Forecasting foreign exchange rates as group experiment: actuality bias and fact-convergence effect within wisdom of crowds

Haruo H. Horaguchi
Faculty of Business Administration, Hosei University, Tokyo, Japan

Abstract

Purpose – This article examines the accuracy and bias inherent in the wisdom of crowd effect. The purpose is to clarify what kind of bias crowds have when they make predictions. In the theoretical inquiry, the effect of the accumulated absolute deviation was simulated. In the empirical study, the observed biases were examined using data from forecasting foreign exchange rates.

Design/methodology/approach – In the theoretical inquiry, the effect of the accumulated absolute deviation was simulated based on mathematical propositions. In the empirical study, the data from 2004 to 2011 were provided by Nikkei, which holds the "Nikkei Yen Derby" competition. In total, 3,657 groups forecasted the foreign exchange rate, and the first prediction was done in early May to forecast the rate at the end of May. The second round took place in June in a similar manner.

Findings – The average absolute deviation in May was smaller than that in June. The first round of prediction was more accurate than the second round one. Predictors were affected by the observable real exchange rate, such that they modified their forecasts by referring to the actual data in early June. An actuality bias existed when the participants lost their diverse prospects. Since the standard deviations of the June forecasts were smaller than those of May, the fact-convergence effect was supported.

Originality/value – This article reports novel findings that affect the wisdom of crowd effect—referred to as actuality bias and fact-convergence effect. The former refers to a forecasting bias toward the observable rate near the forecasting date. The latter implies that predictors, as a whole, indicate smaller forecast deviations by observing the realized foreign exchange rate.

Keywords Wisdom of crowds, Social influence effect, Foreign exchange rate, Actuality bias, Fact-convergence effect

Paper type Research paper

1. Introduction

The wisdom of crowd effect is defined as an estimating effect in which the aggregate of many estimates tends to be closer to the true value than the individual estimates of the participants (Lorenz *et al.*, 2011, p. 9020). Surowiecki (2004) introduced an article by Galton (1907), who reported an attempt to estimate the weight of an ox in which 787 people participated. While the weight of the ox was 1,198 pounds, 1,207 pounds was the average of the expected values. Although the fact that estimations made by many people coincide with the true value has

Conflicts of interest: There are no conflicts of interest to declare. This manuscript has not been published or presented elsewhere in part or in entirety and is not under consideration by another journal. This is a single-authored paper.

been known for a long time (Treynor, 1987), this phenomenon is now called the "the wisdom of crowds effect," named after Surowiecki's (2004) book on collective knowledge management (Horaguchi, 2014). Experimental research and empirical studies on the wisdom of crowds effect have been conducted by social psychologists (Mannes, 2009; Kerr and Tindale, 2011; Rauhut and Lorenz, 2011), cognitive scientists (Mozer *et al.*, 2008; Steyvers *et al.*, 2009; Lee *et al.*, 2011), economists (Kremer *et al.*, 2014), and managerial scientists (Blackwell and Pickford, 2011; Cheon *et al.*, 2012; Da and Huang, 2020).

Our article reports that two novel biases are found in the wisdom of crowds effects. We refer to them as "fact-convergence effect" and "actuality bias." These two biases are defined in this article and hypothesized to verify the experimental data obtained. Our article is influenced by the research conducted by Lorenz *et al.* (2011), who examined the biases inherent in the experimental data for the wisdom of crowds effect. Lorenz *et al.* (2011) referred to this as the "social influence effect" and conducted a series of experiments to ascertain whether the social influence effect biases the wisdom of crowds effect. As we used unique data to ascertain the wisdom of crowds effect and the novel biases, we do not intend to revalidate their scientific experiments. We propose the "fact-convergence effect" and "actuality bias." These two biases are different from the "social influence effect" proposed by Lorenz *et al.* (2011).

The experiments of Lorenz *et al.* (2011) consisted of 12 groups of 12 people; they had 144 participants overall, whom they asked six questions, for example, "What is the population density of Switzerland?" The participants answered each question on the following condition: those who offered answers that had the least deviation from the true value received a reward. Then, Lorenz *et al.* (2011) divided the participants into three groups. After the first estimation, the first group was given the average of the estimates of the 144 participants, and the second group was provided with all the individual responses, although they were not told who gave each response. The third group, a control group, were not provided any information regarding the estimates already made. The results of the estimation are as follows. The first group's estimation converged to the average of the estimates of all the respondents but deviated from the true value. Lorenz *et al.* (2011) referred to this phenomenon as the "social influence effect" and inferred that obtaining social information affects people in a way that they lose the diversity of their ideas.

An inclusionary concept similar to the social influence effect is the anchor effect (Kahneman, 1992; Jacowitz and Kahneman, 1995). The anchor effect refers to the influence of data presented to subjects on their selective decision-making. When the anchor effect is discussed, proximity to the correct answer with respect to the aggregate value of the responses is not an issue. The anchor effects are discussed as in the case of choosing one of several alternatives or in experiments where one price level is used as a reference index to evaluate another price. Where the wisdom of crowds is at issue, people are observed if they can reach true value. Lorenz *et al.* (2011) required the crowd to provide specific estimates as responses to the questions. They checked whether the average of the crowd's estimates was close to the true value. The participants were informed of the average of the crowd's answers but not the true value. Furthermore, Lorenz *et al.* (2011) investigated how the average, as information, affects the next prediction. The social influence effect of Lorenz *et al.* (2011) has been explored in other studies. Mavrodiev and Schweitzer (2021) rigorously discussed the mathematical relationship between the wisdom of crowds and the social influence effect. Afflerbach *et al.* (2021) inquired as to what extent a crowd can provide accurate future predictions. When the participants of an experiment are expected to estimate using stable data such as Switzerland's population density, they are not expected to predict daily changes in population density.

In this article, we inquire how much the crowd can foresee future exchange rates. From 2004 to 2011, the data are provided by Nikkei, which holds the "Nikkei Yen Derby" competition. Each team must consist of at least five students from the same school and must have a teacher or professor as an advisor. The number of participating groups ranged from

361 teams at its lowest to 621 at its highest (Table 1). Thus, the competition was attended by sufficiently many groups. In total, 3,657 groups attended the competition over eight years. More than 21,942 students and teachers participated in predicting exchange rates during our research period. They were asked to forecast twice in May and June. At the end of each round, they were informed of the realized value of the foreign exchange rates. As the participants are informed of the realized value of the foreign exchange rate, the "Nikkei Yen Derby" is different from the experiments conducted by Lorenz *et al.* (2011). The data from the "Nikkei Yen Derby" do not deliver evidence of the "social influence effect" but indicate other types of biases.

Using these data, an empirical investigation was conducted to answer the following research questions. First, can we observe the wisdom of crowds effect? How accurately can the crowd predict foreign exchange rates when possible? Second, what is the forecast bias? If such a bias is observed, what is its effect on the predictor? This article answered these questions. These issues have never been rigorously investigated using large amounts of data. However, to some extent, these research questions are consistent with previous research on exchange rate forecasts. Ito (1990) used micro data from 44 companies that forecast yen-dollar exchange rates. These companies had "wishful expectations" in favor of the performance of their own countries' exports. Therefore, third, we can put forward the following research question. If we aggregate the forecasts of people who do not have "wishful expectations," such as students, can we get neutral forecasts for exchange rates without "wishful expectations?" Will we still observe biases in such cases? This article answers these questions. Data from the students' group that forecasted foreign exchange rates in the "Nikkei Yen Derby" were used. The research results indicate that the social influence effect is not the only bias inherent in crowd prediction.

