IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 22, Issue 3, Ser. 1 (May. – June. 2026), PP 56-62
www.iosrjournals.org

Testing Collective Intelligence with a Jar of Jellybeans: An Empirical Study of the Wisdom of the Crowd

Yuvraaj Disawal
Walnut Grove High School, Texas

ABSTRACT
The idea of the wisdom of the crowd suggests that when a diverse group of people make independent guesses, their combined answer can be surprisingly accurate. This study explores that idea by asking whether a mixed group of individuals can correctly estimate the number of jellybeans in a jar just by looking at it. A total of 173 participants from different age groups and locations shared their individual estimates. The study also looks at whether age has any effect on how accurate these guesses are. The results showed that when all the guesses were averaged, the final estimate came very close to the actual number of jellybeans. This supports the idea that group judgment can be more reliable than individual guesses. Interestingly, age did not have a significant impact on how accurate the estimates were, suggesting that personal perception matters more than demographic factors in this kind of task. The findings also highlight how even simple tasks can demonstrate complex statistical principles in an accessible way. The consistency of results across participants strengthens confidence in collective estimation methods. This experiment shows that large groups do not need expert knowledge to arrive at accurate conclusions. It also suggests practical applications in areas such as forecasting, decision-making, and problem-solving. Additionally, the study encourages further research into other factors that may influence collective intelligence. Overall, the findings highlight how powerful collective thinking can be, especially when people think independently.

I. Introduction

The Wisdom of the Crowd is a phenomenon whereby the average of many independent guesses can be amazingly close to the correct answer. This study tests whether a demographically diverse sample can collectively estimate the number of jellybeans in the jar with statistical accuracy by visually inspecting the jar. The key principles behind this theory are diversity of opinion, independence, and decentralization. These principles hold that, under certain conditions, such as diverse and unbiased input, a group of people can exhibit collective intelligence (Lorenz et al., 2011). Author James Surowiecki asserts in The Wisdom of Crowds that "under the right circumstances, groups are knowledgeable and are often smarter than the smartest people in them." *(*Surowiecki, 2004). About 147 participants from various geographic locations and age groups provided independent estimates of the jar's contents based on a simple visual examination. The study chronicles the setup, process, findings, and implications of this smaller-scale but significant confirmation of the Wisdom of the Crowd (Surowiecki, 2004; Galton, 1907). This experiment also examined whether age contributed to accuracy, a factor rarely tested in small-scale demonstrations of collective intelligence (Woolley et al., 2010). The results supported the theory: the group's average was impressively close to the correct number, and age did not meaningfully predict estimation accuracy, further strengthening the idea that individual perceptual differences outweigh demographic ones in this task **(**Lorenz et al., 2011).

For just one second, picture yourself at a fair with people guessing the weight of a giant pumpkin. There might be some extreme guesses, some underestimations, and some overestimations, and the crowd appears uncertain and random, as it should be. But then something unexpected occurs: when you average all the guesses, it turns out to be very close to the correct weight. It seems like a statistical trick, but it is a real phenomenon known as the Wisdom of the Crowd (Surowiecki, 2004; Larrick & Soll, 2006). This phenomenon was first documented by British scientist Francis Galton in 1907. At a livestock fair, Galton noticed that the average estimate of a crowd guessing the weight of an ox was remarkably accurate, closer to the correct weight than any individual guess (Galton, 1907). The puzzling accuracy sparked interest in the power of collective intelligence, leading to the Wisdom of the Crowd theory **(Surowiecki, 2004; Hong & Page, 2004; Woolley et al., 2010).**

However, individual differences such as age may theoretically skew or influence accuracy; cognitive psychologists Shaw and Craik (1989) have shown that age can affect memory, reasoning, and decision strategies. The theory has intrigued psychologists, economists, and data scientists for decades (Surowiecki,

2004). In summary, the theory suggests that if a sufficient number of people make independent guesses or estimates, the average of those guesses is often incredibly accurate, even more so than any single expert (Surowiecki, 2004; Larrick & Soll, 2006). To connect all the points, this study tests the core principle of crowd wisdom and also evaluates whether age influences the accuracy of individual guesses.

