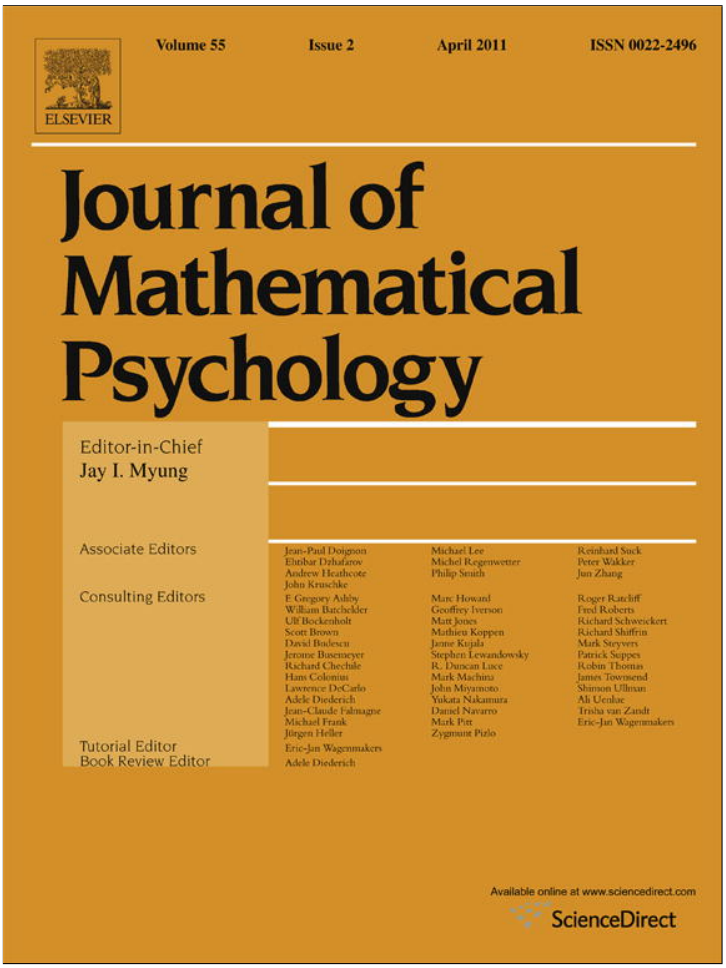

Journal of Mathematical Psychology 55 (2011) 191–197

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

The wisdom of crowds in one mind: How individuals can simulate the knowledge of diverse societies to reach better decisions

Heiko Rauhut [a,*], Jan Lorenz [b,c]

[a] ETH Zurich, Universitätstraße 41, 8092 Zurich, Switzerland

[b] ETH Zurich, Kreuzplatz 5, 8032 Zurich, Switzerland

[c] Center for Social Science Methodology, Carl von Ossietzky Universität Oldenburg, Ammerländer Heerstr. 114 - 118 26129 Oldenburg, Germany

ARTICLE INFO

Article history:
Received 11 March 2010
Received in revised form
15 October 2010
Available online 19 November 2010

ABSTRACT

The joint knowledge of many diverse individuals can outperform experts in estimation and decision-making problems. This wisdom of the crowd has been demonstrated in different societal areas such as internet search engines, political elections or stock markets. Recently, psychologists argued that humans may even simulate a diverse society in their own minds by drawing different answers from their brain (Vul & Pashler, 2008). The underlying idea is that individuals can access different knowledge areas in their brain, whose joint evaluation yields better estimates than their separate consideration. This article presents a mathematical treatment of the wisdom of crowds and two potential mechanisms to quantify the wisdom of crowds in one mind. The implications of both methods are analyzed and applied to new experimental data ($N = 144$), which contain five consecutive estimates from the same individuals. The theoretical and empirical analysis demonstrates limitations of the wisdom of crowds in one mind: Asking oneself several times is on average less powerful than asking only one other individual. This is due to the smaller diversity of estimates of similar individuals and the larger average bias to which they converge. Further, individuals cannot perform independent draws from an "internal distribution". Hence, there may be other mechanisms at work such as talking oneself into believing initial guesses or eliciting progressively wilder ones.

Under the right circumstances, social groups can be remarkably intelligent and statistical aggregates of individuals' decisions can outperform individual's and expert's decisions. Examples of this wisdom of crowds effect range from markets, auctions, political polls, internet search engines to quiz shows (Galton, 1907; Lorge, Fox, Davitz, & Brenner, 1958; Mannes, 2009; Page, 2007; Surowiecki, 2004). Recently, Herzog and Hertwig (2009) and Vul and Pashler (2008) demonstrated a wisdom of crowds effect *within one mind* by an experiment, in which individuals could respond to the same question a second time. The underlying conceptual ideas are that individual estimates are draws from an internal probability distribution (Stewart, 2009; Vul & Pashler, 2008) such that their different estimates represent answers derived from different arguments or bodies of knowledge (Herzog & Hertwig, 2009). More intuitively, one can think of this as sleeping on a decision problem.

While this evidence suggests that individuals are able to simulate a crowd in their brain, it remains open as to whether this *hypothetical* society can compete with a *real* society when individuals ask themselves ad infinitum. Previous data showed that the average of multiple estimates from the same individuals are worse than those from different individuals. This article develops two potential mechanisms for this effect. The first assumes that the joint knowledge of a single individual's brain contains on average a larger error than the joint knowledge of multiple individuals. The second assumes that individuals' estimates converge slower to the truth. We analyze the underlying assumptions of both approaches. We further demonstrate their implications by using empirical examples from new experimental data, which contain five consecutive estimates from the same individuals. The data analysis demonstrates that the first method better fits the data. In particular, this method allows one to analyze how many own guesses are equivalent to asking someone else and to approximate the gained value of asking oneself ad infinitum by extrapolating the time trend to the limit.