2. Theory of wisdom of crowd effect
2.1 Permissible absolute deviation
The four propositions are described below. They are theoretically derived and assume that there is no bias in the assumptions of the theory. These propositions are necessary to observe the psychological bias of the crowd in the next section. To claim that the observed data are biased, we must explain the non-biased state theoretically. Herzog and Hertwig (2009) explained why the wisdom of crowds effect exists. Accumulated absolute deviation plays an important role in inducing the wisdom of crowds effect. Figure 1 gives the following example. Suppose a true value is 100, and this number is unknown to the estimators. It is assumed that absolute deviation is allowed with a deviation no greater than 10, which implies that the answer ranges from 90 to 110. The lower limit of 90 is the minimum, and the upper limit of 110 is the maximum. Suppose the first estimate is 110. While the second estimate can be more inaccurate, it cannot be so inaccurate that the average of all the responses is beyond the specified margin of error. Thus, the estimation range allows the second answer to be any number between 70 and 110. For example, a second estimation of 70 gives an average response of 90, which demonstrates an absolute deviation of 10 from the true number of 100. Next, the third estimation can improve the accuracy of the average estimation because it can
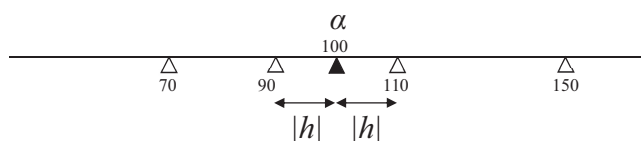

Figure 1.
Forecasting and the accumulated deviation

be any number between 90 and 150. The upper bound of 150 is derived from the multiplication of 110 by 3, followed by the subtraction of 180, which is the sum of the first two estimates. The estimation range increases monotonically under this condition; the absolute deviation must remain within a certain range of the true value.

This estimation process is generalized as follows: Let α be the true value to be estimated. Let $|h|$, an absolute value of $+h$ or $-h$, be the maximum permissible deviation from α. Let the first estimation by the first person be denoted by x_1 and suppose the deviation from the true value is $+|h|$. Therefore, $x_1 = \alpha + |h|$ is the estimated value. The second person provides an estimation with a deviation of $-|h|$. The respondents are numbered by the order of their responses. We assume that each person whose response number is even gives the estimation of $-|h|$ and those whose response number is odd provide a positive estimation with a deviation of $+|h|$.

Let us denote the function that gives the average of x_1 and x_2 as $\text{Ave}\left(\sum_{i=1}^{2} x_i\right)$. Then, the following is an average that satisfies the condition of the permissible range of accuracy from α,

$$\text{Ave}\left(\sum_{i=1}^{2} x_i\right) = \alpha - |h| = \frac{\alpha + |h| + q_2}{2}.$$

This is simplified to $q_2 = \alpha - 3|h|$, where q_i is the permissible limit of the estimated deviation by the ith person ($i = 1, 2$). Thus, for example, we obtain $q_2 = 70$ when $\alpha = 100$ and $|h| = 10$.

Furthermore, $\text{Ave}\left(\sum_{i=1}^{3} x_i\right) = \alpha + |h| = \frac{(\alpha+|h|)+(\alpha-3|h|)+q_3}{3}$, where $q_3 = \alpha + 5|h|$. By iterating these processes, we obtain, for example, $q_4 = \alpha - 7|h|$, $q_5 = \alpha + 9|h|$, and $q_6 = \alpha - 11|h|$. Thus, the general form of this numerical sequence is $q_i = \alpha + (-1)^{i+1}(2i-1)|h|$, where the absolute deviation $|h|$ is accumulated and deviates from the mean α as i increases.

Figure 2 presents the case of the 21 participants. The effect of the accumulated absolute deviation was simulated under the following assumptions: the true value was set at 100, and the maximum value of the range of a given forecast was 101. This forecast was estimated by participants who gave odd response numbers. The minimum was set at 99, estimated by participants who gave even response numbers. The mean of the responses was at 100 ± 1, and the estimations deviated from this number; examples of given estimations are 101, 97, 105, 93, . . ., 61, and 141. The mean of each set of estimations was either 101 or 99. Figure 3 depicts a simulation involving 124 participants, which was conducted under the same conditions as the previous simulation. As absolute deviations can be accumulated, the permissible range for a new estimation widens. Eventually, the range includes zero and negative numbers. However, as a negative number is not a feasible estimation of the foreign exchange rate, the estimation may have the following property: the average estimation by the crowds may tend to have a wider deviation in the range above the true value α. Thus, Proposition 1 is derived. This verifies that a larger number of participants n allows wider deviations in their estimations.

Proposition 1. If the absolute terms of deviation $|h|$ from the true value α are measured, then having a larger number of participants allows wider deviations in the estimations.

Proof. As demonstrated in the earlier example, the participants whose response numbers were odd had $q_1 = \alpha + |h|$, $q_3 = \alpha + 5|h|$, $q_5 = \alpha + 9|h|$, . . ., whereas the participants whose response numbers were even had $q_2 = \alpha - 3|h|$, $q_4 = \alpha - 7|h|$, $q_6 = \alpha - 11|h|$, ... First, we consider participants whose response numbers are
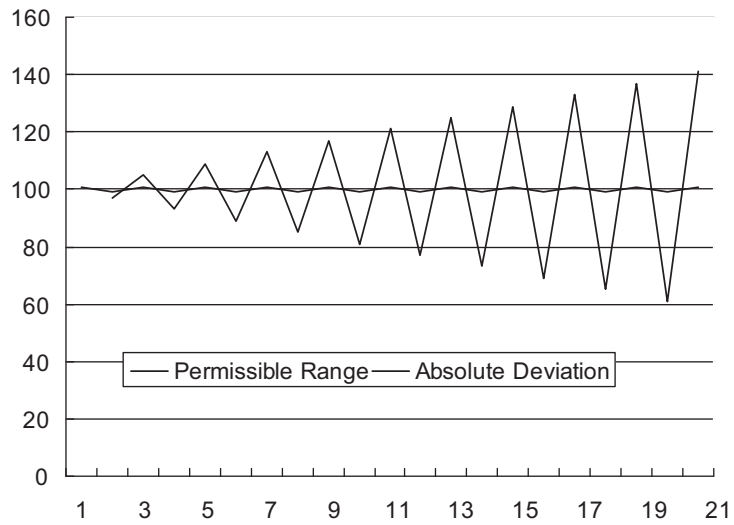
Figure 2.
A simulation with 21
estimates
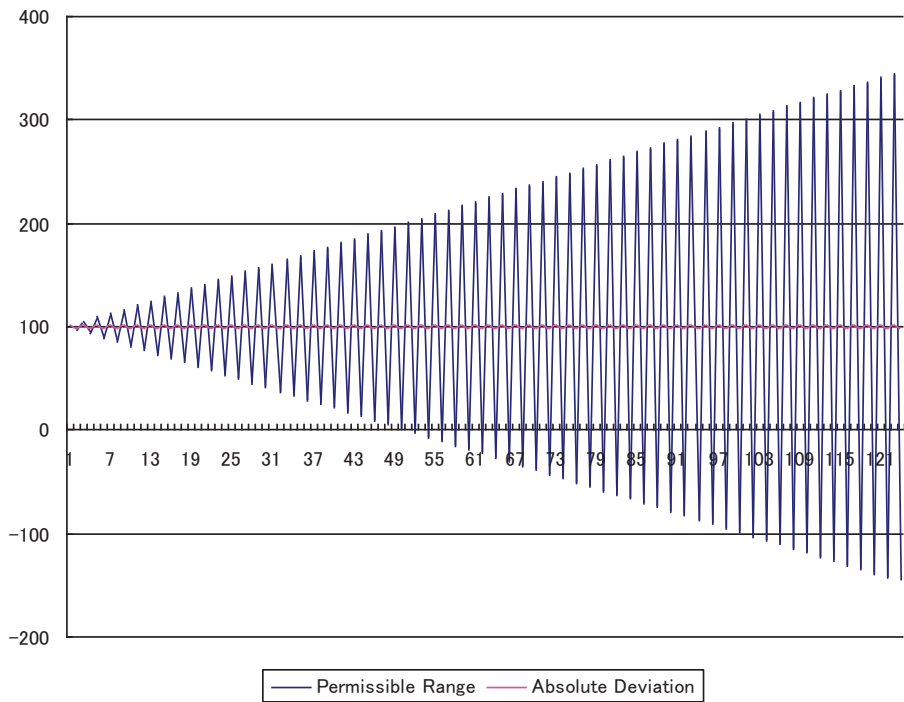
Source(s): Author

Figure 3.
An estimation with
124 estimates

Source(s): Author

odd. This group has a fixed true value of α, and the differences between the estimations increase by $2i-1$ for each response of q_i. Therefore, the general estimation of the ith participant, where i is odd, has the following form: $q_{\text{odd}} = \alpha + (2i-1)|h|$.

