq 0.66$ and $0.05 \leq q \leq 0.2$; $R^{2*} \geq 0.92$), which enabled us to describe $\langle T \rangle$ and $\Delta\langle Accuracy \rangle$ as $\ln\langle T \rangle = \beta_{Time}N + \varepsilon_{Time}$ and $\ln\Delta\langle Accuracy \rangle = \beta_{Acc}\ln N + \varepsilon_{Acc}$, respectively. β_{Time} and β_{Acc} represent coefficients and the ε_{Time} and ε_{Acc} denote intercepts in the regression models.

Given them, the derivative of the logarithm of the jury efficiency can be described as $\frac{d}{dN}\ln(\Delta\langle Accuracy \rangle/\langle T \rangle) = \beta_{Acc}/N - \beta_{Time}$, and thus, the jury efficiency should peak when N is β_{Acc}/β_{Time}. This indication was validated by a significant correlation between β_{Acc}/β_{Time} and $N_{efficient}$ ($R^{2*} = 0.78$, coefficient of variation = 0.14; Fig. 3a).

Jury size and two community factors. Based on this simple expression of $N_{efficient}$, we examined how the efficient jury size was affected by the two community properties —F_{Major} and q. We first calculated β_{Acc} and β_{Time} in wide ranges of F_{Major} and q (Fig. 3b) and then estimated correlation coefficients between them.

β_{Acc} was negatively associated with F_{Major} ($r \leq -0.96$, $P < 10^{-3}$, $P_{Bonferroni} < 0.05$; e.g., the left panel in Fig. 3c) but not with q ($|r| < 0.14$, $P > 0.76$; e.g., the right panel in Fig. 3c). Conversely, β_{Time} was positively correlated with q ($r \geq 0.97$, $P < 10^{-3}$, $P_{Bonferroni} < 0.05$; e.g., the right panel in Fig. 3d) but not with F_{Major} ($|r| < 0.086$, $P > 0.85$; e.g., the left panel in Fig. 3d).

These results suggest that the larger opinion homogeneity in a community decreases the efficient jury size by dampening the beneficial effects of collective decision-making on the verdict accuracy, whereas the stronger anti-conformity tendency in the community also reduces $N_{efficient}$ by accelerating the increase in the deliberation time (Fig. 3e).

Why do not jury sizes shrink or expand? If the two community factors affect the efficient jury size in such seemingly independent manners, $N_{efficient}$ could overly shrink (e.g., $N = 2$ or 3) or excessively expand (e.g., $N = 100$). However, it is difficult to see such extremely small/large juries in real-life settings.

