The wisdom of the crowd versus the power of the swarm.

Master Thesis: Behavioral Economics

Track: Financial Economics

Name: Magoulas Charilaos

Student number: 616762

Thesis Supervisor: Francesco Capozza

Second Assessor:

Date: February 2023

Abstract

The present thesis revolves around Artificial Swarming Intelligence, the intention behind this study is to examine whether human-swarming methods prevail over the Wisdom of Crowd (WOC) approach when it comes to the accuracy of group estimates in general knowledge questions. Groups of individuals were tested on percentage type questions ("what percentage of the world is x"), first, in a blind-survey environment where every participant inserted his response in isolation. Afterwards, participants formed a swarm, and were once again asked to provide a response for the same question, while being able to observe and therefore be influenced by everyone else's responses in real time. This swarming process was found to improve the overall group accuracy of responses relative to the WOC approach, through the reduction of errors (error of the mean, mean error, and individual errors), based on a bootstrapping analysis.

Keywords: Group decision making, Collective intelligence, Artificial Swarm Intelligence, Slider swarm, Human swarming, Wisdom of Crowds.

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics, Erasmus University Rotterdam or Unanimous A.I.

Acknowledgements

I would like to thank Erasmus University for providing me with the proper guidance to construct this thesis, Unanimous A.I, for giving me the chance to test their innovative platform and train me on it, and my family, without which I would not have had the opportunity to complete this thesis and conclude my studies.

Table of Contents

1. Introduction

In everyday language we often refer to the idiom “two heads are better than one” to highlight an improvement in intelligence and decision making as a result of cooperation and consideration of multiple opinions. This phenomenon is something very focal to human experience and society, as is often argued that it can open the way for more optimal and ultimately errorless decisions in both everyday life as well as in more formal and high stakes settings. A result of synergies between varying informational levels and insights, this effect can and is utilized in multiple disciplines and fields of study for its emergent properties.

Aristotle was one of the first to point it out in his work “Politics” (Waldron, 1995), where he argued towards the idea by using potluck dinners as an example, he noticed that the perfect meal consists of the rare combination of all individual dishes, rather than any single one in particular. Since then, the theory has been used in other major fields, in politics, through the electoral process known as sortition (Dowlen, 2009), in the court of law through Condorcet’s jury theorem (Condorcet, 1785), in scientific consensus under open peer-review concepts (Millard, 2011), and even in nature where collective intelligence properties have been observed in bacterial colonies (Moreno-Gamez, 2022). Based on these intuitions, it comes to no surprise how such a versatile and vital process caught the attention of great thinkers and mathematicians alike. Although Aristotle was one of the first to observe it, it was Galton (1907) who introduced it to the field of statistics through his very famous ox weighting experiment, where at a country fair, he gathered 787 guesses from a crowd regarding their predictions for a certain ox’s weight. He then proceeded to aggregate these guesses and extract the median which he named “*Vox Populi*” (the voice of the people), what he found was that the median landed remarkably closer to the true weight of the animal, compared to the 787 individual guesses. After Galton, almost 100 years later, Surowiecki (2004) (based on Galton’s example) calculated the mean guess of the crowd’s estimate to find an even more accurate weight estimation for the ox. While other important papers have examined the phenomenon of forecast aggregation in the past (Bates and Granger, 1969, Clemen, 1989) it was Surowiecki who brought mainstream appeal to it through his books, as well as by coining a new term, *“The Wisdom of Crowds”* (WOC).

In its simplest form, WOC theory revolves around crowd intelligence and more specifically, how a group of individuals known as the collective can ultimately create a more sound and unbiased opinion (or forecast) compared to any individual or expert within the crowd itself. In this manner, the crowd can "amplify" the intelligence of its own members and produce more reliable results. This is done by aggregating group opinions, usually through the measurement of a mean/median, although additional statistical methods have been proposed, sometimes showing an even greater improvement in accuracy (Mannes, Soll and Larrick, 2014, Palan, Huber and Senninger, 2019). The primary mechanism under which this phenomenon is established, comes from the "cancelation" of individual biases, given a sufficient number of differentiating opinions. If we take as an example the case of Galton above, in his experiment, some individuals would have produced a really high estimation of the ox's weight, while others a considerably low one, through aggregation, these predictive errors will cancel or at least smooth each other out. If this process is applied to a scale with a large enough number of forecasts, we can ultimately get an estimate with highly reduced noise, which stands closer to the true weight of the animal. This idea behind forecast combination and the wisdom of the crowd approach has found a wide array of applications, from forecasting inflation rates (Zarnowitz, 1984), to predictions of election outcomes (Boon, 2012), stock market returns (Chen et al., 2014), and even to study gene regulatory networks in computational biology (Marbach et al., 2012).

The theory has become all the more relevant in recent decades, especially with the emergence of the world wide web, through which the possibility of collecting and aggregating a large variety of opinions becomes easier. Sometimes a cacophony of online viewpoints can be summarized through aggregation and create a harmonic outcome which holds great accuracy and unbiasedness. Another vital aspect of the world wide web is the interaction that occurs between individual users. This fairly recent interaction ability has brought tremendous change in terms of interconnectedness, where we are now able to participate and engage with the content on our screen. The main purpose of such an interaction is to serve as a digital dialogue between users which can exchange information and ideas more freely and conveniently. This interaction has become even more intense in the last decade, considering the rise of social media platforms such as YouTube, Facebook, Instagram, Reddit, etc., whose reach has expanded as society grows towards more data-centric paths. This aforementioned interactive aspect has been in the forefront of recent expansions in the WOC theory and has given rise to a new area of study, adjacent to it, such is the idea of swarming.

To properly understand the concept, we should retrace to the famous field naturalist Edmund Selous, who very eloquently in 1905 while bird watching, expressed the following on the movement of bird flocks: "*They must think collectively, all at the same time, or at least in streaks or patches — a square yard or so of an idea, a flash out of so many brains*" (as cited in Couzin, 2007). This is where the inspiration for the term "Swarming" came from, it was coined by Beni and Wang (1993), and it refers to a collective decision-making process, inspired by natural swarming events of biological systems such as insect colonies (Marshall et al., 2009) and animal groups (Couzin, 2008). In these types of systems, a number of autonomous agents or particles inside the swarm are interacting with their environment (consisting of other autonomous particles). An important prerequisite for swarming systems is the absence of a central authority (or node) through which power is exerted onto other agents/particles. If this holds true for the process, self-organization can emerge, and new properties of collective cognition arise and serve towards the completion of an objective, such as predator protection in flocks of birds, nest hunting in bees or decision making in humans. Thus, as opposed to the original WOC theory, in which Surowiecki assumed that for the aggregating effect to hold, there should be no interaction between agents, swarming disregards this assumption and even accentuates the interactive aspect in decision making or forecasting. This is where platforms like "UNU", an Artificial Swarm Intelligence (ASI) system draws inspiration from, utilizing the effects of human swarming, by exposing participants to these interactive/dynamic systems, it produces aggregate opinions (or predictions) on several different topics. By acknowledging the interplay between human agents, we can take advantage of certain qualities that derive from cooperation, disagreement, and conviction, with the final objective of constructing more flawless and accurate decision-making procedures.

Such collective systems are bound to become all the more relevant in the future, especially if we are to take into account the unprecedented experience of the Covid-19 pandemic since March of 2020, where companies and organizations alike, introduced remote work on a scale never seen before, a trend that is very likely to continue further in the future (Robinson, 2022). This highlights an even greater necessity for dynamic interactive systems such as swarming, which enables remote communication and decision making in businesses, without the need for physical presence (Metcalf, Askay and Rosenberg, 2019).

In this paper, the main interest lies on the idea of experimenting and putting to the test this new swarming theory and whether it could actually lead to a more accurate group prediction pattern as opposed to individual predictions, as well as compared to the more classical WOC theory (under the independence assumption). The main research question, in simplified terms, will revolve around the query of whether swarming is effective and whether it can ultimately lead to more errorless group judgements as other researchers have argued in the past (Rosenberg, 2015).

This paper's main innovation lies in the idea of whether swarms are more accurate when it comes to general knowledge questions on percentages, similar to the questions posed by Vul and Pashler (2008) in their paper. The second innovative aspect of this paper revolves around the brand new version of the swarming interface offered by Unanimous A.I, named "Slider-Swarm", which only one paper has tested thus far (Domnauer, Willcox and Rosenberg, 2022).

In order to assess the effectiveness of swarming, the prior mentioned platform "Slider Swarm" was utilized, to perform an online experiment which will be presented in the methodology section. Through this, the effect of swarming will be assessed based on absolute errors from the correct answers, similar to the methodology deployed by Vul and Pashler (2008).