The summation of the estimations of the participants whose response numbers are odd is: $S_{\text{odd}} = \sum_{i=1}^{n} q_{\text{odd}} = \frac{\{\alpha+|h|+\alpha+(2n-1)|h|\}n}{2} = n(\alpha + n|h|)$.

Similarly, the sum of the estimations of the participants whose response numbers are even is:

$$q_{\text{even}} = \alpha - (2i - 1)|h| = \alpha - 2i|h| + |h|;$$

$$S_{\text{even}} = \sum_{i=2}^{n+1} q_{\text{even}} = \frac{\{\alpha - 3|h| + (\alpha - 2(n+1)|h| + |h|)\}n}{2} = n(\alpha - n|h| - 2|h|).$$

As the summation is measured in absolute terms, we can add the summations computed in this proof to obtain the following:

$$S = S_{\text{odd}} + S_{\text{even}} = n(\alpha + n|h|) + n(\alpha - n|h| - 2|h|) = 2n(\alpha - |h|).$$

hence, if $\alpha - |h| \neq 0$, then S deviates from zero as n increases. Q.E.D.

From Proposition 1, we can derive Proposition 2, which states that a larger number of participants allows the convergence of the forecasts to the true value α.

Proposition 2. If the total sum of the absolute deviations S has an upper bound, the deviation $|h| > 0$ from the true value α converges to α as n approaches infinity.

Proof. We use $\overline{S}$ to denote the total sum of the absolute deviation, which has an upper bound. According to the final equation in the proof of Proposition 1, we obtain:

$$\overline{S} = 2n(\alpha - |h|).$$

This equation can be rewritten as follows:

$$|h| = \alpha - \frac{\overline{S}}{2n}.$$

Clearly, $\overline{S}/2n$ approaches zero when n approaches infinity. Therefore, $|h|$ converges to α as n approaches infinity. This process may be written as follows:

$$h \to \alpha = \lim_{n \to +\infty} \left(\alpha - \frac{\overline{S}}{2n} \right).$$

Q.E.D.

As a mathematical intuition, one can imagine a graph where α is sandwiched between $+|h|$ and $-|h|$ and the area between $+|h|$ and $-|h|$ is represented by the bar graph. As n is infinite, the area of the graph is blackened totally by the bars between $+|h|$ and $-|h|$.

Alternatively, one could imagine the situation where n in this equation increases and is equal to the total number of people in the markets. In such a case, there would be as many predictors as participating in real foreign exchange market transactions. Then, the forecast and the actual transaction would lead to the same result. The reason the total sum of the absolute deviation $\overline{S}$ is confined cannot be derived from a mathematical inquiry. Instead, this

can be explained through various socioeconomic factors. Regarding the foreign exchange rates covered in this article, the lower and upper bounds of forecasting are such that they will never be zero or negative, and it will normally take a few months to double the current rates. Institutional forces, such as economic trends perceived by tacit knowledge (Polanyi, 1958, 1966) of traders, limit the upper bound of the foreign exchange rate when they make predictions. Consequently, one can assume that the resulting values of the experiment are based on stochastic functions that result from their maxima or minima of fluctuation. The simulation result in Figure 3 illustrates a situation in which there is an urn containing 124 balls with printed values, such as 101, 99, 103, 345 and −147. In this explanation, the order in which the numbers appear is from small to large, but this is not an essential condition for the proposition to hold. Thus, we obtain Proposition 3 to infer the standard deviation theoretically.

Proposition 3. When the mean absolute deviation $|h|$ is the upper bound of the estimation regarding the average α, the standard deviation σ increases as n increases.

Proof. The maximum permissible absolute deviation at the initial setting is defined as $|h|$, the deviation range is defined as σ, and the number of estimations is set at n.

Chebyshev's inequality is $\Pr(|X - E(X)| \geq |h|) \leq \frac{\text{Var}(X)}{|h|^2}$ (DeGroot, 1984, pp. 227–229).

As we apply the sample mean $E(\overline{X_n}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n\mu$ and variance. $\text{Var}(\overline{X_n}) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$, Chebyshev's inequality can be rewritten as $\Pr(|\overline{X} - \mu| \geq |h|) \leq \frac{\sigma^2}{n|h|^2}$.

We can rewrite this inequality such that $\Pr(|\overline{X} - \mu| \geq |h|) \cdot n|h|^2 \leq \sigma^2$.

The latter inequality holds if n, $|h|^2$ and σ^2 are positive. A simulation of Proposition 1, such as in Figure 3, demonstrates the locus of the equality condition.

Thus, we can rewrite

$$\Pr(|\overline{X} - \mu| = |h|) \cdot n|h|^2 = \sigma^2,$$

to get

$$|\sigma| = |h|\sqrt{n\left(\Pr\left(|\overline{X} - \mu| = |h|\right)\right)}.$$

As we only consider the condition that the absolute deviation is less than or equal to $|h|$, that is $\Pr(|\overline{X} - \mu| = |h|) = 1$, we can obtain $\sigma^2 = n|h|^2$, which implies that $|\sigma| = |h|\sqrt{n}$.

Consequently, $|\sigma|$ is a monotonically increasing function of n. Q.E.D.

It is known that the law of large numbers holds in a situation where Chebyshev's inequality holds. Furthermore, when the law of large numbers holds, the absolute deviation increases with an increasing number of estimators. Thus, we can derive Proposition 4 to indicate that the standard deviation diverges infinitely.

Proposition 4. Given Chebyshev's inequality, a weak law of large numbers allows an infinitely divergent standard deviation.

Proof. Suppose a stochastic variable x has an average μ and variance σ^2, and let λ be an arbitrary constant. Therefore, Chebyshev's inequality $\Pr(|X - E(X)| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$ holds, and furthermore, $\Pr(|X - E(X)| \leq \lambda\sigma) > 1 - \frac{1}{\lambda^2}$ holds. The weak law of large numbers requires that the sample average $\overline{x}$ is a consistent estimator of μ because $\overline{x}$ has an average μ and variance $\frac{\sigma^2}{n}$. Subsequently, we obtain

$\Pr\left(|X - E(X)| \leq \lambda \frac{\sigma}{\sqrt{n}}\right) > 1 - \frac{1}{\lambda^2}$. Additionally, we define another arbitrary constant $\varepsilon = \frac{\lambda\sigma}{\sqrt{n}}$ such that: $\Pr(|X - E(X)| \leq \varepsilon) > 1 - \frac{\sigma^2}{\varepsilon^2 n}$.

If n increases indefinitely, $\lim\limits_{n\to\infty} \frac{\sigma}{\varepsilon^2 n} = 0$ and $\lim\limits_{n\to\infty} P(|\overline{x} - \mu| < \varepsilon) = 1$. As the arbitrary constants λ and ε are fixed, $\varepsilon = \frac{\lambda\sigma}{\sqrt{n}}$ can be rewritten as $\sigma = \frac{\overline{\varepsilon}}{\lambda}\sqrt{n}$; hence, $\sigma^2 = \left(\frac{\overline{\varepsilon}}{\lambda}\right)^2 n$.

Thus, σ^2 increases infinitely as n increases. Q.E.D.

2.2 A higher degree of accuracy

Figure 4 illustrates the result of a simulation in which the radius of the permissible range of forecasts decreases to zero. In this case, the variance reached its maximum value in the middle of the simulation. This simulation has the following assumptions: as the number of participants increases, $|h|$ decreases. Each new participant must produce an estimation that renders the mean estimate more accurate by at least one hundredth, and subsequently, the estimated answer gradually becomes more accurate. Here, the first estimation is set equal to 101.00 and the second estimation is set at 99.00. Therefore, the average estimate determined from the estimations of the first three participants will have an upper bound of 100.99. Similarly, after the fourth estimator's response, the overall estimation will have a lower bound of 99.01. The width of the range of the respondents' margin of error for their estimations will decrease until it reaches the value of 100, which is attained after 200 participants have given their estimations.