To solve this contradiction, we investigated associations between F_{Major} and q in real-life data sets. We simulated opinion dynamics on three real-life large-scale social networks (Leskovec and Sosic, 2016) and traced the changes in the opinion homogeneity F_{Major} when the anti-conformity index q was set as a constant.
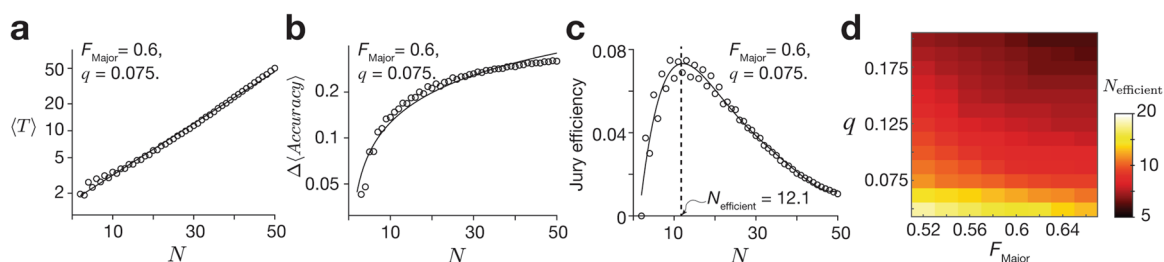
Fig. 2 Brute-force search for efficient jury size. In an example case ($F_{Major} = 0.6$ and $q = 0.075$), the deliberation time $\langle T \rangle$ and verdict accuracy change $\Delta\langle Accuracy \rangle$ increased along with the jury size N (**a** and **b**), and the jury efficiency showed a peak at $N = 12.1$ (**c**). The y-axes in the panels **a** and **b** are in a logarithmic scale. We conducted such a search for $N_{efficient}$ in a brute-force manner in broader ranges of F_{Major} and q.
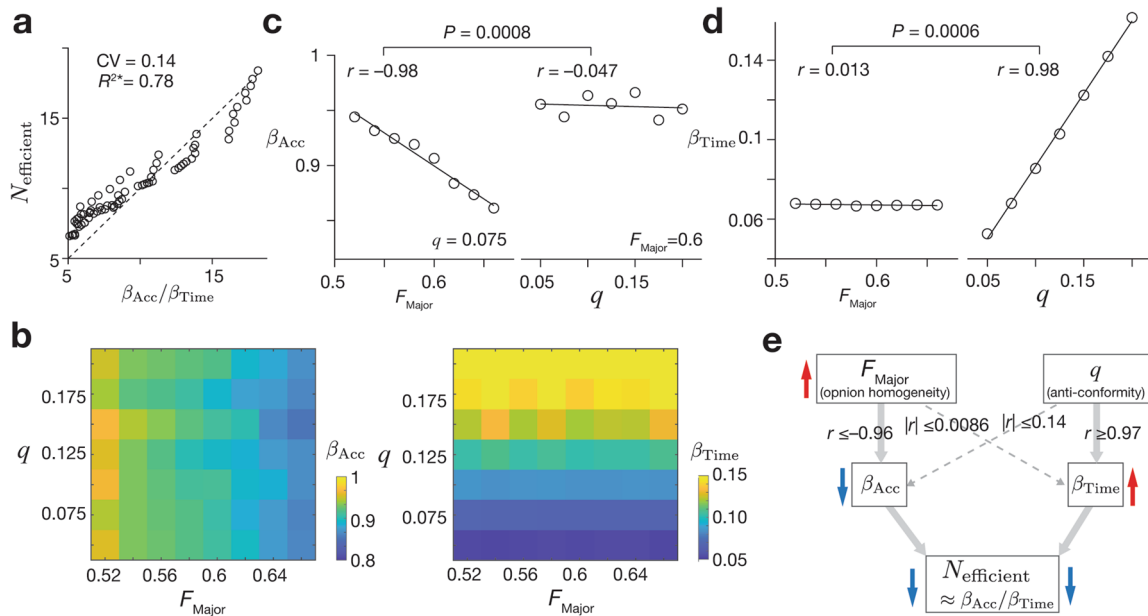
Fig. 3 Determinants of efficient jury size. a The most efficient jury size $N_{efficient}$ could be approximated by β_{Acc}/β_{Time}, where β_{Acc} and β_{Time} are regression coefficients in $\ln \Delta \langle Accuracy \rangle = \beta_{Acc} \ln N + \varepsilon_{Acc}$ and $\ln \langle T \rangle = \beta_{Time} N + \varepsilon_{Time}$. CV represents the coefficient of variation and R^{2*} denotes the adjusted coefficient of determination in a regression model using $N_{efficient} = \beta_{Acc}/\beta_{Time}$. **b–e** We calculated β_{Acc} and β_{Time} in broader ranges of F_{Major} and q (**b**) and estimated correlation coefficients between them. The panels **c** and **d** show the exemplary results of the correlation analyses. β_{Acc} was specifically correlated with F_{Major} (**c**), whereas β_{Time} was exclusively correlated with q (**d**). These observations indicate that larger F_{Major} decreases $N_{efficient}$ by reducing β_{Acc}, whereas larger q decreases $N_{efficient}$ by increasing β_{Time} (**e**). In the panels **c** and **d**, r shows a correlation coefficient and P indicates the statistical significance in a test of the significance of the difference of the correlation coefficient.
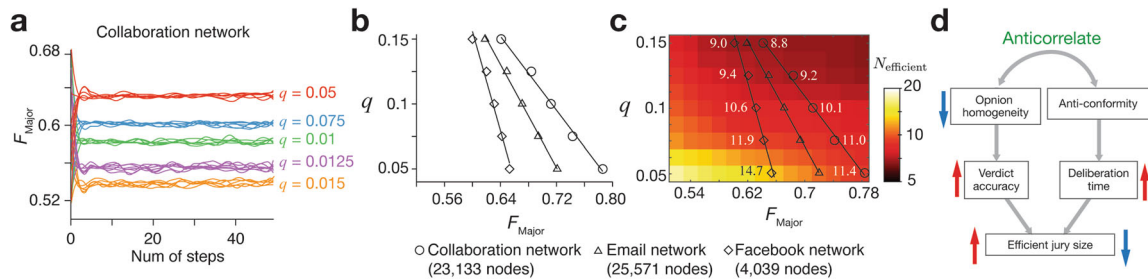
Fig. 4 Efficient jury size in real-life settings. We examined associations between the opinion homogeneity F_{Major} and anti-conformity tendency q in the three real-life large-scale social networks (Leskovec and Sosic, 2016). In all the three networks, F_{Major} converged toward a q-dependent specific value even when the initial F_{Major} was largely different (**a**). The convergence F_{Major} was negatively correlated with q (**b**). By applying such an inverse association between F_{Major} and q, we could narrow down $N_{efficient}$ into 11.8±3.0 (**c**). These results imply a statistical mechanism, which avoids over-shrinking/expansion of the jury size in the real-life society (**d**).