1. A general treatment of the wisdom of crowds

Let us consider the random variable X representing estimates either from a crowd of diverse people from which estimates are sampled at random, or the hypothetical 'crowd within one mind' (Vul & Pashler, 2008) from which an individual samples its estimates. The true value for the question is labeled 'truth'.

* Corresponding author.
E-mail addresses: rauhut@gess.ethz.ch (H. Rauhut), post@janlo.de (J. Lorenz).

0022-2496/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2010.10.002

The mean of the squared deviations of X with respect to the truth

$$\mathrm{MSE}(X) = \mathrm{E}((X - \text{truth})^2) \tag{1}$$

represents the *average individual error*.

The *population bias* is the squared deviation of the expected value of X with respect to the truth

$$\mathrm{Pop}(X) = (\mathrm{E}(X) - \text{truth})^2. \tag{2}$$

The '*diversity prediction theorem*' (Krogh & Vedelsby, 1995; Page, 2007) states that the population bias is the average individual error minus the variance,[1] or equivalently

$$\mathrm{MSE}(X) = \mathrm{Var}(X) + \mathrm{Pop}(X). \tag{3}$$

The proof is essentially using the linearity of the expected value and $\mathrm{Var}(X) = E(X^2) - E(X)^2$:

$$\begin{aligned}\mathrm{MSE}(X) &= \mathrm{E}(X^2 - 2X\text{truth} + \text{truth}^2)\\ &= \mathrm{E}(X^2) - 2\mathrm{E}(X)\text{truth} + \text{truth}^2\\ &= \mathrm{E}(X^2) - 2\mathrm{E}(X)\text{truth} + \text{truth}^2 + E(X)^2 - E(X)^2\\ &= \mathrm{E}(X^2) - \mathrm{E}(X)^2 + \mathrm{E}(X)^2 - 2\mathrm{E}(X)\text{truth} + \text{truth}^2\\ &= \mathrm{Var}(X) + (\mathrm{E}(X) - \text{truth})^2\\ &= \mathrm{Var}(X) + \mathrm{Pop}(X).\end{aligned}$$

Let us now consider a sequence of estimates X_t, where X_t has the same distribution as X. Now, we define the new random variable *belief of the crowd* of T sampled estimates as the arithmetic mean[2] of the first T estimates:

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t. \tag{4}$$

The mean of the squared deviations of the belief of the crowd $\bar{X}_T$ with respect to the truth, $\mathrm{MSE}(\bar{X}_T)$, represents the *average collective error* of a sample of T estimates.

Obviously, it holds that $\lim_{T\to\infty} \mathrm{MSE}(\bar{X}_T) = \mathrm{Pop}(X)$. Therefore, the population bias can also be called the limit collective error. Further on, it holds

$$\mathrm{Pop}(\bar{X}_T) = \mathrm{Pop}(X) \tag{5}$$

for all T. For the variance it holds

$$\mathrm{Var}(\bar{X}_T) = \frac{1}{T}\mathrm{Var}(X) \tag{6}$$

due to the Bienaymé equation $\mathrm{Var}\left(\sum_t X_t\right) = \sum_t \mathrm{Var}(X_t)$ and the fact that $\mathrm{Var}\left(\frac{1}{T}X\right) = \frac{1}{T^2}\mathrm{Var}(X)$.

Thus, by putting $\bar{X}_T$ into (3) and using (5) and (6), we see that the average collective error of T aggregated estimates follows the hyperbola

$$\mathrm{MSE}(\bar{X}_T) = \frac{\mathrm{Var}(X)}{T} + \mathrm{Pop}(X). \tag{7}$$

That means, the average collective error converges to the population bias with an increasing number of considered estimates T.[3]

[1] Page (2007) introduced the diversity prediction theorem as follows: The collective error is the average individual error minus the group's diversity. Thus, the larger the group's diversity, the smaller is the collective error compared to the average individual error. In our terminology, the collective error is the population bias and the group diversity is the variance.

[2] Other measures of aggregation than the arithmetic mean may be more appropriate to elicit the wisdom of crowd. This typically depends on the distributional form of the data.

[3] That explanation has also been given in Vul and Pashler (2008). However, it is actually not related to the central limit theorem, which is not used in the derivation of this result.

2. The wisdom of crowds in one mind

We can utilize the above described general treatment of the wisdom of crowds for the analysis of the so-called *wisdom of crowds in one mind*. Vul and Pashler (2008) argue that single individuals can simulate a "crowd" of two persons within their own brain, whose joint evaluation yields better estimates than single estimates. Let us consider $(X_i)_{i\in\mathbb{N}}$ to be a sequence of estimates from the same person. Vul and Pashler (2008) found that the average individual error of the two estimates $\mathrm{MSE}(\bar{X}_2)$ is on average smaller than each of the single errors $\mathrm{MSE}(X_1)$ and $\mathrm{MSE}(X_2)$, lending support for the notion of the wisdom of crowds in one mind and for the hypothesis that individual estimates are sampled from an internal distribution of estimates composed of different fields of knowledge.

Vul and Pashler (2008) suggested a method to quantify to what extent the crowd within one mind can compete with a crowd of different persons. The general idea is to compare the error of two averaged estimates from the same individual with the error of two averaged estimates from two randomly chosen different individuals. Their theoretical framework extends to more than two estimates from the same individual.

In the following, we present our method which is derived straightforwardly from the mathematical framework presented in the former section. Then, we describe the method of Vul and Pashler (2008) and explicate its underlying implicit assumptions more formally.