The paper will proceed with the literature review which includes the latest and most relevant studies and research, as well as a historical rundown of swarming. Furthermore, the methodology and the experiment performed will be explained in detail. After the methodology, the examination and interpretation of the results will be attended. Next, the paper will proceed with a small discussion section where limitations of the study will be mentioned, as well as future research opportunities that could be performed in the field. Lastly, the study will conclude by providing a small summary of the findings.

2. Literature review

As mentioned before, WOC and swarming phenomena are based on the notion of collective intelligence when it comes to decisions or predictions. They involve accuracy and how an aggregation of (or an interaction between) group members can lead to more optimal opinions and errorless predictions.

Much research has been done on the field, starting off with the investigation of dynamic social interaction systems, with two pivotal papers, first by French (1956) and then by DeGroot (1974), after whom the French-DeGroot averaging model was established. It represents a focal step in the study of influence systems and social interaction dynamics on which swarming is based. The model studies social networks, in which feedback loops are present, under which members constantly update their beliefs to accommodate the opinions of different group members, to which they ascribe varying weights. This keeps on happening until some form of convergence is reached. The model examines these possible convergence paths for a system and how consensus among a group can be formed under the right conditions, such as aperiodicity, even though there might exist subjectivity of opinions.

Another important field in the development of swarming systems is that of cellular automata, introduced by von Neumann and Burks (1966). According to Wolfram (1984), "*Cellular automata are mathematical models for complex natural systems containing large numbers of simple identical components with local interactions. They consist of a lattice of sites, each with a finite set of possible values. The value of the sites evolves synchronously in discrete time steps according to identical rules. The value of a particular site is determined by the previous values of a neighborhood of sites around it.*" (p. 1). There have been several contributors to this field of cellular automata, such as John Horton Conway and his deviceful "Game of Life" (Gardner, 1970) which brought mainstream appeal to the theory, beyond solely academic circles. Another great contributor is Stephen Wolfram, he managed to solidify the knowledge available at the time by revealing ubiquitous statistical properties that govern automata structures, and by connecting the theory with that of dynamical systems as well as a more formal theory of computation (Wolfram, 1983). He also greatly expanded our knowledge by organizing cellular automata into four distinct categories based on the predictability levels of their final spatial outcomes given their initial state (Wolfram, 1984). Other processes that closely resemble special cases of cellular automata (Matache and Heidel, 2005) have been used to study social consensus and collective opinion formation in human networks, some examples are Boolean networks (Green, Leishman and Sadedin, 2007) and Brownian particles (Schweitzer and Holyst, 2000).

An extremely influential paper on the field of swarming, pertaining to the graphical aspect of it, is that of Reynolds (1987). In his paper, he managed to produce computer animations that mimic birds flock-like behavior. He achieved that by including a number of real-world flocking parameters in his own computer-generated simulations, thus paving the way for visual representation of swarms in human-swarming technologies. His approach incorporated behavioral analysis for producing the movement of flocks, which cleared the path for many graphical simulations of flocks, herds, and ultimately human swarms. Part of this analysis has to do with "*Collision Avoidance*" (between nearby flock members), "*Velocity Matching*" (of nearby flock members) and "*Flock Centering*" (tendency for nearby flock members to stick together). He also introduced the idea of "*Prioritized acceleration allocation"*, a system under which the flock decides which direction to move towards and how to avoid environmental obstacles, under the myriad of individual inputs that internally sway it, similarly to modern swarming platforms. Following Reynolds, other researchers based on his work, proposed new swarm-like models (Vicsek et al., 1995), as well as new attempts to quantify the movements behind flocking and swarming, with the most notable being those of Huth and Wissel (1992), Toner and Tu (1998), and finally Li, Lukeman and Edelstein-Keshet (2008).

Moving on, towards Beni and Wang (1993), who first coined the term "*Swarm Intelligence*" in a 1989 NATO hosted workshop for robots and biological systems. Their pivotal paper, which was based on research performed by Stephen Wolfram, investigated the idea of whether Cellular Robotic Systems (CRS) can prove to possess qualities of natural world "swarm intelligence". The concept of CRS represents a group of finite and independent robots working under a specified cellular space to achieve a certain task. Each individual agent/robot works autonomously and must cooperate with other robots to accomplish this predefined task. Under CRS, there exists no central authority or mechanism to synchronize and guide the independent robots, forcing them to work together for the detection of a solution. If in CRS, we were to replace robots with humans, we can observe a process which closely resembles that of human-swarming systems, with decentralization and cooperation among members serving as tools towards an accurate prediction or an optimal decision. This point is also shared by the authors who propose that the engineering of swarming and swarm intelligence could possibly be applied in a "bionic" manner, such as animal societies and insect colonies, on which human swarming is based, as it will be explained later on. The theory of "Swarm Intelligence", achieved even greater popularity, with the emergence of processes such as “Particle Swarm Optimization” (Kennedy and Eberhart, 1995) or “Ant Colony Optimization” (Dorigo and Gambardella, 1997) where individual agents, nodes, particles etc. were used to establish problem solving algorithms in computational simulations. These processes have also been applied to social settings (Kennedy, 1997), paving the way for new approaches to amplifying human intelligence by combining it with computational intelligence (Kirschenbaum and Palmer, 2015).

A major step in the theory behind human swarming involves the research of animal (or insect) swarms as decision making systems. One of the most prevalent swarming systems in the natural world are those of honeybees, with Camazine et al. (1999), Seeley and Buhrman, (1999), Seeley and Buhrman (2001), being some of the first papers to study decision making among honeybee hives in nesting spot selection. Soon after, swarming naturally cascaded towards other types of biological organisms with various internal complexities and environmental settings. For example, Conradt and Roper (2003) studied decision making in animal groups by establishing and contrasting two alternative systems, despotism (tyrannical rule) and democracy (communal or majority rule). Couzin et al. (2005) studied decision making in animal herds and how they manage to maintain cohesion on migration routes when varying levels of information for the “right path” exist among group members. Sumpter et al. (2008), studied the same phenomenon of how consensus is reached in flocks of fish and whether accuracy of decisions is positively related to group size. On the other hand, Couzin (2008), drew parallels between the mechanisms and intricacies behind collective decision making in animal groups and those that govern cognition. He proposed that individual animals interacting with each other, containing differing amounts of informational input, resemble neuronal assemblies, which allow for collective properties to emerge, something considered unattainable in isolating environments. Similarly, Marshall et al. (2009) investigated how cognitive decision making in primates resembles social consensus formation and the transfer of information that happens in insect colonies, and more specifically for nest-hunting tasks.

With so much research on swarming and so many attempts trying to deconstruct or quantify the concept, it comes as no surprise how eventually ideas and methods of human swarming came to fruition. For example, one of the first papers on collective decision making and consequently human swarming, is that of Dyer et al. (2007), which studied real time movement of human crowds. They performed an experiment in which they composed groups with eight participants each. In these groups, they divided members into two distinct categories, “naïve” and “informed” ones. Naive subjects were instructed to closely follow the group, while informed ones were instructed to lead the group towards a certain location, without using verbal cues or gestures of any kind. Through this, they wanted to mimic swarm movements found in flocks of birds and observe how human crowds can possibly resemble those, in evacuation scenarios of fire or other types of safety hazards. A good summary of swarming breakthroughs (at the time of the paper) was produced by Krause, Ruxton and Krause (2009) who offer a small rundown of relevant research that had been done on both animal and human swarming intelligence systems up to that point. Another key paper on human swarming is that of Rosenberg (2015), where he introduced the “UNUM” platform, for collaborative decision making in online environments. Through this, a small group of individuals would come together in an online setting and collaboratively answer questions posed by the moderator (researcher). Contrary to polling methods, this swarming system forced group members to cooperate and direct an animated puck (in a two-dimensional space) towards the answer that the group evaluates to be the correct one. Each member individually exerts a pulling force through his computer mouse onto the puck, tugging it towards the answer he/she deems correct. The result of all individuals simultaneously applying force to the puck in a dynamic fashion (resembling a digital dialogue) leads the group towards convergence and consensus. The research found that human swarms tend to settle on the correct answer more often, compared to individuals answering the same set of questions through polling/survey aggregating methodologies which pertain to the WOC theory.