In all the three social networks, the opinion homogeneity converged toward a q-dependent specific value even when the initial F_{Major} was widely varied (e.g., Fig. 4a). Moreover, the convergence value of the opinion homogeneity $F_{Major\_converge}$ were negatively correlated with the anti-conformity index q ($r \geq 0.98$, $P \leq 0.0034$, $P_{Bonferroni} < 0.05$; Fig. 4b). When we consider this F_{Major}–q inverse correlation together with the results of the brute-force search for $N_{efficient}$ (Fig. 2d), the efficient jury size was narrowed down into a range from 8.8 to 14.7 (11.8 ± 3.0) in the three real-life social networks (Fig. 4c).

These results show that the inverse correlation between the opinion homogeneity and anti-conformity tendency in a community can be one of the key mechanisms that avoid the over-shrinking and excessive expansion of the jury size (Fig. 4d).

Discussion

This numerical study examined opinion dynamics during jury deliberation with the majority-vote model, calculated the

efficiency of the jury system and searched for the most efficient jury size. We first found that such an efficient jury size was determined by two community factors via different manners: the larger opinion homogeneity in a community made the efficient jury size more compact by affecting the verdict accuracy, whereas the larger anti-conformity tendency decreased the efficient jury size by accelerating the increases in the deliberation time. These two community factors were inversely correlated with each other, which prevented over-shrinking and excessive expansion of the efficient jury size. By bringing all these findings into real-life networks, the most efficient jury size was narrowed down to 11.8 ± 3.0, which is close to actual jury sizes in most of the jury trial systems (Ellsworth, 1989).

This study has demonstrated that even a simple statistical model can provide an account for the jury sizes seen in different countries; however, such simplicity could impose limitations.

Sociologically, a series of studies have suggested that the jury sizes must have been chosen and changed through a series of societal and political events (Ellsworth, 1989; Hans, 2008;

Maccoun, 1989), which were not covered in this research. We did not consider a variety of the jury trial systems, either: in some communities, their jury verdicts are not made by unanimous votes but by majority votes with different criteria (Hans, 2008); some jury systems include professional judges (Hans, 2008).

Also, the model we adopted here may be too simple in statistical and psychological contexts. The majority-vote model with noise has been used to understand various opinion dynamics (Chen et al., 2017; Costa and de Souza, 2005; Lima, 2010, 2012; Melo et al., 2010; Vilela and Moreira, 2009; Vilela et al., 2012; Vilela and Stanley, 2018), but the current form does not consider effects of some social and psychological factors, such as the existence of strong opinion holders in the juries (Vilela and Stanley, 2018) and individual differences in the anti-conformity tendency (Costa and de Souza, 2005; Vilela et al., 2012). This study assumed dichotomy in the opinion distribution; therefore, the current model took into account neither more nuanced differences in opinions (Lima, 2012; Melo et al., 2010) nor effects of confidence levels of individual opinions (Bahrami et al., 2012; Bahrami et al., 2010).

In addition to these limitations, the current results may be able to be validated analytically. A previous work on the majority-vote model on complete graphs used the heterogeneous mean-field theory and successfully identified the critical noise for both the infinite-size and finite-size networks (Chen et al., 2015). Another study adopted the master equation approach and obtained the exact equation for the probability of a given opinion pattern (Fronczak and Fronczak, 2017). Combining these analytical findings would bring us more clear, precise and expandable expressions about quantitative associations between the anti-conformity index, opinion homogeneity in the community, verdict accuracy and deliberation time.

These sociological, statistical, psychological and analytical concerns would have to be investigated in future studies; but the fact that even a simple toy model can explain the actual jury sizes may imply that the number of jurors might have been being implicitly optimised in different communities with different cultures based on a certain common mechanism.

Data availability

The data sets of the three real-life network structures are available in Stanford Large Network Data set Collection (https://snap.stanford.edu/data/). The current work generated neither original data nor novel mathematical model. The codes used here are available from the author upon reasonable requests.