2.1. Method 1: how many more times one has to ask oneself compared to asking others

Let Y be the random variable of estimates from randomly chosen different people and X be the random variable of estimates from a single individual. Further on, let $(Y_i)_{i\in\mathbb{N}}$ and $(X_i)_{i\in\mathbb{N}}$ be the corresponding sequences of sampled values.

We define T_T^* to be *the average number of different individuals one needs to ask to achieve the same improvement than asking oneself T times*. This number need not be an integer and one should expect $T_T^* < T$. Graphically it is the projection of a data point of $\mathrm{MSE}(\bar{X}_T)$ horizontally to the hyperbola of $\mathrm{MSE}(\bar{Y}_T)$ and down to the T-axis (cf. red lines in Fig. 1 and the figure in Vul and Pashler (2008) for T_2^*). Vul and Pashler (2008) measured T_2^* empirically to be 1.11 for intermediate guesses and 1.32 for delayed guesses. This can be interpreted as "instead of asking myself twice, I need to ask only 1.11 (or 1.32 in the delayed condition) different people".

In our framework we can explicate T_T^* with the equation

$$T_T^* = \frac{\mathrm{Var}(Y)}{\mathrm{MSE}(\bar{X}_T) - \mathrm{Pop}(Y)} \tag{8}$$

$$= \frac{\mathrm{Var}(Y)}{\frac{\mathrm{Var}(X)}{T} + \mathrm{Pop}(X) - \mathrm{Pop}(Y)}. \tag{9}$$

The equation is derived from (7) for the distribution of estimates from different people $\mathrm{MSE}(\bar{Y}_T) = \frac{\mathrm{Var}(Y)}{T} + \mathrm{Pop}(Y)$ by replacing T with T_T^* and $\mathrm{MSE}(\bar{Y}_T)$ with $\mathrm{MSE}(\bar{X}_T)$, followed by solving for T_T^*. Taking $T \to \infty$ we reach

$$T_\infty^* = \frac{\mathrm{Var}(Y)}{\mathrm{Pop}(X) - \mathrm{Pop}(Y)}, \tag{10}$$

which is the *number of different individuals one needs to ask to reach the same improvement than asking oneself ad infinitum* (cf. magenta line in Fig. 1). In other words, T_∞^* measures how many estimates from different individuals one single individual can simulate on average by asking oneself.

One practical problem consists in computing the variance and the population bias for each individual, because we usually do not have enough estimates from the same person to generate a reliable estimator. We will address this problem later when we turn to the statistical analysis of our experimental data.

2.2. Method 2: how much slower do additional own estimates improve compared to asking others

Vul and Pashler (2008) invented the graphical idea on which Eq. (9) is based. However, mathematically they proposed to estimate $\lambda \in [0, 1]$ instead, which is "the proportion of an additional guess from another person that an additional guess from the same person is worth". Asking oneself T times should thus correspond to asking different people $1 + \lambda(T - 1)$ times. Formally this can only mean that $\mathrm{MSE}(\bar{X}_T)$ follows a hyperbola as[4]

$$\mathrm{MSE}(\bar{X}_T) = \frac{\mathrm{Var}(Y)}{1 + \lambda(T - 1)} + \mathrm{Pop}(Y). \tag{11}$$

The parameter λ can be computed from this equation for a given value of T, $\mathrm{MSE}(\bar{X}_T)$, $\mathrm{Var}(Y)$, and $\mathrm{Pop}(Y)$. The description of Vul and Pashler (2008) suggests the hypothesis that λ is a parameter which is independent of T.

2.3. Discussion of the two methods

There are two implicit assumptions in Eq. (11), which seem problematic in light of the statistical theory given in Eq. (7): First, the method of Vul and Pashler (2008) implies that asking oneself ad infinitum yields the same quality as asking different people in the limit of time. Second, the method implies that multiple estimates of the same individual are more diverse than those from different individuals.

The method neglects that individuals may have a different population bias they converge to than a crowd of different individuals.[5] In other words, $\mathrm{Pop}(X)$ does not need to coincide with $\mathrm{Pop}(Y)$. Although, this is not explicitly stated by Vul and Pashler (2008) it would otherwise not make sense to propose a parameter λ which is independent of T. In essence, Eq. (11) assumes that an individual's mean of squared errors converges to the same population bias as a crowd of different individuals; however, with a slower convergence. This can only happen if the variance of estimates from the same individual is larger than the variance from different individuals. An assumption which is implausible.

Method 1 resolves this implausibility by adding the individual population bias as an independent parameter. It is thus an extended version, which includes another parameter, which we think is necessary to avoid the overestimation of the wisdom of crowds effect within single individuals.

Despite our theoretical argument that the underlying assumption of method 2 is implausible, it is still an empirical question which method better predicts the reduction of average collective errors for more than two estimates. It would be possible that the population bias of the crowd within an individual is on average close to the population bias of different individuals. Furthermore, it is an empirical question whether a sequence of estimates for the same question from the same person can be seen as a sequence of independent draws from an internal distribution. These questions can only be answered empirically by new experimental data.