This sort of technology has been utilized in various fields with promising results, for example Rosenberg, Hornbostel and Pescetelli (2018), found that swarms are much better at distinguishing between genuine and fake smiles in people, compared to when the same swarm members answer in isolation. This means that the swarm outperformed both individual answers as well as when those were aggregated, resulting in greater efficiency where detection of smile authenticity is concerned. An adjacent study by Rosenberg et al. (2018) examined whether small swarms (3 to 6 members) could better excel at the “Reading the Mind in the Eyes” (RME) task, compared to individual guesses. The RME is a test which involves cut-out images of different people’s eyes which participants try to assess in terms of emotion, for example a stare might convey anger, fear, desire, conviction etc., this test is often employed to proxy for “social intelligence” (Vellante et al., 2012). What the authors found was that swarming resulted in lower RME score error rates (15.29%) relative to averaging individual RME scores (31.54%), as well as when establishing a plurality vote criterion (25.94%), providing additional support for the notion that swarming amplifies a group's intelligence. This increased intelligence in groups was also presented by Rosenberg and Baltaxe (2016) who tested human swarming when it came to setting political priorities. What they found was that once a set of respondents formed a swarm, they were able to better distinguish between political issues that directly affected their everyday lives, as opposed to long term or broader issues that did not immediately impact them. Individuals also expressed that the priorities produced by the swarm, more realistically represented their individual priorities, compared to a simple survey aggregation (WOC) approach. Political decisions often relate to making the societally optimal decision, a theme which Rosenberg and Willcox (2018) examined. They ran an experiment, in which subjects were presented with three alternative personal monetary payoffs, one of which represented the "Social Optima" decision (unknowingly to the players) which maximized collective payoff for the whole group. Participants first got to vote individually for the payoff of their choosing (similar to a simple democratic voting process), without knowing other players' payoff values. Afterwards, they were placed in an artificial swarming intelligence platform, under which they now had to settle on a consensus decision. What the results showed was an 82% chance of the group converging on the socially optimal decision under swarming conditions, surpassing the other three voting methods (Plurality: 63%, Condorcet pairwise: 60%, Borda count: 58%) employed for the analysis of the individual votes.

On the business side of things, Willcox, Rosenberg and Schumann (2019) employed human swarming to examine sales forecasting events in fashion retailer entities. They once again tried to establish a comparison of traditional polling methodologies (WOC) and swarming, by ranking sweaters that the crowd (or the swarm) believe will attain higher net sales. They concluded that swarming systems outperform traditional survey aggregation techniques by producing more accurate forecast rankings when it comes to sales of different sweaters. Another business-related study was performed by Metcalf, Askay and Rosenberg (2019), who introduced an ASI platform for strategic business decisions, among small groups of colleagues, such as "Should we enter a new market?", "What product should we launch?" etc. They found that this new swarming way of approaching business related decisions enabled groups to locate primary and secondary goals more accurately and clearly, through means of anonymity (allowing everyone to state their opinion without repercussions) and convergence (by giving everyone's opinion equal power). From businesses to markets, Rosenberg, Pescetelli and Willcox (2017), conducted a human swarming-based experiment with the goal of predicting financial market outcomes. By gathering groups of self-identified "active traders", they elicited predictions on the future directional changes of major market indices relating to stocks and gold over the next three days. The measurements had to do with "*individual accuracy*", relating to the average prediction of all participants, "*crowd accuracy*" relating to the most popular prediction and "*swarm accuracy*" which measured the answers produced through swarming. The outcome of the study pointed towards an increase in accuracy for the group under swarming conditions, in all the four indices that were tested. The swarm managed to outperform both individual median guesses as well as the most popular answers. In some cases, the swarm produced an increase of 43% in terms of accuracy, compared to the median group accuracy where members acted separately. Another study by Rosenberg et al., (2021) this time on volatile assets such as "cult-stocks" (or "meme stocks") and cryptocurrencies, investigated whether Return on Investment (ROI) could be improved using ASI systems. Initially, participants were asked to provide their personal three-day price change forecasts for the assets in question, further on, they were instructed to either bet short or long in those assets, using a virtual bankroll. The individuals also gave their opinions on which assets will experience the largest price increase and decrease respectively. Afterwards, the same participants were allocated into human swarms, on an ASI platform, and produced a new set of forecasts for the same questions, this time by working together under swarming conditions. The experiment showed that over a nine-week period, the swarm managed to produce a cumulative ROI of 22.7%, which translates to a roughly 2.3% weekly return. This figure exceeded both the other two investment strategies' weekly returns of considering the median (0.96% ROI) or the most popular (1.6% ROI) picks.

In sports, and more specifically in American football, Rosenberg, Baltaxe and Pescetelli (2016), in their study, had a swarm of 29 randomly selected American football fans, compete against a survey/crowd of 469 football fans for predicting "Prop Bets" outcomes. The results of the study showed that the swarm outperformed the crowd by correctly predicting the outcomes of 13 out of 19 bets, while the crowd managed to get 9 correct answers. This difference led to a 21% improvement in accuracy when predicting betting outcomes for the swarm relative to the crowd, a very promising result, especially considering the size difference between the two. Crossing over to the other side of the Atlantic Ocean, Rosenberg and Pescetelli (2017) studied football-forecasting but this time in English Premier League (EPL) matches. They gathered groups of 25 to 31 participants and had them (over five weeks) make outcome predictions on 50 different EPL games (win, tie or lose), first as isolated players, and then collectively as human swarms, through the “UNU” platform. They compared the accuracy of the swarm with that of a survey by the same members and that of an EPL predicting supercomputer called SAM, finding that the swarm managed to outperform both by 17% and 8% respectively in terms of correct guesses for all 50 matches. The researchers concluded that an improvement was observed to the magnitude of 131% for when football fans formed a swarm as opposed to having them separately fill in a survey and then aggregating their responses. From the ball to the puck, Rosenberg and Willcox (2018) applied swarming methods in hockey, with the aim of examining whether human swarms could outperform Vegas betting markets, proxied by a famous online betting website. They compared the two in simple game forecasting as well as in identifying which games can be predicted with a high confidence of outcome. In both cases, swarming managed to outperform the market, and achieve an ROI of 170%, by turning 100$ to 270$ in a 20-week period, providing additional evidence for the utility and relevance of swarming. Similarly, Willcox et al., (2019) performed an equivalent study, this time in National Basketball Association (NBA) games. They gathered subjects which self-identified as “NBA enthusiasts” and had them predict outcomes for a total 238 basketball games by participating in a human swarm. The authors then compared their results in terms of forecasting accuracy to Vegas odds (extracted from an online sport betting website). What they found is that the human swarm managed to out-predict the Vegas odds market by 5%, a sizable improvement.

Related studies have been carried out in the field of medicine, where Rosenberg et al., (2018) recruited eight radiologists in order to assess a total of 50 X-ray images, for the probability of the patients having pneumonia. They first had to provide their individual estimate and then work together as a team, by being connected through a swarming platform, in order to conjointly settle on a probability. Moreover, deep-learning software for pneumonia diagnosing was used in the study, serving as a benchmarking index. Their results pointed towards an outperformance by the radiologist swarm, relative to both individual diagnoses as well as the software, based on three criteria, mean absolute error (from the “ground-truth”), binary classification (only considering X-rays with at least 50% diagnostic probability of pneumonia) and Receiver Operating Characteristic (ROC) analysis. A closely related study by Shah et al., (2021), explored the possible effect of introducing an ASI approach for radiologists, in the process of grading knee magnetic resonance (MR) exams for meniscal lesions. Initially, the authors conceived two different cohorts of resident and certified radiologists to individually grade the M.R images, while also extracting another set of grades for each cohort, using a majority rule. Next, the authors employed a meniscal lesion detecting AI software to produce its own grade. Finally, they transformed the cohorts into swarms, requiring the radiologists to provide a new set of grades, this time as collective entities. All the various grades (swarm, individual, majority, and software) were then benchmarked in terms of inter reader reliability on the basis of both clinical and radiological standards. The results showed that swarming grades (in both cohorts) surpassed all three alternative grading methods, further supporting the potential of swarming systems in the field of medical diagnoses.

From the field of medicine to the skies of the U.S Air Force, Befort et al., (2018) introduced swarming technology on the question of whether collective intellect can improve accuracy on pilot workload and usability ratings. After performing virtual missions on a Boeing simulator, a group of six pilots had to provide individual ratings for workloads of designated tasks, through Likert and Bedford scales. Afterwards, the group of pilots were placed on a swarming platform, in order to collectively assess the workload of the same tasks. The authors included two "Subject Matter Expert" (SME) pilots to provide personal assessments on the tasks performed by the group of pilots, as a comparison metric. Results pointed towards greater accuracy in ratings when pilots worked in swarms as opposed to individual ratings, with swarming assessment landing closer to those of the two (more objective) SME pilots.