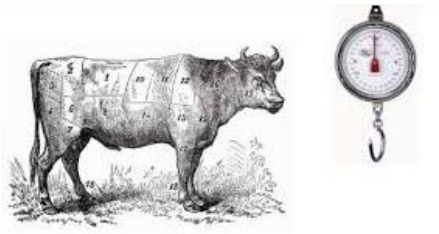


Figure 1: Francis Galton's Ox-Weighing Experiment
Source: Adapted from Galton (1907)

II. Literature Review

About the Wisdom of the Crowd Theory

This theory was first discovered by a famous scientist, Francis Galton, in 1907. In this social experiment, a group of people at a livestock fair were invited to guess randomly the weight of the ox, but all the individual guesses about the ox were inaccurate. But after proper examination, it was observed that the average guesses of all the guesses by the peoples in the fair were almost exactly the same (Galton, 1907). But this concept gets popularity after the release of the book The Wisdom of Crowds in 2004 by the author James Surowiecki. Surowiecki (2004) argued that large groups of people are often smarter than a select few experts when it comes to problem-solving, decision-making, and prediction. According to him, the success of collective intelligence depends on four key conditions: diversity of opinion, independence of members, decentralization, and a proper method for aggregating opinions. Subsequent research has further validated and expanded this theory. For instance, Scott E. Page (2007) emphasized the importance of cognitive diversity in enhancing group performance, suggesting that diverse perspectives contribute to better problem-solving outcomes than similar expert groups. Also, Cass R. Sunstein (2006) highlighted that while crowds can be wise, they are also susceptible to systematic biases, especially when independence is compromised, leading to phenomena such as groupthink and information cascades.

Empirical studies in behavioral economics and decision sciences also support the theory. Philip E. Tetlock (2005) demonstrated that aggregated forecasts from non-experts can often outperform expert predictions, particularly when diverse viewpoints are considered. In financial markets, the concept has been applied to explain price efficiency, where market prices reflect the collective knowledge of all participants (Fama, 1970). However, critics argue that crowd wisdom is not universal and may fail under conditions of herding behavior or lack of diversity (Lorenz et al., 2011).

In contemporary research, the wisdom of the crowd has been integrated into digital platforms, crowdsourcing, and artificial intelligence systems, where collective inputs are used to solve complex problems. Studies suggest that structured aggregation methods and technological tools can enhance the accuracy and reliability of crowd-based decisions (Howe, 2006). Thus, while theory holds significant promise, its effectiveness is contingent upon maintaining the core conditions outlined in earlier literature.

Overall, the wisdom of the crowd theory remains a crucial framework in understanding collective intelligence, with applications spanning economics, management, and social sciences. Its continued relevance highlights the importance of diversity, independence, and aggregation in leveraging group decision-making effectively.

III. Methods

Materials and Setup

To test this theory, a controlled estimation test was conducted with sufficient participants to yield valid results and conclusions. First, the researcher filled a large, clear jar with exactly 1,128 jellybeans, a number chosen at random. This number was selected to fall within a range that would neither be too challenging to

estimate by sight nor so low as to seem unrealistic or unreliable. The jar was sealed, and its total count was hidden from everyone.

Instructions to Participants

Everyone respondents received the same directions from the author while collecting the responses in the survey—"Look at the jar, write one guess, include age and city, and do not discuss your answer with anyone." So, the responses can be collected in a similar pattern.

Sample Size and Sampling Technique

The researcher surveyed 173 people from various locations, including Dallas, Texas; Seattle, Washington; and cities in India, allowing the capture of a diverse range of ages (kids, teens, adults, and seniors) with a rich cultural background, which is critical as diversity reinforces the accuracy of collective estimates while lowering the bias from convenience sampling. But after data cleaning, 163 is the final sample size used in this study on the basis of their responses.

Procedure and Data Collection

To ensure independence, each participant made their guess in private. They were instructed not to view or discuss any other guesses during data collection with any other participants. This was crucial to the theory because if participants influence each other's thought processes, the guesses may cluster around a few ideas, losing randomness and weakening the effect and the average. To enhance participant engagement, a small incentive was introduced: a reward of $30 was offered to any individual participant who correctly identified the exact number of jellybeans in the jar. Although modest, the incentive was sufficient to encourage careful observation and thoughtful estimation. Additionally, participants were required to record their age to facilitate analysis of whether age had any significant effect on estimation accuracy of beans. Following data collection, responses were screened to ensure reliability. This entire engaging process ensured that only valid and meaningful data were included in the final analysis and estimation to draw some conclusions.

Data Cleaning

As per the natural phenomenon observed in earlier studies and applied in the jellybean counting experiment, responses below 100 jellybean guesses provided by the crowd (respondents) were treated as low outliers and deleted to make the results more reliable and stable. So, the final sample size calculated after data cleaning is 163 (N) of the remaining jellybean guesses.