3. Experimental design

We conducted a laboratory experiment with 144 participants from ETH Zürich, consisting of twelve sessions with twelve subjects each. All participants were asked to provide five consecutive answers without any information about the other subjects' estimates. It was left to the subjects as to how they generated five responses to the same question and whether they chose to vary their answers or stick to the same one. Directly after all subjects gave their first response, all subjects were asked to give their second response and so forth until the fifth response. Six different estimation tasks probed their real-world knowledge, such as "What is the population density in Switzerland?" or "How many murders were registered in Switzerland in 2006?" (see Table A.1 for the full list of questions). Subjects received monetary payments taking into account the distance between estimate and true value (0%–10% (1.40 CHF), 11%–20% (0.70 CHF), 21%–40% (0.35 CHF), >40% (0 CHF)). The correct values and the achieved payments were only disclosed after completion of all five responses to avoid informed guesses of the true values. This induced truthful revelation of judgments, resembling a scoring rule (Camerer, 1995). The order of questions was randomized across sessions. This procedure delivered the sample $\tilde{x}$, consisting of 1440 raw data points ($N_t = 5$ time steps, $N_i = 48$ subjects per question, $N_q = 6$ questions) with $\tilde{x}_t^{i,q}$, denoting the t-th answer of subject i to question q.[6]

4. Statistical estimation

For calculating the wisdom of crowds with our empirical data, we aggregate the individuals' estimates to the geometric mean. This takes into account that our empirical distribution is skewed and non-Gaussian and is reflected in the fact that the geometric mean delivers results closer to the truth than the standard arithmetic mean (see Table A.1). Therefore, we normalized and transformed the raw data $x_t^{i,q} = \log \frac{\tilde{x}_t^{i,q}}{\mathrm{truth}^q}$ so that the true values correspond to zero and the arithmetic mean of x delivers the logarithm of the geometric mean of $\tilde{x}$.[7] See Fig. A.1 for visualizations of the empirical distributions of the normalized data and the logarithms of the normalized data.

We will apply our theory to the data in the following way. Assume that we are dealing with question q, then we regard Y^q to be the random variable of asking different randomly selected people the same question q. Thus, we define the random variable to take random values of the set of the first estimates of all subjects $Y^q = x_1^{i,q}$ with $i \in \{1, 2, \ldots, 48\}$. We compute the variance and the population bias for this random variable to extract the hyperbola according to Eq. (7), which will be used as a benchmark in Figs. 1 and 2.

For asking the same individual multiple times, we have 48 random variables $X^{i,q} = x_t^{i,q}$ with $t \in \{1, 2, 3, 4, 5\}$ as the time steps. The five data points of the time steps of each individual determine a random variable from which we can sample at random. Further on, we can compute the variance and population bias for these five estimates for each individual. From these values we can compute

[4] The graphical representation of projecting the empirically measured $\mathrm{MSE}(\bar{X}_2)$ to the hyperbola for $\mathrm{MSE}(\bar{Y}_T)$ underpins that our formal interpretation was implicitly assumed by Vul and Pashler (2008).

[5] "Individual population bias" may sound contradictory. What is meant is the population bias of the crowd within, thus the bias of the distribution of estimates for the same individual.

[6] Note that not all participants were presented with all questions, but with a random subset. This is why there are only 48 instead of 144 subjects. This is due to the fact that we implemented two other experimental treatments in random order, which considered information feedback (in contrast to the present investigation of the treatment without information feedback.

[7] Note that this logarithmic transformation yields a relatively low population bias reported in Figs. 1 and 2. The point here is not to demonstrate a low population bias but rather to transform the data such that it follows more closely a normal distribution. Fig. A.1 provides more details on the raw and transformed data.

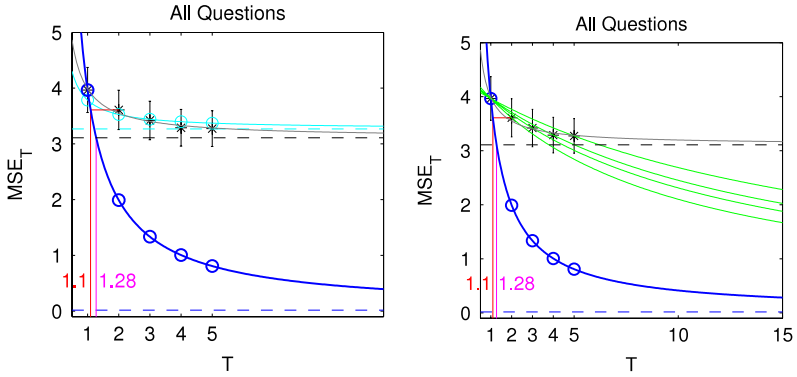

Fig. 1. The wisdom of crowds for asking repeatedly oneself compared to asking others (144 subjects, 288 responses, averaged over all questions). The x-axis denotes T answers from the same respectively different individuals. The y-axis represents the wisdom of crowds, measured by the mean of the squared errors between estimates and truth. The blue circles denote the wisdom of asking different individuals $\mathrm{MSE}(\bar{Y}_T)$. The dashed blue line represents the population bias $\mathrm{Pop}(Y)$. The black stars denote the wisdom of asking repeatedly oneself $\mathrm{MSE}(\bar{x}_T)$. The error bars are standard errors. The left panel shows method 1 to quantify the wisdom of crowd within on mind; in cyan the unordered method based on Eq. (12) and in gray the ordered method based on Eq. (13). Dashed lines indicate the corresponding average individual population bias $\overline{\mathrm{MSE}}(\bar{X}_T)$ and b. The red line illustrates the projection of T_2^*, providing a comparison between the wisdom of one to many minds: Asking oneself twice corresponds with asking 1.1 other individuals. The magenta line projects T_∞^*, the benefit of asking oneself ad infinitum, based on the ordered method. Empirically, asking oneself ad infinitum corresponds with asking 1.28 other individuals. The left panel is a demonstration of method 2 proposed by Vul and Pashler (2008), extended to 15 responses for illustrative purposes. The estimated parameter λ is computed for $\mathrm{MSE}(\bar{x}_T)$ with $T = 2, 3, 4, 5$. The estimated values of λ are further shown in Table A.3. The figure contains the four functions according to Eq. (11). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