Moving on to one of the most recent studies, we redirect our attention towards Domnauer, Willcox and Rosenberg (2022), a pivotal paper in its own term, as it represents the first instance for the utilization of a new version of the UNU swarming platform known as "Slider-Swarm", the platform that this current paper is also based on. The authors performed an experiment in which participants had to assess the likelihood that a virtual "bag" had a majority blue or red colored marbles. After eliciting individual responses on the likelihood, the participants were then logged onto the slider-swarm platform, in order to provide a new estimate, this time collectively. What the results showed was a reduction of the mean brier score in the case of swarming compared to both individual responses, as well as WOC methods such as eliciting the mean of all estimates. Another study from which this paper draws great inspiration, in terms of its methodology and its statistical analysis, is that of Vul and Pashler (2008), who examined the "Crowd Within" phenomenon. This phenomenon argues that crowd diversity of opinion can be internally reproduced in a single person, which occurs with every additional guess that one makes, the average (of all past guesses) will land closer to the truth compared to any one in isolation, similar to WOC theories. Questions used in this current study were greatly inspired by those posed by Vul and Pashler in their own paper.

Considering all the above, the main study of the current paper forms around the absolute errors of the participant's guesses (from the ground-truth) under both WOC and swarming conditions. The first null hypothesis (H1) involves finding no improvement in the error of the group estimate under swarming conditions. The alternative hypothesis has to do with the idea that there actually will be a decrease of error in the group response under swarming, much like Domnauer, Willcox and Rosenberg (2022). Furthermore, the second null hypothesis (H2) involves finding no improvement in mean errors under swarming conditions, compared to the mean error of a group taking a simple survey. In this case, the alternative hypothesis aims to find a statistically significant decrease in mean errors when swarming is enacted.

In simpler terms, for the first hypothesis, step 1 will revolve around eliciting the mean response for each round and then step two, eliciting the error from the truth, to compare for which round it is larger. For the second hypothesis, step 1 and step 2 are reversed, where first, the individual errors for each response will be elicited, and then, for the second step, their means for round 1 (WOC rules) and 2 (swarming rules) will be extracted. It will then be feasible to compare these two mean errors and observe if there is any improvement under swarming conditions in the second round.

3. Methodology

Through this research, the intention of this paper was to expand on the topic of crowd accuracy and more specifically, on whether "Swarms" can outcompete "Crowds" when it comes to providing estimates in the form of percentage type general knowledge questions. The thesis was conducted under guidance and with the approval of the Erasmus Universiteit Rotterdam ethical board.

The nature of the research was experimental, it was conducted by using an online survey tool called "Slider Swarm" managed by Unanimous A.I. The design of the experiment went as follows; Participants had to provide an estimate as an answer on 10 different questions, based on previous relevant work by Vul and Pashler (2008). Every single one of those questions had to be answered in terms of a percentage, so the questions would always begin with the phrase "What percentage of". For each of those questions there were three phases that participants had to go through, in a linear order, meaning that only after all of the phases of a question were finished, would the experiment proceed with the next one.

The first phase was called "THINK TIME" and was introduced to the experiment so that participants had some time to fully read and digest the questions before getting to respond. This thinking time was 5 seconds long, since studies have shown this to be an optimal amount of thinking time for higher quality responses (Rowe, 1974 and Stahl, 1994). Next, the second phase was introduced, the "Individual Deliberation" phase (figure 1), during this, participants were asked to provide their first estimate/guess for the question posed, with a time-limit of 25 seconds. Responses were automatically "locked in" when the timer ended. As soon as this phase was over, the third and final phase began. In this "Group Deliberation" phase (figure 2) participants had to input a second and final guess for the same question. This phase included a treatment effect compared to the last one, in this round, participants were provided with a live feed of other players' guesses, thus creating a dynamic feedback loop which further influenced individuals which in turn influenced group sentiment and so on. The final equilibrium is reached when there is no further update of opinions or, in this case the timer runs out. Every player had access to everyone else's answers without exception, they could readjust their guesses (again within a 25 second countdown) based on the sentiment of the swarm. For both phases, participants were instructed to move at least 2% from their starting position in either direction so to evoke active participation and prevent idleness. They were informed that should they not follow this guideline; their answer would not be counted. The platform elicits its own results after the experiment is done, in those results, it actually excludes the observation of participants who did not follow the 2% rule. Even though this is the case, the platform still records their data, which is why they were not discarded, and it was viable for those to be included in the statistical analysis of this current paper.

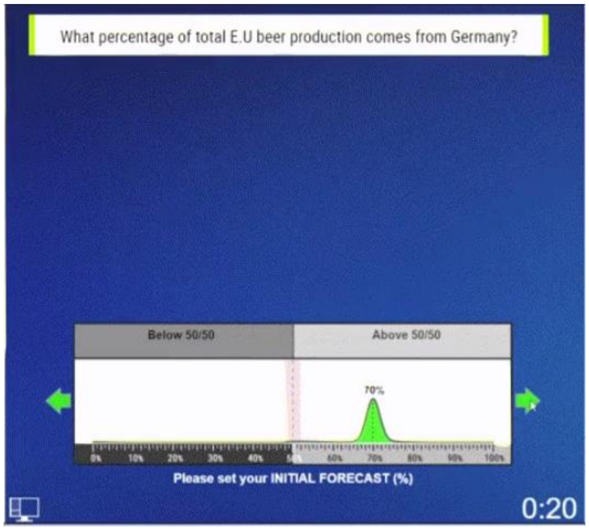


Figure 1: Individual deliberation phase (control condition)

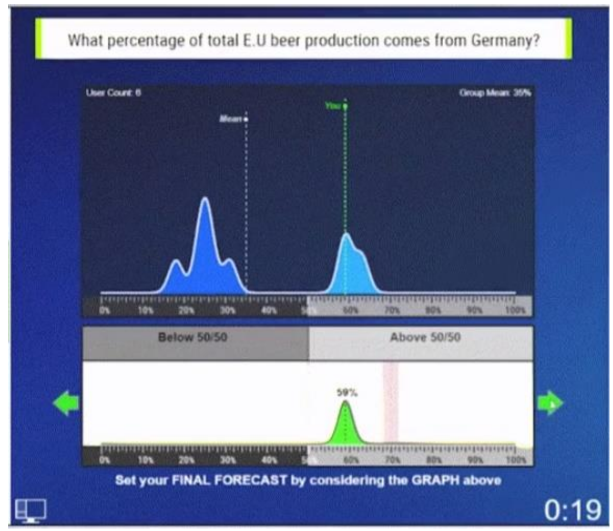


Figure 2: Group deliberation phase (treatment condition)

The variables of interest here have to do with the errors of the participants and how far off they were from the correct answers. In order to extract any possible effect of swarming on the control group, an initial comparison will be performed, between the error of the mean guess of the group in the individual deliberation phase (WOC approach) with the error of the mean guess of the group in the group deliberation phase (swarming approach). Next, another comparison will be performed, between the mean error of the participants' guesses in the individual phase with that of the group deliberation phase. The goal here is to observe any variation on the accuracy of the mean guess between crowds (independency among participants) and swarms (dependency among participants), as well as in the mean error of the participants' guesses in the first and second phase, with the final goal being to reveal any potential improvement on accuracy from swarming. This will be done for all 10 different questions. The variables of concern in this paper are independent and account for within variation since the same set of participants constitute a control condition (Individual deliberation/ WOC) and a treatment condition (Group deliberation/ Swarm).

Participants were gathered through my social circle, friends, family, and acquaintances who were willing to join the swarming session from their own laptops/computers and contribute towards my thesis. Apart from the restriction that participants should be over 18 years of age and that he/she needs to have sufficient command of the English language (in order to understand the instructions given), there were no further limitations on who could join the experiment.

On sample size, in order to roughly estimate the number of participants needed for statistical power, the program G-power was used, the settings were set as the following:

- Statistical test: “Means: Difference between two dependent means (matched pairs)”.
- Test family: “t tests” as the standard deviation of the parent distribution was not known.
- Type of power analysis: “A priori: Compute required sample size – given a, power, and effect size” in order to find a sufficient sample size for the experiment beforehand.
- Tails: “two” and also “one”, so to consider the possibility of any effect taking place, even an increase in error under swarming, and also stay true to focal point of this study which orientates around the idea that swarming in the group deliberation phase contributed towards a reduction and not merely a change of the error of responses.
- Parent distribution: “Normal”
- Effect size dz: “0.5”
- α err prob: “0.05”
- Power (1 - β err prob): “0.95.”

The output for the required sample size was 45 (one-tail) and 54 (two-tails) for normally distributed data (t-test matched pairs) and 52 (one-tail) and 62 (two-tails) for non-normally distributed sample data (matched pair sign-test) following the “15% rule” of Lehman (1998) for non-parametric tests. Additionally, on their paper, Domnauer, Willcox and Rosenberg (2022), used 3 groups of 30 to 36 participants each, using the same swarming platform, while Rosenberg and Baltaxe (2016), in their paper, used a swarm with a size of 43 individuals. Finally, other similar studies (Rosenberg, Baltaxe and Pescetelli, 2016) have shown that even with a sample of just 29 members individuals in a Swarm formation outperform themselves (compared to a WOC independence condition) by 1.72 standard deviations, in terms of accuracy. If a 0.5 effect size is replaced with 1.72 in G-Power the output results in a sufficient sample size of 6, which is considerably less than this paper’s sample of 48. Considering all the information above, the number of observations for all questions, which stands at 48 individuals, is evaluated as sufficient for properly assessing statistical significance under the current hypotheses.