Let us consider the third estimation as an example. Given that the first and second estimations are 101.00 and 99.00, respectively, we want to find the overall estimation, which has become more accurate after the second estimation. To have 100.99 as the average of the participants' responses, the sum of the given estimations must be 302.97, which is the average of the participants' responses. As the first and the second estimations are already set at 101.00 and 99.00, the third participant's estimation must be no greater than 102.97. Figure 4 illustrates the case of 200 participants. The permissible range of estimation demonstrates
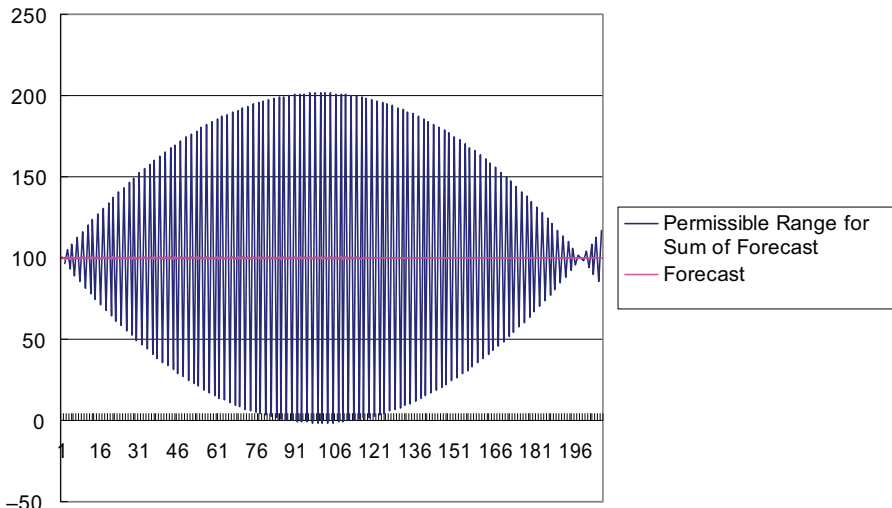
Figure 4.
The allowable
deviation $|h|$ decreases
by one hundredth after
each forecast: the case
of two hundred
forecasts

Source(s): Author

saturation, reflecting the narrowing range of estimation attained after approximately 100 participants. However, even in this restricted example, the permissible range widens between 200 and 0 when there are approximately 100 estimators for whom the data reaches the upper and lower bounds. By considering the foreign exchange rate of the Japanese yen, which was between 70.00 and 135.00 during the 1990 and 2000s, this model of the accumulated absolute deviation can cover the feasible range of the exchange rate's movement.

Theoretically, the wisdom of crowds effect is observed when the mean is calculated by alternately adding and subtracting successive participants' estimates. The deviation is measured in absolute terms with the cancellation of the positive and negative values. Given an accumulated absolute deviation, estimations that are either too high or too low are canceled out. An incorrect guess is canceled out by the accumulated absolute deviation from others' responses. Thus, accumulated absolute deviations play an important role in generating the wisdom of crowds effect. The above exposition theoretically shows that the law of large numbers allows a wider variance by forecasters and allows the mean to converge to the true value. However, the fourth section of this paper reports two types of biases in the data from the "Nikkei Yen Derby," where the variance was smaller, and the mean got closer to an observable foreign exchange rate. In the next section, we examine the wisdom of crowds effect.

3. An empirical study on the wisdom of crowd effect
3.1 Timeline of the Nikkei Yen Derby
To investigate whether the wisdom of crowds effect exists in forecasts of the future foreign exchange market, an interesting and relevant experiment was conducted in Japan. Nikkei, one of the biggest newspaper companies (known as the Nikkei Index in the Japanese stock market), has been holding an annual competition for its readers since 2000. The competition is called the "Nikkei Yen Derby," and many student groups forecast the foreign exchange rate during the competition. The participants comprise students from junior high school, high school, college, technical college, university, or graduate schools. As a condition, each participating team must consist of at least five students belonging to the same school, with a teacher or professor as an advisor. The teams must predict the yen-dollar exchange rates on two occasions. The team with the smallest sum of their absolute deviations from the actual market values wins the competition. Nikkei provided the author with data from 2004 to 2009, and the data from 2010 to 2011 were retrieved from Nikkei's website. Nikkei exhibits all participants' performances on its website for a certain period after each competition.

Figure 5 illustrates the timeline of the Nikkei Yen Derby. For example, the 10th Nikkei Yen Derby was held in 2010. The first deadline was Monday, May 11, and the participants were asked to predict the final exchange rate on Monday, May 31. Similarly, the deadline for the second prediction was Tuesday, June 8, and the participants had to predict the final exchange rate as of Wednesday, June 30. The 11th Nikkei Yen Derby in 2011 set the first deadline on
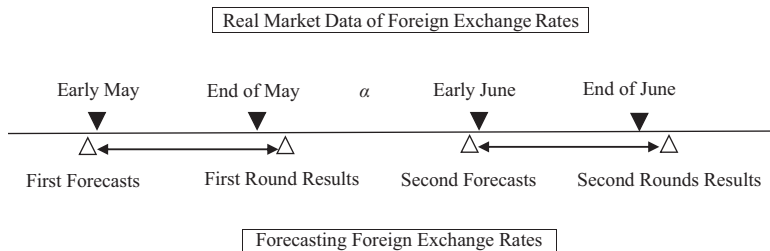

Figure 5.
Forecasting timeline for the Nikki Yen Derby

Tuesday, May 10, for forecasting the final exchange rate on Tuesday, May 31. The second prediction was made by Monday, June 6, to predict the final price on Thursday, June 30. All predictors had approximately three weeks to make their predictions. As is customary in Japan, predictions were made on the value of yen per dollar.

3.2 The accuracy of prediction and causes of random walk
Table 1 summarizes the results of the Nikkei Yen Derby from 2004 to 2011. For instance, the real value at the end of May, 91.47 yen per dollar, was the exchange rate at the end of May 31, 2010. The average prediction of the participating teams, the "Average Forecast in May," was 93.37 yen per dollar, and the difference was 1.90 yen. The average predicted value in May demonstrated a 2.07% difference. The value at the end of June 30, 2010, was 88.65 yen per dollar, and the average prediction was 92.26 yen per dollar. Thus, there was an absolute deviation of 3.61 yen; that is, the predicted value deviated by 4.07% from the actual value. Figure 6 illustrates the distribution of the forecasts in May and June 2010. The forecast for the end of May 2010 had the smallest absolute deviation, 1.90 yen, or 2.07%, during the period listed in Table 1.

Table 1 lists the mean absolute deviations for May. For all eight years, it is 1.43 yen, and the mean absolute deviation in June for the same period is 2.09 yen. All the forecasts in Table 1 have deviations smaller than 5% from the actual values of the foreign exchange rates. As such, these deviations are smaller than the 5% error commonly used in statistical tests. However, considering that the group estimated and assembled by Galton (1907) correctly reported the ox's weight, the deviations in the Nikkei Yen Derby may not be so small. While the true weight of the ox was 1,198 pounds, the average expected value was 1,207 pounds; there was a 0.75% mean absolute deviation of the projected weight.