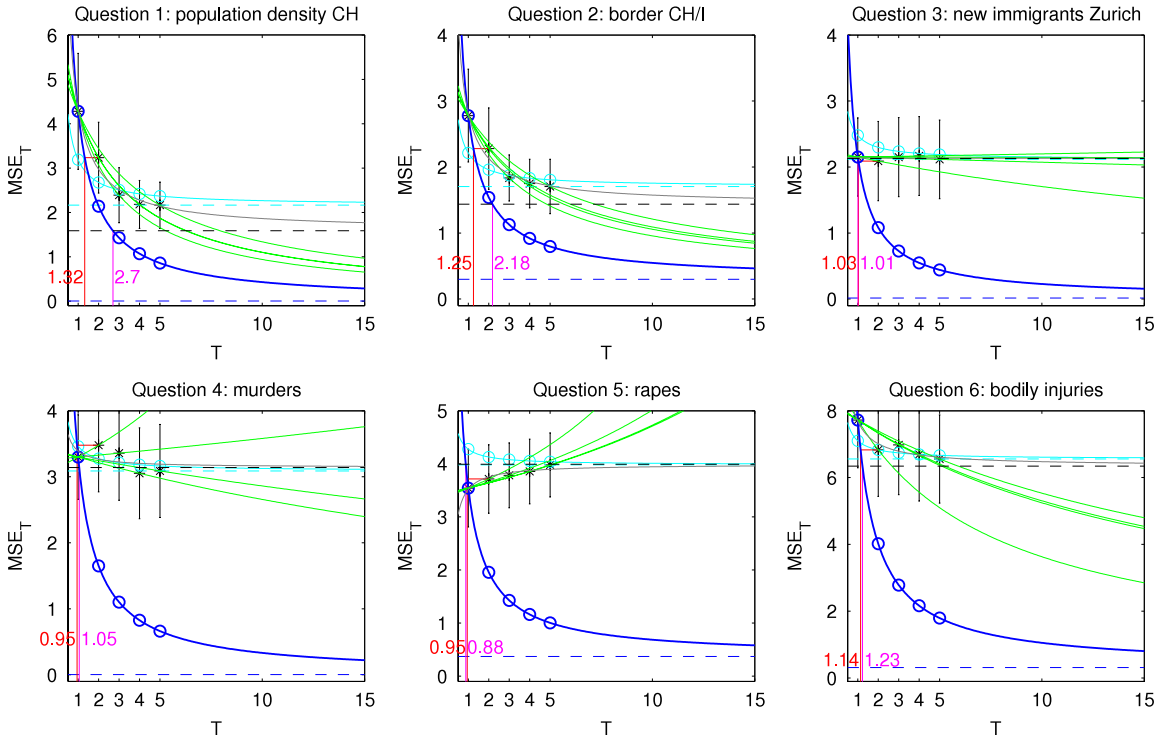

Fig. 2. The wisdom of crowds for asking repeatedly oneself compared to asking others (144 subjects, 288 responses). This figure shows the same analysis as in Fig. 1, but separately for each question. The magenta lines show that there is only a benefit of asking oneself multiple times for some questions. The green lines show that the method of Vul and Pashler (2008) typically overestimates this benefit for larger numbers of multiple estimates from the same individuals. A comparison between gray and cyan hyperbolas suggests that individuals do not sample estimates independently from a stable internal distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$\mathrm{MSE}(\bar{X}_T^{i,q}) = \frac{\mathrm{Var}(X^{i,q})}{T} + \mathrm{Pop}(X^{i,q})$ with Eq. (7). Note that Var and Pop are computed over t here.

This method neglects the order of estimates as they happened in reality and thus it treats the estimates $x_t^{i,q}$ as random draws with respect to time. This relates to the hypothesis that individual estimates are independent random samples from an internal distribution. In order to check the validity of this hypothesis, let us define the real aggregated belief of the crowd within the T'th estimate as $\bar{x}_T^{i,q} = \sum_{t=1}^{T} x_t^{i,q}$ for each i on the sample data. For these values we can compute the mean of squared errors $\mathrm{MSE}(\bar{x}_T^q)$ over all individuals i according to the definition above.

For a comparison of the crowd "within" an individual and a crowd of different individuals we need to define the *average individual bias* and the *average individual variance*. We estimate both in two different ways according to the difference of $\mathrm{MSE}(\bar{X}_T^{i,q})$ and $\mathrm{MSE}(\bar{x}_T^q)$.

We call the first method the *unordered method*. Here, it is assumed independence of estimates and a stable distribution over

Table A.1
List of all questions in the experiment, their corresponding true values and further the geometric and arithmetic means of the responses ($N = 48$ for each question except Question 2 with $N = 47$).

No.	Question	Truth	Wisdom-of-crowd aggregation	
			Geo. mean	Arith. mean
1	How high is the population density in Switzerland in inhabitants per square kilometer?	184	175	6 046
2	How long is the border between Switzerland and Italy in kilometers?	734	428	4 138
3	How much did the resident population in the city Zurich grow from 2006 to 2007?	10 067	8 947	26 927
4	How many officially registered murders took place in Switzerland in 2006?	198	190	1 250
5	How many officially registered rapes took place in Switzerland in 2006?	639	348	1 324
6	How many officially registered assaults took place in Switzerland in 2006?	9 272	16 221	356 875

Table A.2
Statistics for our proposed methods illustrated in Figs. 1 and 2. The parameters a_q and b_q and the goodness-of-fit measures SSE and R^2 have been fit with matlab's curve-fitting functions.