3.1 Sample description

The invitation for the experiment was sent to 60 individuals in total, of which 48 joined. Out of those 48, 49% were males (22) and 51% females (23), additionally, around 7% (3) did not consent on filling the demographic questionnaire, so although they are included in the experiment they are not accounted for in figures 3 and 4 below. When it comes to the age of the participants, as we can see on figure 4, most individuals are between 18 and 44 years old (roughly 78%), with the youngest being 20, while the rest are over 45 years old, with the oldest being 66. The sample's standard deviation is 13.6, the mean age of the sample is 34.4 years old, and the median is 29, which stands very close to the global median age of 30 (Ritchie and Roser, 2019).

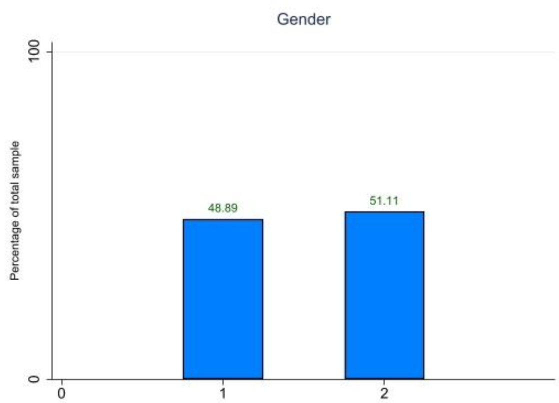


Figure 3: Gender demographics cohorts of sample in terms of Percentage (1 = male, 2 = female).

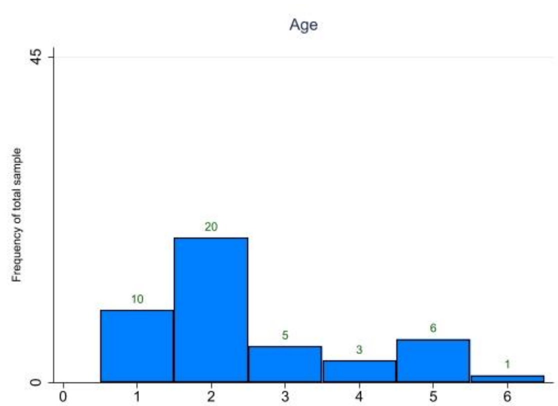


Figure 4: Age demographics cohorts of sample in terms of Frequency. The numbers correspond to the following age groups, 1: 18-24, 2: 25-34, 3: 35-44, 4: 45-54, 5: 55-64, 6: 65-74.

The experiment took in total close to 22 minutes and there was no compensation for the participants except for my personal gratitude and appreciation. Subjects were informed that the experiment represented “a fun little general knowledge quiz” where they would have to answer a series of 10 questions. They were also provided with sufficient information regarding the nature of the platform, the ethical concerns, contact information and instructions for the experiment. Regarding the ethical issues, participants were informed that the data would be shared with the parent company Unanimous A.I for research purposes and that they hold the right to quit the survey at any point by exiting the browser window. Lastly, participants were also informed on the date and time that the experiment will take place, while also receiving a final email for debriefing purposes (after the completion of the experiment) which included the correct answers, an explanation of the experimental goal as well as a final thank you note. All of the aforementioned information can be found in greater detail in the appendix section of this paper.

3.2 Statistical analysis

The first part of any statistical analysis is to examine the sample and hunt for potential outliers which could affect the outcome of the experiment. Due to the different nature of the questions posed, input that could be considered extreme for one question might be reasonable for another. To deal with this, the best course of action would be to manually remove them while considering a number of different factors indicative to the context of each question. For each question and round I would consider:

1. a graphical method, using quantile graphs and histograms to detect outliers,
2. a statistical score estimator such as the Median Absolute Deviation (MAD) (Leys et al., 2013, Maas, 2019), as it serves both normally and non-normally distributed data,
3. the logical feasibility of an answer, relating to common sense,
4. and whether a certain participant with an “extreme” value in the individual deliberation phase, significantly reconsidered his response (at least 2%) in the group deliberation phase towards a “less extreme” value.

As an example, I managed to spot a response of 100% to the question “What percentage of the planet is covered by water” through the use of relevant graphs. After that I would run a Shapiro-

Wilk test to check for the normality of the question replies in both the first and second phase individually as this is important for calculating the MAD score (Basharat, 2017). I would then proceed by producing a MAD[1] score on Stata, for all observations including the potential outlier of 100%. Next, I would contemplate whether this extreme value of 100% logically made sense and whether there could be any misinterpretation with the question from the participant's side, which could justify such a value. Lastly, I would check whether the subject reconsidered his response in the group deliberation phase towards a less extreme and more logically coherent value, in this case anything less than 100%.

When deciding on removing an outlier, for example in the first round, and the outlier did not follow rule 4 above, I would also remove it in the second phase (and vice versa), so to always end up with the same sample size between the two different rounds of each question. After that, I would manually impute these former outliers, now missing values, with the mean of their original sample. This was done so as not to affect the first and focal hypothesis of this paper, which refers to the error of the mean guesses (H1 below). The preservation of accuracy in the mean results is an equivalent preservation of accuracy in its error from the truth, thus unaltering the effect of swarming. It should be noted here that the maximum amount of missing values replaced in a single sample were 2 out of 48 (4% of the sample), so although this switch might introduce some slight bias onto the other two hypotheses, it would be of minimal impact.

At this point, it is crucial to note that accuracy here is used interchangeably as error, and they both refer to the absolute deviation (or distance) of a certain percentage/answer from the ground truth.

3.3 Hypotheses

3.3.1 Group error effect (H1)

The paper's main goal is to study the effect of swarming on group accuracy, through errors, relative to the more conventional Wisdom of Crowd approach. To do this, the appropriate variables must be formed first, which will help distinguish and reveal any potential effect. Thus, the mean

[1] Rejection criterion used was 2.5 times the metric for M.A.D. For eliciting the scale factor "b" the 0.75 quartile value was used, where in case of a normally distributed response sample "b" equals 1.4826 (Leys et al., 2013).

response of the participants in the first round needs to be identified (for all 10 questions). In the first round, all of the four assumptions for the WOC effect were held, such as diversity of opinion, decentralization, a valid aggregation mechanism and most importantly that of independence among group members (Chafetz, 2005). After eliciting the mean, next comes calculating its deviation from the truth. Subsequently, the same calculations are performed for the second round, so to estimate the new mean and new error, which represents the swarming technique, where although the first three WOC assumptions hold, individuals do not act independently of one another.

Due to the assumption of independence (of differences) and normality not holding, it is not recommended to run a standard matched pair t-test for the means. Thus, the ideal course of action here, based on a similar analysis by Domnauer, Willcox and Rosenberg (2022), is to run a matched t-test bootstrap analysis for 1000 repetitions, comparing the two means, so to establish that their difference is statistically significant. After this, it will be possible to compare the error of the mean response in both rounds and conclude which technique fared better, and most importantly whether swarming was more or less effective, leading to a reduced error of the mean. In this sense, the first hypothesis is formed as follows:

Hypothesis 1: There is no difference between the means of the two rounds.

This means that a rejection of the hypothesis will provide evidence of a statistically significant difference of the means and therefore group accuracy, between swarming and WOC conditions. A failure to reject means that swarming does not necessarily produce differing group accuracy, compared to the more traditional wisdom of the crowd approach.

3.3.2 Mean Error and individual error effects (H2 and H2.1)

For the second hypothesis, mean errors will be collated for the first and second round. To do so, a new variable must be constructed, which measures the error for each individual response in each round and for every question. Thus, the deviation variables are established as: "**INACCscoreji**", where "j" can be replaced by either 1 or 2 to dictate the round of the question that is referred to (round 1 = individual deliberation, round 2 = group deliberation) and “i” can be replaced with a word, such as in the table below, to dictate the question that is examined.