	2004	2005	2006	2007	2008	2009	2010	2011	Mean absolute deviation
(a) Average forecast in May	111.09	105.76	113.04	120.07	105.13	98.55	93.37	82.06	
(b) Real value at the end of May	109.55	108.15	111.84	121.61	105.44	96.44	91.47	81.59	
(c) Absolute deviation |(a)−(b)|	*1.54*	*2.39*	*1.20*	*1.54*	*0.31*	2.11	*1.90*	*0.47*	*1.43*
(d) Number of groups	473	411	372	440	463	621	527	437	
(e) Average forecast in June	110.91	107.42	112.02	121.76	106.12	97.10	92.26	81.62	
(f) Real value at the end of June	108.68	110.36	114.65	123.47	105.32	95.55	88.65	80.41	
(g) Absolute deviation |(e)−(f)|	2.23	2.94	2.63	1.71	0.80	*1.55*	3.61	1.21	2.09
(h) Number of groups	458	401	361	431	455	601	515	435	

Note(s): The numbers in underlined italics indicate a smaller "Absolute deviation" when comparing (c) and (g)

Table 1.
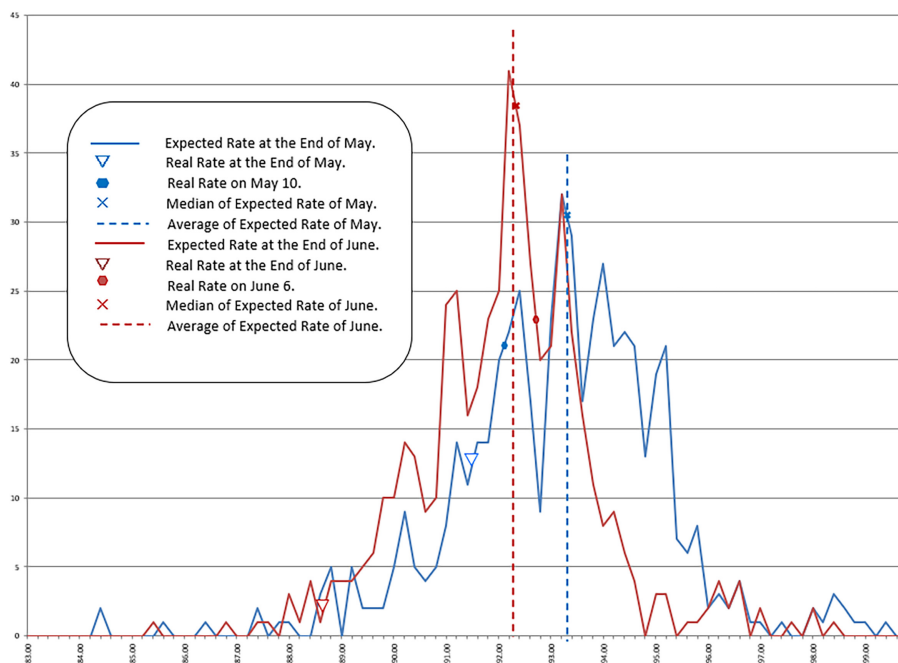The results of the "Nikkei Yen Derby" and wisdom of crowd effect

Figure 6.
The forecast of the exchange rate by groups by 20/100th of Yen per dollar: May 2010 and June 2010

The distributions appear somewhat different if the user subdivides the yen-dollar foreign exchange rate in different ways. Figures 6 and 7 were derived using the same dataset; both illustrate the results for May and June 2010. Figure 6 illustrates the forecast values aggregated within the range of 20/100 yen as the aggregation width. Figure 7 illustrates the predicted value aggregated with the aggregation width set to 1/100 yen. As illustrated in Figures 6 and 7, some spikes result from the concentration of many people's predictions. Blank values, a characteristic of Figure 7, were not predicted by anyone. Figure 7 indicates that some values in the forecasts are concentrated on many people, and some are not. If these predictors meet as sellers and buyers in the forex market, blank values will not close the deal and values with many spikes will lead to an overshooting of the price. In other words, by looking at the distribution of forecasts, it is clear why the foreign exchange market transaction prices move in a zigzag fashion, according to Brownian motion.

We verified the randomness of the forecasts. Figure 8 illustrates the correlation between the rankings made by the forecasting groups in 2010. As evident from the figure, there is no correlation between these forecasts. Moreover, the correlation coefficient is 0.09. We checked the same tendency in the data obtained from the rest of the years. Furthermore, the students' groups that forecasted the foreign exchange rate precisely in May were not good predictors in June. In this sense, the wisdom of crowd effect is paradoxical: even though some groups showed better performances in May, it would be erroneous to assume that any of them will be "expert predictors" in June.

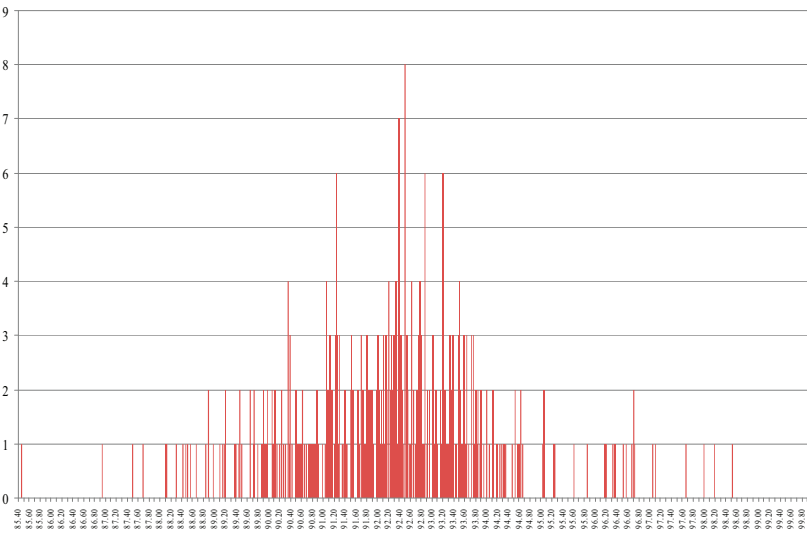## 4. The actuality bias and fact-convergence effect
4.1 Definitions of biases
As the Nikkei Yen Derby asks participants to make forecasts twice in May and June, we can compare the distribution of the forecasts to determine whether some sort of bias exists.
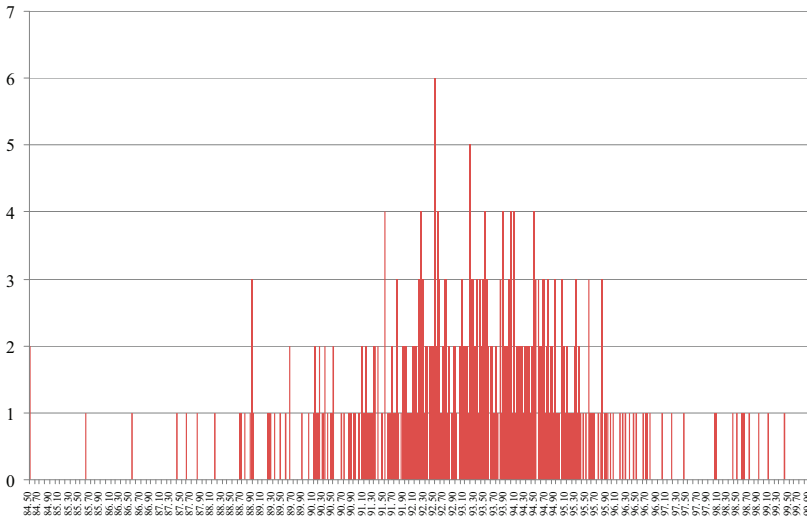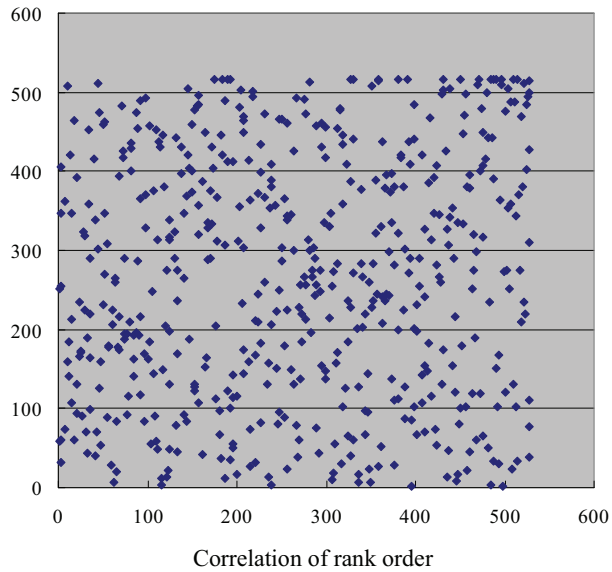
May 2010