Question	1	2	3	4	5	6	All
$\text{Var}(Y^q)$	4.28	2.48	2.14	3.30	3.17	7.41	3.95
$\text{Pop}(Y^q)$	0.00	0.30	0.01	0.00	0.37	0.31	0.02
$\bar{\text{Var}}(X^q)$	1.03	0.51	0.36	0.37	0.30	0.54	0.52
$\bar{\text{Pop}}(X^q)$	2.16	1.71	2.12	3.09	3.98	6.56	3.27
a_q	2.77	1.39	0.01	0.25	−0.47	1.34	0.88
b_q	1.59	1.44	2.13	3.14	3.99	6.34	3.11
SSE (fit a_q, b_q)	0.10	0.03	0.00	0.10	0.01	0.07	0.01
R^2 (fit a_q, b_q)	0.97	0.96	0.01	0.20	0.89	0.91	0.98
T^*_∞ (based on b_q)	2.70	2.18	1.01	1.05	0.88	1.23	1.28
T^*_∞ (based on $\bar{\text{Pop}}(X^q)$)	1.98	1.76	1.01	1.07	0.88	1.19	1.21

time so that we compute $\bar{\text{Var}}(X^q) = \frac{1}{N}\sum_{i=1}^{N}\text{Var}(X^{i,q})$, $\bar{\text{Pop}}(X^q) = \frac{1}{N}\sum_{i=1}^{N}\text{Pop}(X^{i,q})$ and $\bar{\text{MSE}}(X^q) = \frac{1}{N}\sum_{i=1}^{N}\text{MSE}(X^{i,q})$. It is easy to see that it holds

$$\bar{\text{MSE}}(\bar{x}^q_T) = \frac{\bar{\text{Var}}(X^q)}{T} + \bar{\text{Pop}}(X^q). \tag{12}$$

The values for $\bar{\text{Var}}(X^q)$ and $\bar{\text{Pop}}(X^q)$ are given in Table A.2.

We call the second method *ordered method*. Here, the estimation of both quantities is based on real-time data $\text{MSE}(\bar{x}^q_T)$ so that we fit a_q and b_q for the hyperbola-model

$$\bar{\text{MSE}}(\bar{x}^q_T) = \frac{a_q}{T} + b_q \tag{13}$$

with a_q representing the average individual variance and b_q the average individual bias. Thus, we try to estimate parameters for a hyperbola which best fits the averages of real estimates in the order how participants elicited them. The choice of the hyperbola-model is based on the theory behind Eq. (7). Fitted data and goodness-of-fit results are shown in Table A.2.

We applied both methods on the 48 individuals answering the same question five times and also on the set of 288 responses to the six different questions. In the latter case we treat all questions equal which is possible due to the normalization of answers by their true values.

5. Empirical results

We will first report results which are averaged over all questions. We compare the results for the method of Vul and Pashler (2008) in the right panel of Fig. 1 with our proposed method in the left panel. We apply the same graphical representation as Vul and Pashler (2008) in order to compare individuals with crowds. Both methods are conducted using the two alternative statistical estimation techniques, the ordered and the unordered method. In addition to these overview figures, we conduct the graphical analyses separately for all six questions in Fig. 2. Here, we do not divide the figures in two panels but report the different methods in one subfigure.

In the following we summarize three main results we can draw from the figures and the additional data in Tables A.2 and A.3.

Result 1. *Asking oneself ad infinitum does on average not outperform asking only one other person.*

Vul and Pashler (2008) obtain the empirical results of $T^*_2 = 1.11$ in the 'immediate'-condition and $T^*_2 = 1.32$ in the '3-week delay'-condition. Thus, asking oneself once again corresponds in their case on average to asking 1.11 or respectively 1.32 other persons. Our results for the average over all questions demonstrate that asking oneself twice corresponds with asking $T^*_2 = 1.1$ other persons (red line in Fig. 1). Further, asking oneself ad infinitum corresponds with asking $T^*_\infty = 1.28$ other persons (magenta line). When we estimate according to the unordered method we achieve $T^*_\infty = 1.21$. Our findings suggest that decision-makers can indeed make use of societal effects in the sense of averaging their own estimates; however, the effect does on average not outperform asking only one other person (although Fig. 2 shows some values which are slightly higher than two for some questions).

Result 2. *The hypothesis that individuals sample independently from a "mental" distribution when they estimate several times is questionable. In particular, there is only a benefit of asking oneself several times for some questions.*

In general, the hyperbolas based on the unordered and ordered (cyan and gray hyperbolas) method of estimating the average individual variance and population bias deviate. Thus, the unordered method (cyan), which does not consider the order of estimates, does not match the data for real time steps (black stars). Moreover, the benefit of taking the average of multiple own estimates is different for different questions. While there is a benefit for some questions, e.g. questions 1 and 2, the results become even worse for other questions for an increasing number of multiple responses (e.g. question 5). This effect contradicts the idea of the wisdom of crowds in one mind. While a slow, barely visible reduction of errors could be explained with a large error variance of individuals' estimates, increasing errors with an increasing number of estimates cannot be explained with the theory of the wisdom of crowds. It is not possible to yield increasing errors with an increasing number of draws from the same population. This is evidence that other mechanisms are at work such as people talking themselves into believing their initial guesses, working themselves into emotions, or becoming more speculative over time by eliciting progressively