Error round 1.	**Question**	**Error round 2.**
INACCscore1car	*What percentage of E.U population owns a car?*	INACCscore2car
INACCscore1water	*What percentage of the planet is covered by water?*	INACCscore2water
INACCscore1empire	*What percentage of the world's population was under British Empire rule?*	INACCscore2empire
INACCscore1left	*What percentage of the world’s population is left-handed?*	INACCscore2left
INACCscore1urban	*What percentage of the world’s population lives in urban areas?*	INACCscore2urban
INACCscore1tobac	*What percentage of the global population are tobacco users?*	INACCscore2tobac
INACCscore1cov	*What percentage of the global population is estimated to have had covid-19?*	INACCscore2cov
INACCscore1christ	*What percentage of the global population is of Christian faith?*	INACCscore2christ
INACCscore1energy	*What percentage of E.U energy consumption comes from renewable sources?*	INACCscore2energy
INACCscore1rich	*What percentage of global wealth is owned by the richest 1%?*	INACCscore2rich

Table 1: Variables indicating the deviations of participants’ guesses from the correct answer (in absolute terms) for both rounds of a question.

The analysis here consists of 4 steps:

1. The mean of the errors above will be calculated.
2. The difference of the mean errors between the two rounds of each question will be extracted.
3. A matched pair t-test bootstrap for the difference of means of the variables in table 1 will be executed, in order to examine the statistical significance of said difference.
4. Finally, a one-sided matched pair sign-test will be performed, as an additional sub-hypothesis, to further explore the effect of swarming on individual guesses. The sign-test

was chosen, as the error variables (table 1 above) do not adhere to a normal distribution, a prerequisite for a standard matched pair t-test to be performed. Even though the sign-test also assumes independence of observations (something that is not valid in the group deliberation phase), I believe it still holds useful insights (supplementary to the two main hypotheses) for showcasing the ratio between individuals with reduced and those with increased or unchanged errors.

Continuing with the second hypothesis, under the matched pair bootstrap test for the mean errors. In simple terms, the second hypothesis will be the following:

Hypothesis 2: The mean errors between the two rounds do not differ.

If the null is rejected, the idea that the difference elicited in step 2 above is statistically significant can be supported, and therefore the notion that swarming did affect individual accuracy, holds validity (improved or not, will depend on the observed difference). If on the other hand, the null is not rejected, it would mean that there is a high possibility that the difference between the mean errors could be zero and thus swarming did not necessarily affect individual guesses, compared to the WOC approach.

Moving to the sub-hypothesis, through the sign-test, the goal is to observe which sample of errors between the first and second round is higher (by examining their ratio), and thus, whether swarming actually improved individual accuracy. The hypothesis here will be formed as follows:

Hypothesis 2.1: The median of the differences between the errors of the first and second round is zero.

If the null is rejected, it will provide support for the statistical significance of the results, meaning a big majority of individual errors have indeed decreased in the second round under swarming conditions. If the null is not rejected, this will mean that the individual errors between the two rounds are not statistically different and thus swarming did not necessarily lead to an improvement of individual accuracy, compared to the WOC approach.

To quickly summarize, for the first hypothesis (H1), step one will revolve around finding the mean response for each round and then step two will be to elicit their error from the truth, so to compare for which round it is larger. Finally, the significance of this difference will be examined, through bootstrapping. For the second hypothesis (H2), step one and step two are reversed, where the individual errors for each response will be calculated, followed by the second step, elicitation of the mean error for round 1 (WOC rules) and then 2 (swarming rules). Finally, the significance of this difference will be examined again by bootstrapping. It will therefore be possible to compare the mean errors between the two rounds and observe if there is any improvement under swarming conditions in the second round. Finally, under the sub-hypothesis (H2.1), the ratio/proportion of positive and negative individual errors will be examined for each question, between the two rounds through a sign-test. In the next part of the paper, the results will be showcased.

4. Results

Results were elicited for each question posed in the experiment, firstly (H1), the error of the mean was produced in the first-round, where individuals act separately (similar to a blind survey), representing the WOC approach. Then, in the second-round, where individuals interacted through the slider-swarm interface, representing the swarm approach. Secondly (H2), the mean error of individuals guesses was computed, in both rounds, to observe potential improvement. Lastly (H2.1), the individual guesses were contrasted in matched pairs to observe individual improvement. To summarize, the hypotheses refer to testing for the effect of swarming on the error of the mean (H1), the mean error of individual guesses (H2) and finally individual guesses in matched pairs, through the median of their differences (H2.1), with the ultimate aim to find a significant (statistical and possibly economical) reduction on error levels.

4.1 Hypothesis 1

For the first hypothesis (H1), the results for the difference between errors of the means are presented in table 2 below:

Question and correct answers (%)	*1st round mean (WOC) (%)*	*2nd round mean (Swarm) (%)*	*WOC error (pp)*	*Swarm error (pp)*	*Net effect (pp)*
*1. What percentage of the E.U population owns a car? (**56**)*	62.49	63.55	6.49	7.55	+1.06

Question and correct answers (%)	*1st round mean (WOC) (%)*	*2nd round mean (Swarm) (%)*	*WOC error (pp)*	*Swarm error (pp)*	*Net effect (pp)*
*2. What percentage of the planet is covered by water? (**71**)*	71.02	73.26	0.21	2.26	+2.23
*3. What percentage of the world's population was under British Empire rule? (**23**)*	39.75	32.33	16.75	9.33	-7.42***
*4. What percentage of the world's population is left-handed? (**10**)*	25.39	23	15.39	13	-2.39*
*5. What percentage of the world's population lives in urban areas? (**56**)*	58.92	61.44	2.92	5.44	+2.52
6. What percentage of the global population are tobacco users? (***22**)*	40.5	39.02	18.5	17.02	-1.48
*7. What percentage of the global population is estimated to have had covid-19? (**44**)*	61.48	63.6	17.48	19.6	+2.12
*8. What percentage of the global population is of Christian faith? (**31**)*	39.3	35.3	8.3	4.3	-4***
*9. What percentage of E.U energy consumption comes from renewable sources? (**22**)*	32.5	32.5	10.5	10.5	0
*10. What percentage of global wealth is owned by the richest 1%? (**46**)*	48.83	47	2.83	1	-1.83

*Table 2: Error of the mean guess. Error is the deviation of an answer from the ground-truth, in absolute terms, where: WOC error = |correct answer – WOC mean|, Swarming error = |correct answer – Swarm mean| and Net effect = (Swarm error – WOC error). All figures in the first three columns are in percentages, while the latter three are in percentage points. All figures have been rounded to two decimal points; due to this rounding some figures may not exactly match. Here: * = 5%, ** = 1%, *** = 0.1% significance level achieved.*

As shown in the table above, results are mostly mixed, but point towards a considerable effect of swarming on accuracy. Although in some questions the crowd had a slight edge over the swarm, the swarm managed to mostly outperform the crowd. Overall, there is a pattern of improvement under swarming in 5 out of 10 questions, while it did not show signs of improvement on the rest 5, with one being a tie. It is important to note here that 3 out of 5 swarm improvements were found to be statistically significant, while no question in which the swarm performed worse or equal showed statistical significance.

In more detail, the question in which the swarm produced the greater increase in accuracy, was question number 3, were the swarm managed to outperform the crowd and reduce its own error by 7.42 percentage points. This pertains to a relative decrease in error of 44.3%, a result found statistically significant at the 0.1% level. Additionally, the swarm, with a net error of 9.33 percentage points, managed to outperform 62.5% of the total sample in in terms of accuracy, on the first round, while the crowd managed to outperform only 50%, these results are graphically presented in figure 5 below.

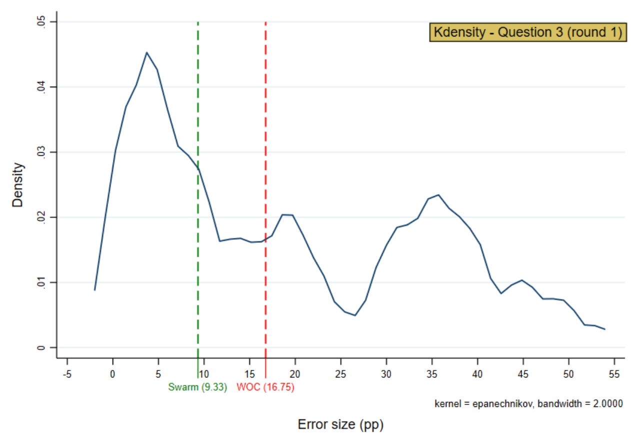


Figure 5: Kernel density plot of individual round 1 errors (q. 3). Represented also, the error for WOC (red) and Swarming (green) techniques.

Similarly, to a lesser extent, the swarm outperformed the crowd in question number 4, where error was reduced to by 2.39 percentage points, this pertains to a relative decrease in error of 15.53%, a result that also proved to be statistically significant at the 5% level. Here, the swarm managed to outperform in terms of accuracy, 46% of the total sample in round 1, while the crowd outperformed 41%, these effects are graphically presented in figure 6 below.