(a)

June 2010

(b)

Note(s): Four data points are not included in the May data, and two data points are not included in the June data because they were outliers
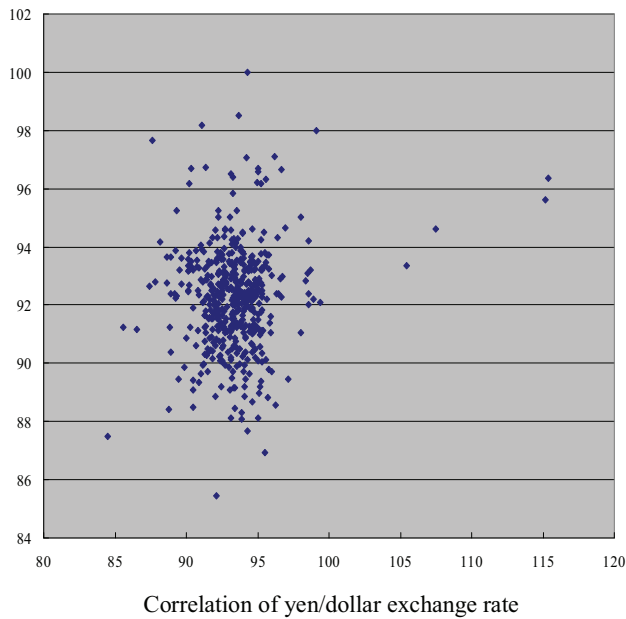
(a)

(b)

Lorenz *et al.* (2011) argued that people are influenced by the information given after the first
round of experiments—the "social influence effect"—due to data from the first round. In our
data of the Nikkei Yen Derby, the groups of students were not provided any information, and

they were informed which group was closest to the foreign exchange rate after the first round. Hence, we cannot call this bias the social influence effect. However, we can expect other biases to exist for those who participated in the Nikkei Yen Derby. We define two classes of biases inherent in forecasting future values.

Definition 1. Actuality bias is defined as having a closer prediction value in the second round to the actual observable value by the due date of forecasting.

Definition 2. The fact-convergence effect is defined as having a smaller standard deviation for the second-round prediction.

4.2 Actuality bias

Proposition 2 suggests that having many participants leads to accurate predictions by the crowds. However, the theoretical results could not be ascertained. Table 1 illustrates the number of participants and deviations from the predictions. The number of participants was unrelated to the accuracy of the predictions. The most accurate prediction was an average estimate made in May 2008 by 463 groups with a deviation of 0.31 yen per dollar. When 621 groups participated in May 2009, their prediction had a deviation of 2.11 yen per dollar. The correlation coefficient between the two variables was only 0.048 over 16 pairs of periods.

Table 1 also illustrates the systematic bias. This can be found in the differences in the absolute deviations by the forecasts given in May and June for each year and that of the average for eight years. For example, in 2011, the absolute deviation in May was 0.47 yen per dollar and that in June was 1.21 yen per dollar. The absolute deviations are smaller in May for seven out of eight years in the period under consideration. Suppose one assumes that the difference between the absolute deviations in May and June is determined merely by chance and the probability of a prediction being too high or too low follows a binomial distribution. In that case, we can calculate the probability of having higher absolute deviations in June than in May seven times out of eight: $_8C_7p^7(1-p)^1 = 8 \cdot \left(\frac{1}{2}\right)^8 = \left(\frac{1}{32}\right) = 0.03125$, where $p = 1/2$. This percentage is lower than the 5% level, which supports the hypothesis that the absolute deviation is random. The wider absolute deviation in June indicates that the crowd did not improve its forecasting capability in June. It did not learn from the experience in May, and the forecasts worsened in June. This may be attributable to some bias. The bias is related to the information on the foreign exchange rate obtained by the participating groups during May or early June.

Table 2 illustrates the absolute difference between the (a) average forecast in May and (b) real value in early May. It also indicates the absolute difference between the (e) average forecast in June subtracted from the (f) real value in early June. As illustrated by the timeline in Figure 5, the students can observe (b) real value in early May and (f) real value in early June before the due date of forecasting. If |(a)−(b)| is close to zero, it indicates that the forecaster is influenced by referring to the actual foreign exchange rate before the due date. When we compare |(a)−(b)| as May data and |(e)−(f)| as June data, we see that |(e)−(f)| was smaller for seven of the eight years. This implies that (e) the average forecast in June was greatly influenced by (f) the real value in early June.

One can infer the reason Table 1 recorded a wider deviation of the forecasted value in June. As the (e) average forecast in June was influenced by (f) real value in early June, as illustrated in Table 2, the forecast had been dragged to the real value. In May, the students were less biased toward the real value before forecasting, and the average forecast in June was not more accurate than that in May. This was due to the bias because of looking at the actual foreign exchange rate before another forecasting round. After the crowd learns reality, they reduce the variety of predictions. Thus, we call it actuality bias.

	2004	2005	2006	2007	2008	2009	2010	2011	Average closeness
(a) Average forecast in May	111.09	105.76	113.04	120.07	105.13	98.55	93.37	82.06	
(b) Real value in early May	110.58	105.32	111.89	119.83	103.08	99.38	92.11	80.64	
(c) Real value at the end of May	109.55	108.15	111.84	121.61	105.44	96.44	91.47	81.59	
(d) Closeness: |(a)−(b)|	0.51	*0.44*	1.15	0.24	2.05	0.83	1.26	1.42	0.99
(e) Average forecast in June	110.91	107.42	112.02	121.76	106.12	97.10	92.26	81.62	
(f) Real value in early June	111.05	107.87	111.74	121.74	105.89	96.71	92.70	80.64	
(g) Real value at the end of June	108.68	110.36	114.65	123.47	105.32	95.55	88.65	80.41	
(h) Closeness: |(e)−(f)|	*0.14*	0.45	*0.28*	*0.02*	*0.23*	*0.39*	*0.44*	0.98	*0.37*

Note(s): "Closeness" indicates the absolute deviation between the average forecast and the real value of the due date of the forecast. Forecast deadlines were set in early May and early June. The italicized numbers indicate that (h) is smaller than (d), or "Closeness" in June is smaller than that in May

Table 2.
The real exchange rates in early May and June and the results of the "Nikkei Yen Derby": actuality bias in May and June

4.3 Fact-convergence effect
Although Propositions 1, 3 and 4 of this paper suggest that having a larger number of participants allows wider deviations in the estimations, the data from the Nikkei Yen Derby presents narrower deviations from the student predictions. Table 3 illustrates the standard deviations of the values forecasted by the student groups. The "May" row illustrates the standard deviation of foreign exchange rates used to predict the future price, and the "June" row illustrates the standard deviation among the second-round forecast. The standard deviations for "June" are smaller than those of "May." In Table 3, the ratio between the two rows is illustrated in the "June/May ratio" row. Every year, the ratio is less than unity. The absolute deviation in June is smaller than that in May, which is eight times out of eight. If we calculate the probability of having this result assuming a binary distribution, it is either zero or one.

We revealed that the fact-convergence effect gives a smaller standard deviation for the second-round prediction. The fact-convergence effect concerns the standard deviation of the predicted value. People tend to be crowded in a smaller range of prediction values. Actuality bias is dependent upon the mean of the predicted values. People tend to predict closer to the observable values at the due date of prediction.

Figures 9 and 10 present the profile of the predictions in 2011. As illustrated in Table 3, the standard deviations were 3.595 and 1.823 in May and June, respectively. The width of the predictions decreased in June, as illustrated in Figures 9 and 10. This decrease suggests that participants in the Nikkei Yen Derby are more influenced by the real exchange rate on the day of their forecasting deadline in June. They learn recent trends in real exchange rates in May, and their predictions converge to the observable rate as they forecast the rate in June.