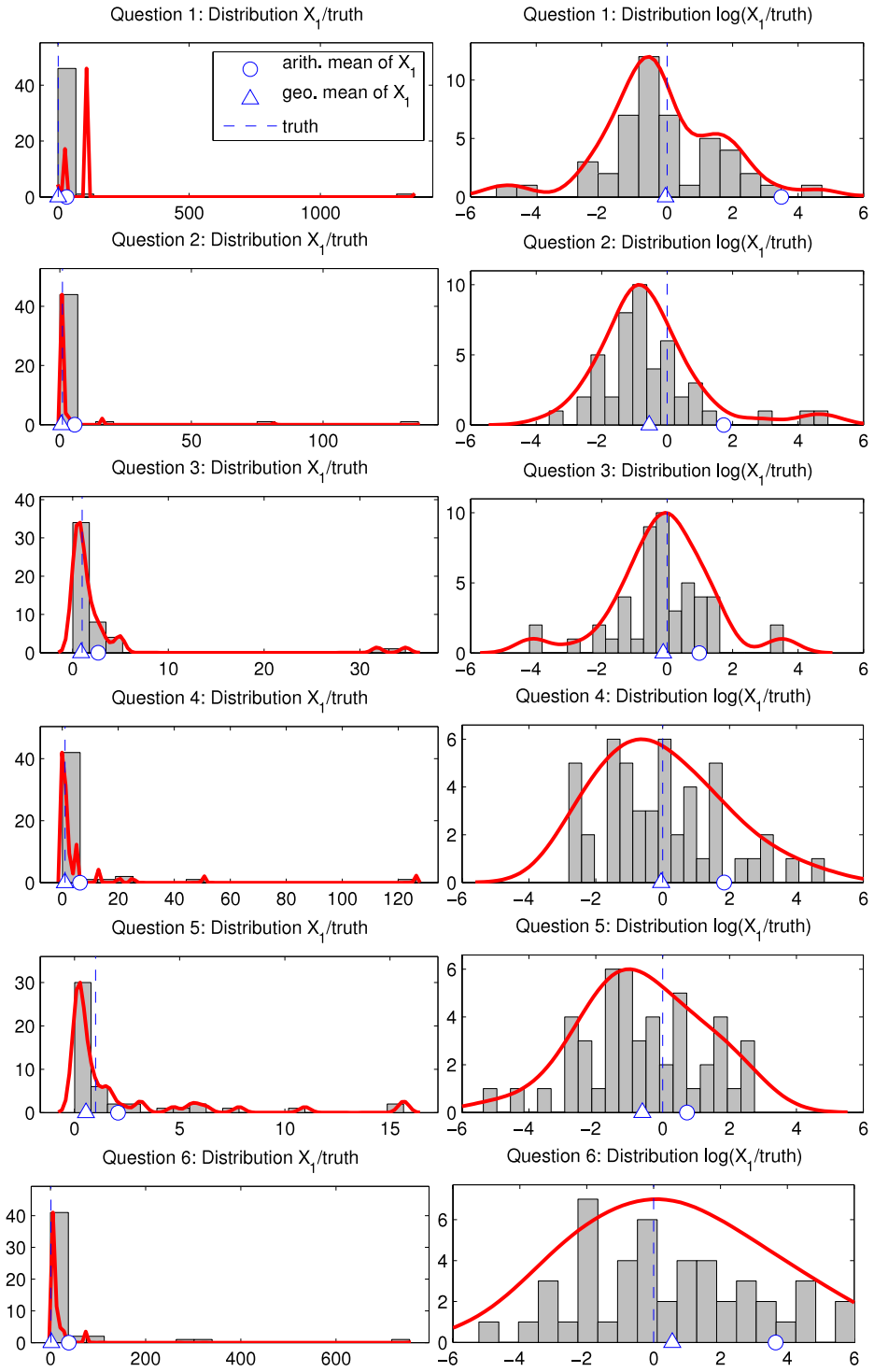

Fig. A.1. Distributions of the first estimates for each question normalized by the true value $\tilde{x}_1^q$/truth and the logarithm of it log($\tilde{x}_1^q$/truth). Gray bars belong to a histogram with 20 equidistant bins in the range. The red line is a standard KS-density plot. The legend at the top left panel holds for the whole left column. The corresponding dots in the right column are respectively the logarithms of the arithmetic mean (circle) and geometric mean (triangle). The truth is at one in the left column and at zero in the right column. (The logarithm of the geometric mean is the arithmetic mean of the logarithms of the values.) Values where normalized to make the panels comparable. For the raw values see Table A.1. ($N = 48$ for each question except question 2 with $N = 47$.)

wilder guesses. In other cases, asking oneself several times seems to add only noise (questions 3 and 4). One reason could be the difficulty of questions; for too difficult questions, individuals may not be able to draw sufficiently different responses which surround the true value. As additional information we report the population bias Pop(x_T^q) and the variance Var(x_T^q) as it changes over time for each question in Fig. A.2.