References

Bahrami B, Olsen K, Bang D, Roepstorff A, Rees G, Frith C (2012) Together, slowly but surely: the role of social interaction and feedback on the build-up of benefit in collective decision-making. J Exp Psychol Hum Percept Perform 38:3–8

Bahrami B, Olsen K, Latham PE, Roepstorff A, Rees G, Frith CD (2010) Optimally interacting minds. Science 329:1081–1085

Chen H, Shen C, He G, Zhang H, Hou Z (2015) Critical noise of majority-vote model on complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 91:022816

Chen H, Shen C, Zhang H, Li G, Hou Z, Kurths J (2017) First-order phase transition in a majority-vote model with inertia. Phys Rev E 95:042304

Costa LS, de Souza AJ (2005) Continuous majority-vote model. Phys Rev E Stat Nonlin Soft Matter Phys 71:056124

Davis JH, Stasser G, Spitzer CE, Holt RW (1976) Changes in group members decision preferences during discussion-illustration with mock juries. J Pers Soc Psychol 34:1177–1187

Ellsworth PC (1989) Are twelve heads better than one? Law Contemp Probs 52:205–224

Ellsworth PC, Getman, JG (1987) Social science in legal decision-making. In: Lipson L, Wheeler S (eds) Law and the social sciences. Russell Sage Foundation, pp. 581–636

Faust WL (1959) Group versus individual problem-solving. J Abnorm Psychol 59:68–72

Fay N, Garrod S, Carletta J (2000) Group discussion as interactive dialogue or as serial monologue: the influence of group size. Psychol Sci 11:481–486

Forsyth J, Macdonnell H (2009) Scotland's unique 15-strong juries will not be abolished. In: McLellan J (eds) Johnston Press (Scotland)

Fronczak A, Fronczak P (2017) Exact solution of the isotropic majority-vote model on complete graphs. Phys Rev E 96:012304

Garrett BL, Crozier WE, Grady R (2020) Error rates, likelihood ratios, and jury evaluation of forensic evidence J Forensic Sci 65:1199–1209

Hamilton VL (1978) Obedience and Responsibility-Jury Simulation. J Pers Soc Psychol 36:126–146

Hans (2008) Jury systems around the world. Cornell Law Faculty Publications. Paper 305

Krapivsky PL, Redner S (2003) Dynamics of majority rule in two-state interacting spin systems. Phys Rev Lett 90:238701

Leskovec J, Sosic, R (2016) SNAP: a general purpose network analysis and graph mining library. ACM Trans Intell Syst Technol 8

Liggett TM (2005) Interacting particle systems. Springer, Berlin; New York, NY

Lima FWS (2010) Analysing and controlling the tax evasion dynamics via majority-vote model. J Phys: Conference Series 246

Lima FWS (2012) Three-state majority-vote model on square lattice. Physica A 391:1753–1758

Maccoun RJ (1989) Experimental research on jury decision-making. Science 244:1046–1050

Masuda N (2014) Voter model on the two-clique graph. Phys Rev E Stat Nonlin Soft Matter Phys 90:012802

Melo DFF, Pereira LFC, Moreira, FGB (2010) The phase diagram and critical behavior of the three-state majority-vote model. J Stat 2010

Nagel SS, Neef M (1975) Deductive modeling to determine an optimum jury size and fraction required to convict. Wash Univ Law Quart 1975:933

Nowak B, Sznajd-Weron K (2019) Homogeneous symmetrical threshold model with nonconformity: independence versus anticonformity. Complexity 2019:1–14

Oliveira MJD (1992) Isotropic majority-vote model on a square lattice. J Stat Phys 66:273

Ross R, Kramer K, Martire KA (2019) Consistent with: what doctors say and jurors hear. Aust J Forensic Sci 51:109–116

Saks MJ, Marti MW (1997) A Meta-analysis of the effects of jury size. Law Human Behav 21:451–467

Stephan C (1974) Sex prejudice in jury simulation. J Psychol 88:305–312

Thomas EJ, Fink CF (1963) Effects of group size. Psychol Bull 60:371–384

Tome T, Oliveira MJD, Santos MA (1991) Non-equilibrium ising model with competing Glauber dynamics. J Phys A 24:3677–3686

Vilela ALM, Moreira FGB (2009) Majority-vote model with different agents. Physica A 388:4171–4178

Vilela ALM, Moreira FGB, de Souza AJF (2012) Majority-vote model with a bimodal distribution of noises. Physica A 391:6456–6462

Vilela ALM, Stanley HE (2018) Effect of strong opinions on the dynamics of the majority-vote model. Sci Rep 8:8709

Warren WL (1973) Henry II. University of California Press, Berkeley

Werner CM, Strube MJ, Cole AM, Kagehiro DK (1985) The impact of case characteristics and prior jury experience on jury verdicts1. J Appl Soc Psychol 15:409–427

Young DM, Levinson JD, Sinnett S (2014) Innocent until primed: mock jurors' racially biased response to the presumption of innocence. PLoS ONE 9:e92365

Author contributions

TW designed this study, conducted the numerical investigations and wrote the manuscript.

Competing interests

The author declares no competing interests.

Additional information