	2004	2005	2006	2007	2008	2009	2010	2011
May	4.225	1.960	2.913	9.490	3.838	2.817	2.542	3.595
June	*2.651*	*1.664*	*1.921*	*9.447*	*2.779*	*2.171*	*1.856*	*1.823*
June/May ratio	0.6274	0.8488	0.6597	0.9955	0.7240	0.7705	0.7302	0.5070

Note(s): The numbers in underlined italics indicate a smaller "Standard deviation" when comparing May and June

Table 3.
Standard deviation of forecasted value in May and June: fact-convergence effect

Expected Rate at the End of May.
Real Rate at the End of May.
Real Rate on May 10.
Median of Expected Rate of May.
Average of Expected Rate of May.
Expected Rate at the End of June.
Real Rate at the End of June.
Real Rate on June 6.
Median of Expected Rate of June.
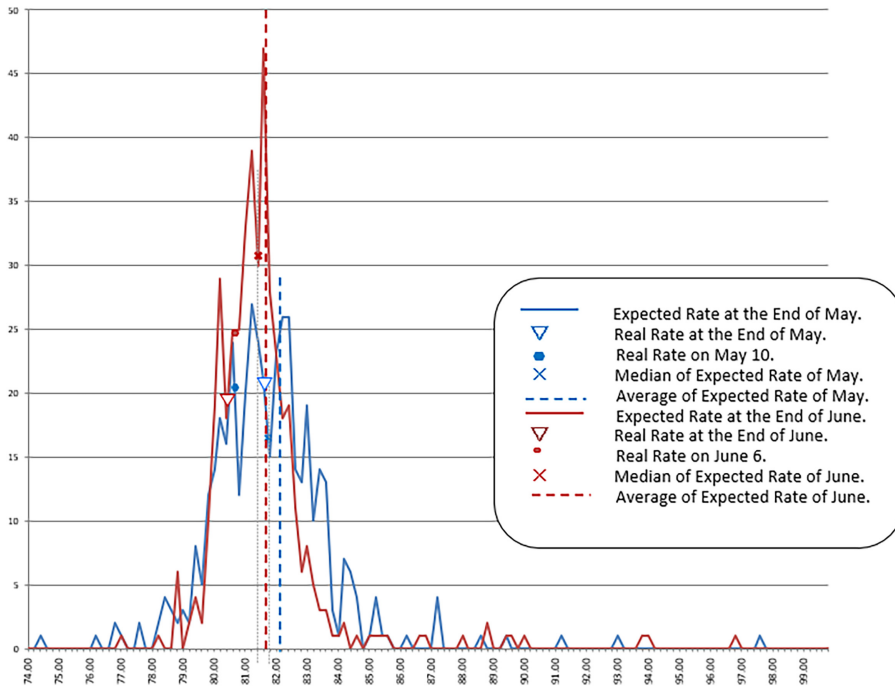Average of Expected Rate of June.

Figure 9.
The forecast of the
exchange rate by
groups: range taken by
20/100th of Yen per
Dollar, May 2011 and
June 2011

The convergence in June is observed toward the rate on the due date when the student groups submit their predictions to Nikkei. The exchange rate in early June of the due date does not guarantee prediction accuracy for the end of June. The fact-convergence effect can explain why the student groups made worse predictions in June than in May. The wider deviations of June predictions in Table 1 may be due to the convergence of estimates after one round of predictions.

5. Discussions and limitations
The social influence effect proposed by Lorentz *et al.* (2011) raises the question of who makes up a group and who undertakes the "social" interactions. Their experiments were conducted with 144 participants. They found a bias among some groups of participants that were given information on estimated values, such as the average or individual records of their answers. They found that some groups made up of 144 participants were biased by having information about the answers of other groups. Lorentz *et al.* (2011) claim that they examined "social" influence rather than 144 participants divided into three groups. They investigated the influence of other respondents in each group on their limited number of participants. They circulated the participants' answers in their experiments after the first round. Their data and findings, based on 144 participants, could be termed as groups but not appropriately be termed "social" in the sense that society involves members undertaking social experimental interactions.

If subjects are provided with some data, and if it influences their decision-making, everything could be understood as encompassed by the concept of the anchor effect. We thus need to elaborate the definition of the anchor effect in relation to forecasting foreign exchange rates. For example, the anchor effect is measured differently in the case of a small number of
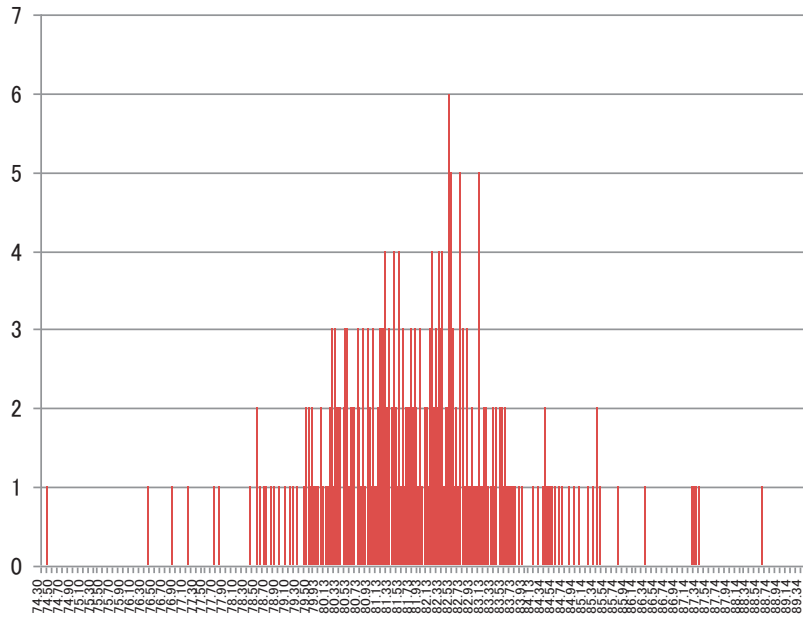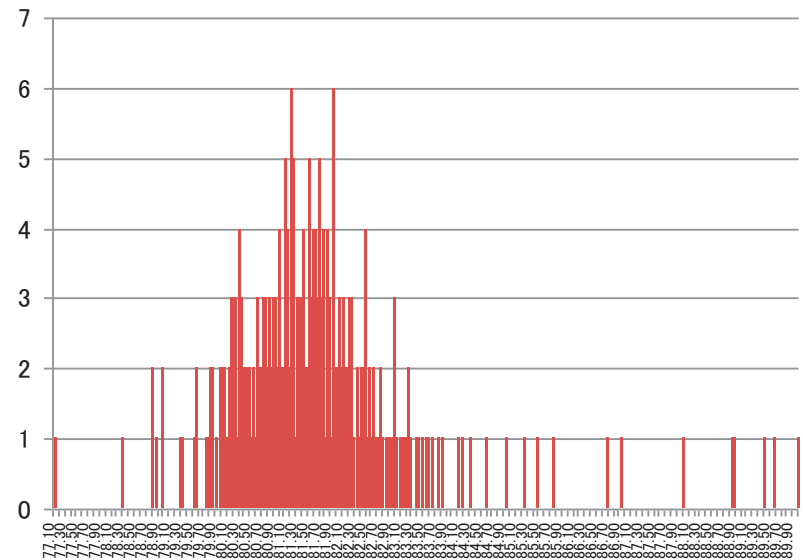
May 2010

(a)

June 2010

(b)

Figure 10.
The forecasts of the
exchange rate by
groups: range taken
by 1/100th of Yen
per dollar

Note(s): Six data points are not included in the May data, and three data points
are not included in the June data because they were outliers

discrete decision choices and the case of continuous data with a large range. In the latter case, the median, mean and variance can be measured. In the former case, it is meaningless to measure them. In this sense, the actuality bias and the fact-convergence effect are important subsets of the anchor effect.