Result 3. *The suggested method by Vul and Pashler (2008) of how to calculate the benefit of asking oneself several times overestimates the effects.*

Fig. 1 and Table A.3 demonstrate further that there is no globally uniform λ and that the method of Vul and Pashler (2008) overestimates the effect of asking oneself multiple times. We demonstrate this in more detail with analyzing each question

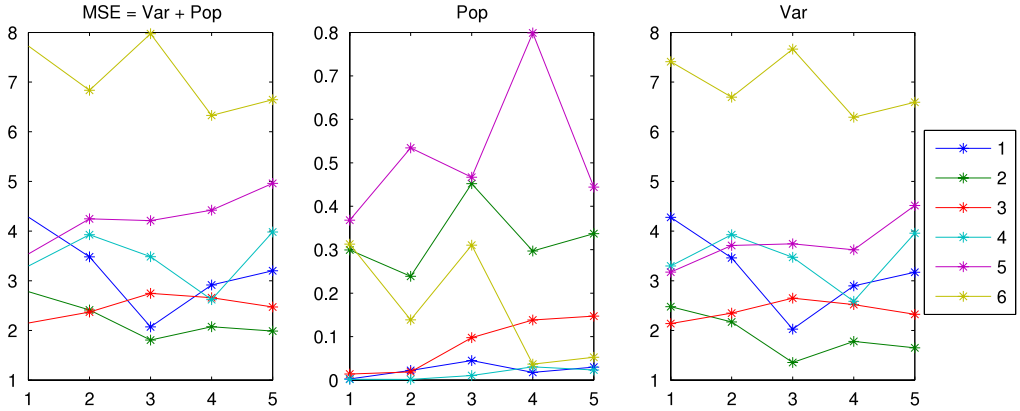

Fig. A.2. Population bias Pop(x_T^q), variance Var(x_T^q), and mean of squared errors MSE(x_T^q) for questions 1–6. Note, that MSE(x_T^q) is not the same as MSE($\bar{x}_T^q$) from Fig. 2.

Table A.3
A table of the computed values of λ according to Eq. (11) of Vul and Pashler (2008) for the data points at $T = 2, 3, 4, 5$. The functions are shown as green lines in the plots. There is clearly no globally valid λ as suggested by Vul and Pashler (2008).

Question	1	2	3	4	5	6	All
λ_2	0.32	0.25	0.03	−0.05	−0.05	0.14	0.10
λ_3	0.40	0.31	0.00	−0.01	−0.04	0.06	0.08
λ_4	0.32	0.24	−0.00	0.03	−0.03	0.05	0.07
λ_5	0.24	0.19	0.00	0.02	−0.03	0.05	0.05

separately in Fig. 2. Let us consider questions 1 and 2, for which the effect works best. As can be seen from Fig. 1 and Table A.3, the proportional effect of asking oneself exactly three times is better than Vul and Pashler's theory predicts. However, with an increasing number of multiple estimates from the same individual, the effect is worse than predicted by the method of Vul and Pashler (2008).

6. Discussion

This article confirmed the psychological notion that individuals are capable of simulating a diverse society in their own mind. Thus, individuals may apply the wisdom of crowds effect for their estimation of vaguely known facts by conducting multiple reconsiderations and averaging them. However, our empirical results demonstrate that the effect is more limited than previous analyses suggested. Even if individuals ask themselves ad infinitum, their simulated crowd returns on average worse results than if they would only ask one other person. Further, our analyses demonstrate that the previously proposed measure by Vul and Pashler (2008) of the wisdom of crowds in one mind only holds for two individual reconsiderations. This measure overestimates the power of the wisdom of crowds for multiple reconsiderations, lending more support for our claim that the proposed psychological effect is weaker than previously assumed.

Furthermore, our results suggest that the method of how to elicit multiple responses in the experiment does not change the results considerably. In our treatment, people were from the beginning well aware that they will be asked the same question five times. Further, we paid for every of the five estimates based on its closeness to the truth. This may have stimulated people to elicit different values to ensure at least some profit in the case of uncertainty. In contrast, Vul and Pashler (2008) ensured that the subjects did not know that they will be asked to respond to the same question once again. Further, in one condition, they asked for the second response three weeks later, which improved the accuracy of the average of both estimates. Herzog and Hertwig (2009) used the method 'consider the opposite', which yielded better average results than eliciting second responses without this technique. Thus, there seem to be conditions for which the accuracy of the wisdom of the crowd in one mind differs; however, the basic effect is similarly triggered by any of these mechanisms.

Our intention here is to provide an analytical framework and to stimulate subsequent analyses. Questions arise as to what extent and under which conditions individuals are able to simulate a diverse society in their mind. It would be interesting to disentangle more successful from less successful sampling strategies. For instance, individuals who can draw highly independent estimates from their own knowledge without ordered patterns or correlations between their estimates should be more successful. Further, the difficulty and the emotional extent of the estimation problems may affect individuals' capacity to apply the mechanism.

Acknowledgments

This paper benefited from comments by Dirk Helbing, Michael Mäs, Ryan Murphy, Karl-Dieter Opp, Frank Schweitzer, Mark Steyvers, Edward Vul, one anonymous reviewer and from research assistance by Hanna Thorn and Silvana Jud. We thank the ETH Zürich Competence Center 'Coping with Crises in Complex Socio-Economic Systems' (CCSS) through ETH Research Grant CH1-01-08-2 and the ETH Foundation for partial support.

Appendix. Supplementary information

See Tables A.1–A.3 and Figs. A.1 and A.2

References

Camerer, C. (1995). Individual decision making. In *The handbook of experimental economics*. Princeton University Press.
Galton, F. (1907). Vox populi. *Nature*, *75*(1949), 450–451.
Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*(2), 231–237.
Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. *Advances in Neural Information Processing Systems*, 231–238.
Lorge, I., Fox, D., Davitz, J., & Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920–1957. *Psychological Bulletin*, *55*(6), 337–372.
Mannes, A. E. (2009). Are we wise about the wisdom of crowds? The use of group judgments in belief revision. *Management Science*, *55*(8), 1267–1279.
Page, S. E. (2007). *The difference: how the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.
Stewart, N. (2009). Decision by sampling: the role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology*, *62*(6), 1041–1062.
Surowiecki, J. (2004). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday Books.
Vul, E., & Pashler, H. (2008). Measuring the crowd within: probabilistic representations within individuals. *Psychological Science*, *19*(7), 645–647.