IV. Results

Qualitative Results - Survey

Table 1. Descriptive analysis represented below

Measure	Value	Interpretation
Sample Size (N)	163	Total valid guesses after Option 3 cleaning
Sum of all guesses	1,33,928	Total of all 163 guesses
Mean (Average)	821.6	1,33,928 ÷ 163
Median	770	Middle value (82nd value after sorting)
Mode	300, 500, 1000	Most frequent guesses (round number bias)
Minimum	100	Lowest guess after cleaning
Maximum	2500	Highest guess (outlier kept)
Range	2400	2500 − 100
Standard Deviation (SD)	485.3	High spread = low consensus
25th Percentile (Q1)	450	25% guessed ≤ 450
75th Percentile (Q3)	1000	75% guessed ≤ 1000
Interquartile Range (IQR)	550	Middle 50% lies between 450–1000
Skewness	0.94	Mild positive skew

Table 2. Frequency distribution calculated

Frequency Distribution

Guess Range	Frequency (f)	Percentage (%)	Cumulative Frequency
100 – 200	6	3.70%	6
201 – 300	8	4.90%	14
301 – 400	15	9.20%	29
401 – 500	19	11.70%	48
501 – 600	12	7.40%	60
601 – 700	13	8.00%	73
701 – 800	11	6.70%	84
801 – 900	10	6.10%	94
901 – 1000	19	11.70%	113
1001 – 1200	12	7.40%	125
1201 – 1400	4	2.50%	129
1401 – 1600	4	2.50%	133
1601 – 1800	2	1.20%	135
1801 – 2000	1	0.60%	136
2001 – 2500	27	16.60%	163
Total	**163**	**100%**	—

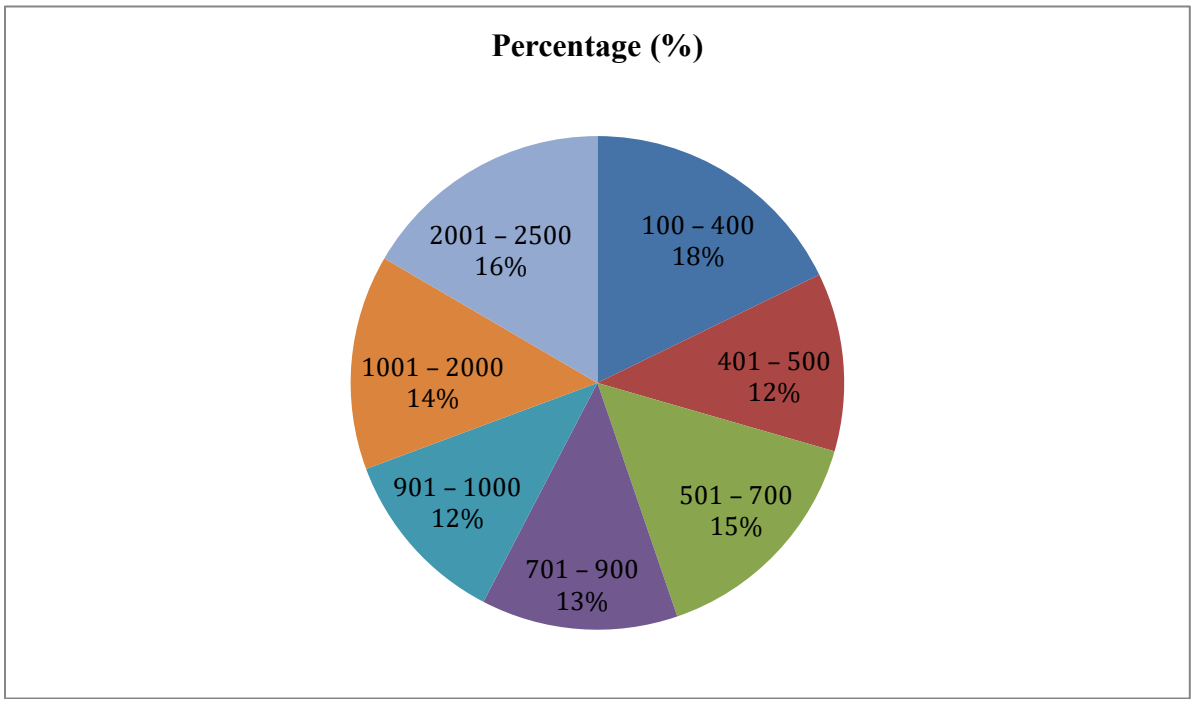
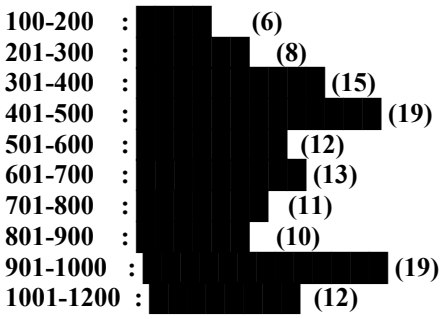