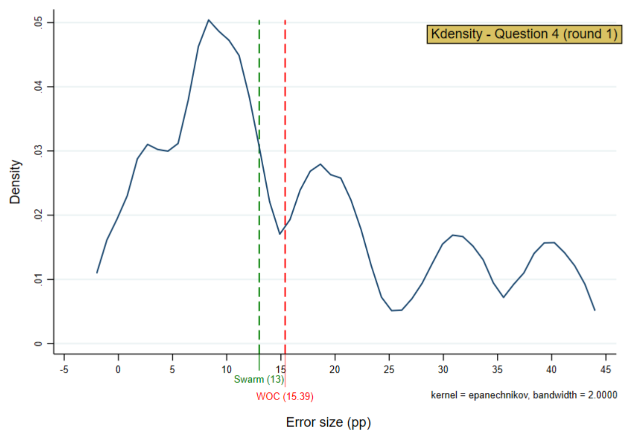


Figure 6: Kernel density plot of individual round 1 errors for question 4. Represented also, the error for WOC (red) and Swarming (green) techniques.

Finally, the swarm managed to outperform the crowd on question number 8, where error was reduced by 4 percentage points, this pertains to a relative decrease in error of 48.19%, an effect statistically significant at the 0.1% level. The error of the swarm, at around 4.3 percentage points, was smaller than the individual errors of 77% of the first-round sample, while the error of the crowd at 8.3 percentage points was smaller than 62% of individual errors. This result is graphically presented in figure 7 below.

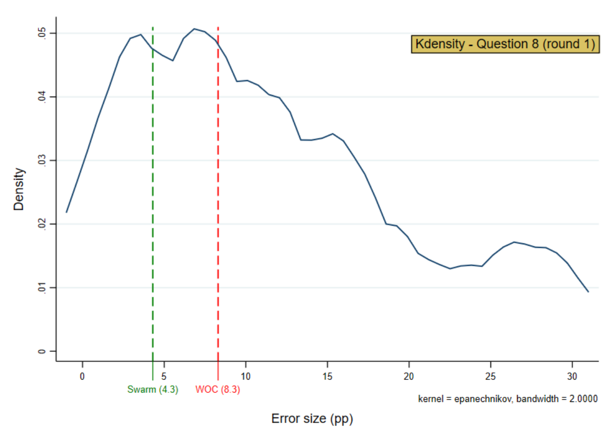


Figure 7: Kernel density plot of individual round 1 errors for question 8. Represented also, the error for WOC (red) and Swarming (green) techniques.

Another interesting result is in question number 2, where the crowd managed to achieve an error of 0.21%, a negligible figure, while increasing its error under swarming, to 2.26%. On the one hand, this could be explained by the effect not being statistically significant. On the other hand, it could also be due to the instruction given to participants, to “move at least 2% in either direction for your answer to register”, which could produce a situation where although the answer was well known throughout the group members, they had to deviate from the truth as a swarm (roughly 2 percentage points difference), in order to abide by this rule. This could present one of the drawbacks of the platform, where if a very well-known question is posed by the moderator, the swarm could deviate further from the truth rather than verge towards it, under this instruction.

This second theory is also supported based on the histograms of the question, presented in figure 8 below, where it is clearly displayed that in the first round, there is high concentration of individual guesses around the mean (identical to the truth). On the other hand, there is a higher degree of dispersion around the mean in the second round. This could be a sign of participants being forced to readjust their guesses (by 2%) in either direction (left or right), in order to abide by the instructions given, even though they already held an accurate estimate for the question. Another trend which can asserted in the graph, is the highest density bar (at the 71% mark), where before, it settled at roughly 0.155, which means that 15.5% of the group landed extremely close to the mean/truth, something that is indicative of the question being relatively common knowledge. In contrast, after swarming (round 2), the density bar close to the correct answer (71%) more than halves to 0.7 or 7% of the sample. The new highest density bar (at the new mean) is close to roughly 0.12 , which means that 12% of the group were close to the mean, showing greater dispersion than round 1, as participants might have moved towards either direction.

This is also supported by a more algebraic approach for dispersion, such as the standard deviation. The return of Stata for the standard deviations for rounds 1 and 2 respectively are 12.96 and 7.89. The standard deviation in the second round is much lower, which seems to be counter intuitive for the 2% rule theory mentioned above. This might be the case here, due to extreme values (more common in the first round) affecting the standard deviation. So, what would happen if we were to only account for values in a range close enough to the mean. To do this, by first performing a Wilkinson-Shapiro test, Stata provides us with a rejection of normality for both rounds, and thus the best plan is to use the MAD score (instead of a Z-score), taken from the outlier analysis in section 3.2 above. For a rejection criterion 2 times the MAD of each round, a threshold value of 2*5.77 = 11.54 (first round) and 2*3.87 = 7.74 (second round) is found. If these cutoff points are applied around the mean, for both rounds, in a manner such that every $observation < (mean - threshold\ value)\ or\ observation > (mean + threshold\ value)$ is dropped, new standard deviation levels for each round are elicited, 4.42 (round 1) and 5.03 (round 2). Although the two standard deviations are now closer, the ranking order has changed, where the more dispersed distribution now refers to the second round. This could be an additional sign of the 2% rule having an effect by spreading estimates further away from the truth (either left or right), in the case of a very common knowledge question. It should be noted here, that for this particular question, 37.5% of the sample (18 out of 48) had a readjustment size of exactly 2% (in either direction/absolute terms), the highest among all questions. This could potentially signify that those participants "were forced" to move under the rule, rather than out of genuine social influence.

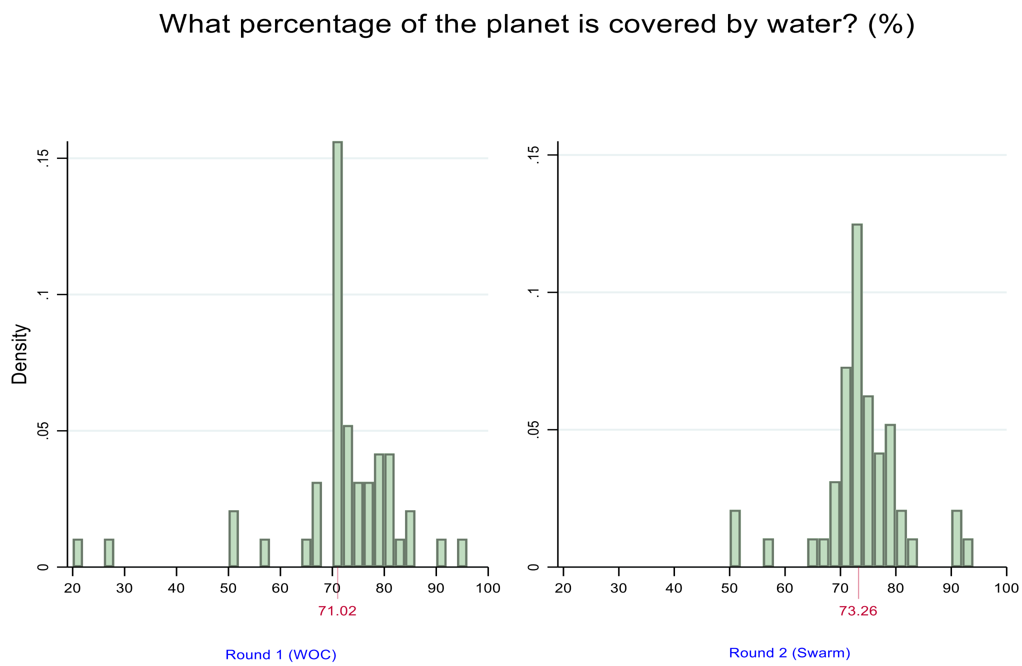


Figure 8: Histograms for question number 2, round 1 (left) and round 2 (right). Colored in red are their means.