We defined two novel biases in forecasting foreign exchange rates: the actuality bias and the fact-convergence effect. Actuality biases are found in our daily lives. An actuality bias implies that we base our predictions on the facts we observe. Therefore, the predicted value is not merely the average value of individual predictions in the sense that the prediction is made by regressing to the observed value. For example, if a neighbor dies at a young age, we recognize the importance of life insurance from that reality. In that case, the premium for life insurance may be higher than the amount required, or one may wish to have multiple life insurance policies. As another example, recorded prices, such as the highest and lowest stock prices in the last ten years or changing land prices over the last 20 years, can have biases for future forecasts. The average value predicted by crowds is biased toward data values based on past facts.

The existence of the fact-convergence effect demonstrates how difficult it is to maintain diverse opinions. Predicting the future is an important issue for the leadership of an organization, but if the predictions in the first round are successful, they tend to make similar predictions in the second round. This suggests that crowds may mistakenly think that the correct answer in the first round can guarantee success in the second round. The existence of the fact-convergence effect could also explain a crash in stock prices or exchange rates (Le Bon, 1895). The crash may be caused by the disappearance of diverse forecasts and market participants making the same forecasts. Maintaining diversity might also contribute to the diversification of future forecasts.

The limitation of this study is that the forecasts were summarized at one point in each month. Dynamic predictions require forecasts by week, day, hour, minute and second. Collecting data for cloud populations is costly, even with modern technological methods. However, with the development of the Internet, smartphones, and the Internet of things, it may be possible to aggregate data to achieve continuous forecasting in the future.

6. Conclusions: the dynamics of learning crowds
A simple thought experiment may suffice to explain why the wisdom of crowds effect exists in the foreign exchange market. If all the traders in the market participated in the Nikkei Yen Derby, then the forecasted value of the yen would converge to the actual foreign exchange value. Although it is commonly inferred that the wisdom of crowds effect is related to the law of large numbers, this article is novel in presenting formal proof of this connection. Proposition 1 of this article explains that if the deviation from the true value is measured using absolute deviation, having a larger number of participants allows wider deviations in the estimations. From a theoretical point of view, one cannot assume that this accumulated deviation has any limitation. However, empirical analysis of the Nikkei Yen Derby data reveals the following facts. By stating these facts, we can now summarize the answers to the three research questions mentioned in the introduction.

(1) We can observe the wisdom of crowds effect such that the average absolute deviation in May during an eight-year period was 1.43%, and this deviation in June was 2.09% during the same period. The accuracy of the predictions was higher in the first round of prediction.

(2) Participants were affected by information on the real exchange rate through early June, such that they modified their forecasts by referring to the actual data in early June. This suggests that an actuality bias exists in the forecasts, and the participants lost their diverse prospects. Forecasts made in May and those made in June indicated no correlation with each other.

(3) We still observed bias when students who do not have "wishful expectations" predicted the future foreign exchange rates. The standard deviations of the June forecasts were smaller than those of May, and the data supported the fact-convergence effect.

These results indicate that the participants were influenced by observable facts. Whereas the social influence effect by Lorentz *et al.* (2011) exists when the experiment organizer provides the subjects with the information, only a researcher or an organizer such as Nikkei can calculate the average of these predictions in the Nikkei Yen Derby. Therefore, we still need to examine whether the fact-convergence effect necessarily leads to herding behavior because such a contest as the Nikkei Yen Derby does not give the participating groups a chance to observe other participants' behavior.

References

Afflerbach, P., van Dun, C., Gimpel, H., Parak, D. and Seyfried, J. (2021), "A simulation-based approach to understanding the wisdom of crowds phenomenon in aggregating expert judgment", *Business and Information Systems Engineering*, Vol. 63 No. 4, pp. 329-348.

Blackwell, C. and Pickford, R. (2011), "The wisdom of the few or the wisdom of the many?: an indirect test of the marginal trader hypothesis", *Journal of Economics and Finance*, Vol. 35 No. 2, pp. 164-180.

Cheon, C., Kim, Y. and Yoon, S. (2012), "Can we predict exchange rate movements at short horizons", *Journal of Forecasting*, Vol. 31 No. 7, pp. 565-579.

Da, Z. and Huang, X. (2020), "Harnessing the wisdom of crowds", *Management Science*, Vol. 66 No. 5, pp. 1847-1867.

DeGroot, M.H. (1984), *Probability and Statistics*, 2nd ed., Addison-Wesley Publishing Company, Reading.

Galton, F. (1907), "Vox populi", *Nature*, Vol. 75 No. 1949, pp. 450-451.

Herzog, S.M. and Hertwig, R. (2009), "The wisdom of many in one mind: improving individual judgements with dialectical bootstrapping", *Psychological Science*, Vol. 20 No. 2, pp. 231-237.

Horaguchi, H.H. (2014), *Collective Knowledge Management: Foundations of International Business in the Age of Intellectual Capitalism*, Edward Elgar, Cheltenham.

Ito, T. (1990), "Foreign exchange rate expectations: micro survey data", *American Economic Review*, Vol. 80 No. 3, pp. 434-449.

Jacowitz, K. and Kahneman, D. (1995), "Measures of anchoring in estimation tasks", *Personality and Social Psychology Bulletin*, Vol. 21 No. 11, pp. 1161-1166.

Kahneman, D. (1992), "Reference points, anchors, norms, and mixed feelings", *Organizational Behavior and Human Decision Processes*, Vol. 51, pp. 296-312.

Kerr, N.L. and Tindale, R.S. (2011), "Group-based forecasting?: a social psychological analysis", *International Journal of Forecasting*, Vol. 27 No. 1, pp. 14-40.

Kremer, I., Mansour, Y. and Perry, M. (2014), "Implementing the 'wisdom of the crowd'", *Journal of Political Economy*, Vol. 122 No. 5, pp. 988-1012.

Le Bon, G. (1895), *La psychologie des foules, Presses Universitaires de France, Paris, 1988 (translated in Japanese, Gunshūshinri, Kodansha Gakujutsu Bunko*, Tokyo, 1993).

Lee, M.D., Zhang, S. and Shi, J. (2011), "The wisdom of the crowd playing the price is right", *Memory and Cognition*, Vol. 39 No. 5, pp. 914-923.

Lorenz, J., Rauhut, H., Schweitzer, F. and Helberg, D. (2011), "How social influence can undermine the wisdom of crowd effect", *Proceedings of the National Academy of Sciences*, Vol. 108, pp. 9020-9025.

Mannes, A.E. (2009), "Are we wise about the wisdom of crowds? The use of Group judgments in belief revision", *Management Science*, Vol. 55 No. 8, pp. 1267-1279.

Mavrodiev, P. and Schweitzer, F. (2021), "The ambiguous role of social influence on the wisdom of crowds: an analytic approach", *Physica A: Statistical Mechanics and Its Applications*, Vol. 567 No. 1, pp. 1-14, 125624.

Mozer, M.C., Pashler, H. and Homaei, H. (2008), "Optimal predictions in everyday cognition: the wisdom of individuals or crowds", *Cognitive Science*, Vol. 32 No. 7, pp. 1133-1147.

Polanyi, M. (1958), *Personal Knowledge: Towards a Post-Critical Philosophy*, University of Chicago Press, Chicago.

Polanyi, M. (1966), *The Tacit Dimension*, Peter Smith, Gloucester, MA, Reprinted by Doubleday & Company, 1983.

Rauhut, H. and Lorenz, J. (2011), "The wisdom of crowds in one mind: how individuals can simulate the knowledge of diverse societies to reach better decisions", *Journal of Mathematical Psychology*, Vol. 55 No. 2, pp. 191-197.

Steyvers, M., Lee, M., Miller, B. and Hemmer, P. (2009), "The wisdom of crowds in the recollection of order information", *Advances in Neural Information Processing Systems*, Vol. 22, pp. 1-10.

Surowiecki, J. (2004), *The Wisdom of Crowds*, W.W. Norton & Company, New York.

Treynor, J.L. (1987), "Market efficiency and the bean jar experiment", *Financial Analysts Journal*, Vol. 43 No. 3, pp. 50-52.

Corresponding author
Haruo H. Horaguchi can be contacted at: horaguch@hosei.ac.jp