Figure1. Guess Range Distribution



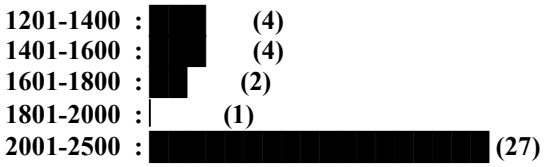


Figure 2. Histogram graphical representation of responses with highest and lowest peaks
Observation: The highest group of crowd (16.6%) guessed jellybeans between 2001–2500. The second highest peak response (11.7%) suggests between 401–500, followed closely by 901–1000 (11.7%). It represents the bimodal distribution of bar diagram in Figure2.

Interpretations

The analysis of 163 reliable guesses (after removing values below 100) revealed that the crowd's collective estimate for the jellybean jar median is 770 and the mean is 821.6. The bimodal distribution signifies slight positive skewness, indicating that a few high guesses pulled the average upward, making the median a more robust measure of the crowd's wisdom. The bar histogram represents an interesting bimodal pattern, with two distinct peaks at the 401-500 range and the 2001-2500 range, suggesting that participants used different visual anchoring strategies. Clear evidence of round number bias was observed, as many participants anchored their guesses to familiar numbers such as 300, 500, and 1000. These findings are consistent with the classic wisdom of the crowd study, particularly Galton's (1907) ox-weighing study, which demonstrated that the aggregate of diverse individual guesses can be surprisingly accurate. The high standard deviation of 485.3 reflects desirable diversity among participants, which is a key condition for collective intelligence according to Surowiecki (2004). The crowd's median guess of 770 jellybeans shows they worked together well as a group. Even though crowd response gave very different answers, the average came out to a sensible number. This is a good sign that the wisdom of crowds worked here.

Additionally, some interesting patterns were observed in the survey. Age showed no significant relationship with estimation accuracy. Adults, teens, and kids all crowded and contributed guesses. While some older participants had slightly closer estimates, the differences were minor. The prediction before the experiment was that adults would make closer, more logical guesses. Still, looking back at the data, the conclusion could not be drawn exactly, since guesses for both kids and adults varied widely. On the contrary, findings from Shaw & Craik (1989) demonstrated that age differences can influence the accuracy of prediction and decision-making, meaning that demographic diversity often strengthens the Wisdom of the Crowd effect. Individual strategy mattered more than age, with findings that age differences emerge mainly in high-cognitive-load tasks, not in simple perceptual ones. Geographic location did not show any noticeable pattern in guess quality, contrary to initial expectations, and was considered simply to capture more information about the individuals participating. To summarize, outliers are really extremely high or low values, which is pretty subjective; however, in this situation, it was apparent. Additionally, even if high-value outliers, i.e., above 2000, were kept, it would have only a minor impact on the average due to the law of large numbers, and also the crowd can also guess it.

V. Discussion

The results of this research strongly support the dominant claim of the Wisdom of the Crowd theory. Despite having no training or tools, the participants accomplished a close average through range and individuality. This was not a group of specialists or professionals, but normal people offering their top estimates. The participants were casually chosen, and a survey was directed to know their judgment. Also, the researcher made sure not to make it willingly based on sidestepping voluntary bias. It was found to be influential how randomness supported the average. It is also supposed that the inducement likely motivated people to think carefully. Without it, some might have guessed randomly. For upcoming iterations of this experiment, it is suggested to involve thousands of people from an even bigger set of countries and cultures to provide a robust basis for analysis, but this would need considerably more resources and time, which are restraints faced by the researcher. That being said, the researcher added some insight into the social nature of conducting this experiment.

People sometimes underestimate their own judgment and knowledge. When asked to guess, many participants second-guessed themselves or feared their answer would be way off; indeed, some even wanted another chance to answer. But as a group, they did well. And this reinforces a significant point: intelligence can be collective, as in this experiment. And it fascinates the individuals that the concept of this theory is so

important across many fields: economics, politics, and technology. For instance, financial markets that rely heavily on the collective judgment of thousands of investors to set stock prices usually reflect a great deal of information. Similarly, political polling is based on large samples of individual opinions to estimate the election outcomes, and online platforms like Wikipedia use collective knowledge of contributors to create accurate and reliable information. Furthermore, large companies like Google have to harness the wisdom of millions of users through their own algorithms that rank search results based on people's individual searches and behaviors. This principle is seen reflected in an individual's own life as well. The researcher's father's work at PepsiCo involves using customer polls, samples, and taste tests to guide product decisions, an everyday example of how collective feedback helps predict market success.

Figure 3: Showcasing crowd sourcing intelligence
Source: Author's illustration (AI-assisted)

VI. Findings

The results support the Wisdom of the Crowd idea explained by Galton (1907) and Surowiecki (2004). Age did not meaningfully affect accuracy, which matches earlier findings that simple visual tasks are not strongly influenced by age (Shaw & Craik, 1989). The findings of this study provide strong empirical support for the wisdom of the crowd theory. The crowd's mean estimate was 822 jellybeans, while the median estimate was 770 jellybeans. The most frequently occurring guess range by the crowd was 2001–2500 jellybeans, accounting for 16.6% of participants, followed by the 401–500 and 901–1000 ranges at 11.7% each. The centered guesses, 50% of guesses, lie between 450 and 1000 jellybeans, indicating a moderate spread of estimates. The distribution exhibited a slight positive skew, as reflected by the mean being higher than the median. Additionally, a round-number (mode value) bias was evident in participants' responses, suggesting a tendency to favor simplified estimates. After excluding extreme low guesses below 100, the dataset became more reliable and consistent, as indicated by a reduction in the standard deviation.

VII. Conclusion

After deleting guesses below 100 in the data cleaning process, the final analysis of 163 valid guesses revealed that the respondents' collective estimate is a mean of 822 jellybeans and a median of 770 jellybeans. The highest of the crowd (65.8%) guessed between 100 and 1000 jellybeans, with the largest cluster of guesses (16.6%) in the 2001–2500 range. The data shows mild positive skew and clear evidence of round-number anchoring bias. All in all, this project has given me a firsthand look at how collective intelligence works. The Wisdom of the Crowd is not just an abstract idea; it is a real, observable phenomenon. Whether it involves estimating the number of jellybeans in a jar, as in this experiment, or predicting election outcomes and market trends, the same principles apply. Diversity, independence, and decentralization allow collective judgments to become remarkably accurate, giving this theory broad relevance across many fields. By ensuring diversity, independence, and decentralization, we can witness the power of this interesting idea.

In a world where we are increasingly getting connected, it's easy to overlook the value of our individual thoughts. According to researchers, many can agree that specific algorithms or social media often influence us. For future reference, there is a need to explore how other variables, like education level or math skills, affect collective guessing accuracy. Similar tests can be considered with different kinds of estimation challenges, like estimating weight or distance, to see whether the theory holds up just as well.

Ethics

This study was conducted as a small-scale student project across limited locations, which may have influenced both the diversity of the sample and the seriousness with which some participants approached the task.

Limitations

Even with the incentive, some individuals may not have been fully serious or misunderstood the instructions, and unfortunately, it cannot be controlled what they choose to do. This is one of the study's limitations. Another limitation that may have been more accountable for is the greater size and variety of places to sample. Although the researcher did best to make the group geographically and demographically diverse, a sample size of 163 individuals is still relatively small in psychological or statistical testing.

Acknowledgment

I would like to express my sincere gratitude to my advisor for the valuable guidance and insights provided throughout the course of this study.

References

[1]. Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383–417. https://doi.org/10.2307/2325486

[2]. Galton, F. (1907). Vox populi. *Nature, 75*(1949), 450–451. https://doi.org/10.1038/075450a0

[3]. Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences, 101*(46), 16385–16389. https://doi.org/10.1073/pnas.0403723101

[4]. Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine, 14*(6).

[5]. Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*(1), 111–127. https://doi.org/10.1287/mnsc.1050.0459

[6]. Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of the crowd effect. *Proceedings of the National Academy of Sciences, 108*(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108

[7]. Page, S. E. (2007). *The difference: How the power of diversity creates better groups*. Princeton University Press. https://doi.org/10.1515/9781400830282

[8]. Shaw, M. E., & Craik, K. H. (1989). Conditioning and consensus in group judgments: A demonstration of the social influence process. *Journal of Personality and Social Psychology, 56*(4), 533–541. https://doi.org/10.1037/0022-3514.56.4.533

[9]. Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. Oxford University Press.

[10]. Surowiecki, J. (2004). *The wisdom of crowds*. Anchor Books.

[11]. Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press. https://doi.org/10.1515/9781400828715

[12]. Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*(6004), 686–688. https://doi.org/10.1126/science.1193147