4.2 Hypothesis 2

For the second hypothesis (H2): the results for the difference between mean errors are presented in table 3 below:

Question and correct answers (%)	*1st round mean error. (WOC)*	*2nd round mean error. (Swarm)*	*Net effect*
*1. What percentage of the E.U population owns a car? (**56**)*	13	12.23	-0.77
*2. What percentage of the planet is covered by water? (**71**)*	7.72	5.53	-2.19
*3. What percentage of the world's population was under British Empire rule? (**23**)*	19.17	10.75	-8.42***
*4. What percentage of the world's population is left-handed? (**10**)*	15.91	13.35	-2.56*
*5. What percentage of the world's population lives in urban areas? (**56**)*	13.04	8.98	-4.06***
*6. What percentage of the global population are tobacco users? (**22**)*	19.42	17.44	-1.98
*7. What percentage of the global population is estimated to have had covid-19? (**44**)*	20.98	22.35	+1.37
*8. What percentage of the global population is of Christian faith? (**31**)*	11.36	6.55	-4.81***
*9. What percentage of E.U energy consumption comes from renewable sources? (**22**)*	13.88	11.25	-2.63*

Question and correct answers (%)	*1st round mean error. (WOC)*	*2nd round mean error. (Swarm)*	*Net effect*
*10. What percentage of global wealth is owned by the richest 1%? (**46**)*	22.13	18.63	-3.5**

*Table 3: Mean errors for the two rounds. Where: Net effect = (Swarm mean error – WOC mean error). All figures (apart from the correct answers) are in percentage points, and have been rounded to two decimal points, some figures may not exactly match due to rounding up. Here: * = 5%, ** = 1%, *** = 0.1% significance level achieved.*

In the table above, even though once again mixed results arise, for the effect of swarming on mean errors, a clear improvement in the vast majority of questions is observed. Considering the error of the mean (H1), which only decreased in 5 out of 10 questions, here, the mean error decreased in 9 out of 10 questions in total. The amount of statistically significant improvements doubled, to 6 questions, compared to three under the original hypothesis. More specifically, for question number 3, the swarm exhibited the largest decrease in mean error, compared to the crowd, to the magnitude of 8.42 percentage points, this pertains to a relative decrease in error of 43.92%, with the result being statistically significant at the 0.1% significance level. For question number 4 the swarm produced reduced errors compared to the crowd, with a 2.56 percentage point statistically significant decrease at the 5% significance level, this pertains to a relative decrease in error of 16.09%. Furthermore, in question number 5, the swarm compared to the crowd, managed a 4.06 percentage point decrease in mean error, this pertains to a relative decrease in error of 31.13%, an effect statistically significant at the 0.1% significance level. For question number 8, a reduction of the mean error occurs under swarming, compared to the crowd, to the magnitude of 4.81 percentage points, this pertains to a relative decrease in error of 42.34% an effect statistically significant at the 0.1% significance level. Moving on to question number 9, the swarm once again achieved a reduced error of 2.63 percentage points, compared to the crowd, this pertains to a relative decrease in error of 18.95%, an effect statistically significant at the 5% significance level. Finally, in question number 10, the swarm outperformed the crowd by reducing its mean error by a magnitude of 3.5 percentage points, this pertains to a relative decrease in error of 15.82%, an effect statistically significant at the 1% significance level. Based on the results above, a general trend of improvement in accuracy is clearly emerging, where two out of three questions with reduced mean error effects were found to be statistically significant. Contrary to that, only one case of a question with an increased mean error was produced, which was not statistically significant at any level, much like the first hypothesis, where some cases of increased error were observed.

4.2.1 Hypothesis 2.1

For the sub-hypothesis (H2.1): the results pertain to a one-sided matched pair sign-test between the individual errors of the first and second round. Only one tail was examined, due to already existing evidence from the two main hypotheses (H1 and H2), which leaned towards a reduction of errors, rejecting any effect of statistically significant increases in errors. Results are presented in table 4 below:

Question and correct answers	*Number of individual errors (decreased)*	*Number of individual errors (increased)*	*Number of individual errors (unchanged)*	*Sign-test results (p-value)*
*1. What percentage of the E.U population owns a car? (**56**)*	22	22	4	0.5598
*2. What percentage of the planet is covered by water? (**71**)*	18	13	17	0.2366
*3. What percentage of the world's population was under British Empire rule? (**23**)*	33	8	7	0.0001***
*4. What percentage of the world's population is left-handed? (**10**)*	27	15	6	0.0442*
*5. What percentage of the world's population lives in urban areas? (**56**)*	30	14	4	0.0113*
*6. What percentage of the global population are tobacco users? (**22**)*	24	21	3	0.383

Question and correct answers	*Number of individual errors (decreased)*	*Number of individual errors (increased)*	*Number of individual errors (unchanged)*	*Sign-test results (p-value)*
*7. What percentage of the global population is estimated to have had covid-19? (**44**)*	20	26	2	0.849
*8. What percentage of the global population is of Christian faith? (**31**)*	35	9	4	0.0001***
*9. What percentage of E.U energy consumption comes from renewable sources? (**22**)*	29	14	5	0.0158*
*10. What percentage of global wealth is owned by the richest 1%? (**46**)*	27	16	5	0.0631

*Table 4: Comparison of individual errors proportion. Here: * = 5%, ** = 1%, *** = 0.1% significance level achieved.*

Moving over to the sub-hypothesis, under the sign-test, for the comparison of individual matched errors, between the first and second round. In total, 8 out of 10 questions had a greater proportion of individuals which reduced their errors under swarming conditions, compared to those that increased them. Even more interesting is the fact that in 6 out of 10 questions, the proportion of individuals that reduced their errors when in swarm mode, was greater than the combination of those who increased their errors and those whose errors remained unchanged. Out of those six questions, five saw a statistically significant result for these ratios, based on the sign-test. Once again, there was not any statistically significant result that saw a greater ratio of increased compared to decreased errors. On an individual question basis, question number 3 saw 33 participants reduce their errors when in a swarm, compared to 8 who increased them and 7 with unchanged errors, this ratio was found to be statistically significant at the 0.1% significance level. This means that 68.75% of the total sample managed to reduce its error when in swarming formation, in the second round. Similarly, in question 4, 27 individuals reduced their error under swarming conditions, 15 increased them, and for 6 they remain unchanged between the two rounds., this proportion was found statistically significant at the 5% significance level. This means that 56.25% of the total sample managed to reduce its error when in swarming formation, in the second round. For question 5, 30 individuals managed to reduce their errors under swarming conditions, 14 increased them and 4 did not change their guess between the two rounds, this proportion was found to be statistically significant at the 5% significance level. This means that 62.5% of the total sample managed to reduce its error when in swarming formation, in the second round. Moving on to question number 8, swarming had a statistically significant effect at the 0.1% significance level, with 35 reduced errors, 9 increased and 4 unchanged. This means that 72.92% of the total sample managed to reduce its error when in swarming formation, in the second round. Finally, in question number 9, a statistically significant ratio of 29 reduced, 14 increased and 5 unchanged errors, was found at the 5% significance level. This means that 60.42% of the total sample managed to reduce its error when in swarming formation, in the second round. The only question in which individuals performed worse under swarming than under WOC conditions was question number 7, which once again did not produce statistical significance.

5. Discussion

Research on group thinking and collective decision making has been fascinating researchers for decades now. Theories such as the wisdom of the crowd have become even more relevant today with the rise of the internet and social media. Although collective intelligence represents a highly researched field, a new direction has been established in the past few years. From studying beehives and bird flocks, to simulating them in virtual environments and quantifying them through computational models, this direction leads us to today, where human based swarming systems are becoming even more relevant with every passing day. These types of systems examine how group decision making can greatly benefit from mechanisms and properties based on animal and insect swarms, systems such as the "UNU" platform: A platform that gathers online groups of individuals to answer questions and converge on consensus solutions, opinions, or predictions, by exposing members to the influence of others. The current experiment revolved around this concept and examined the hypothesis of whether swarms can outperform simple crowds (through means of aggregation) when it comes to accurately assessing general knowledge probability themed questions. What the main analysis showed, was that although results were mixed, those that achieved statistical significance, always supported an increase in accuracy under swarming conditions, relating to: a reduction in error of the mean guess (H1), a reduction in the mean group error (H2), as well as reductions in individual errors (H2.1). The main results coincide with previous literature, where swarming has been shown to reduce error and provide more accurate responses. Going into more detail, for the first hypothesis (H1) and the most fundamental, it was found that the greater effect was to a magnitude of 7.5 percentage points improvement. This represents a roughly 44% relative decrease in error, from the first round, a similar effect size to that of Rosenberg, Pescetelli and Willcox (2017), which examined accuracy amplification between the swarm and the crowd, for predicting stock market and commodities index forecasts. Similarly, the roughly 15% improvement on error for the swarm compared to the crowd for the first hypothesis, in question 4, resembles another finding from the same study. Similar studies have also supported the efficiency of swarming systems, one of those is by Rosenberg, Baltaxe and Pescetelli (2016), who found positive effects of swarming on American football betting, compared to a much larger crowd. Another study on sports betting by Rosenberg and Pescetelli (2017), showed an increase in accuracy when the same set of individuals forecasted match outcomes under swarming conditions, compared to individually doing so. Other studies that have offered similar results to this current paper, include Rosenberg et al. (2018) and Shat et al. (2021) which concluded that x-ray and M.R imaging diagnosing accuracy increased when a group of radiologists worked together as a swarm, relative to when individually assessing them. Finally, this papers' results are also support by Domnauer, Willcox and Rosenberg (2022), who exhibited an improvement under their hypothesis for the swarm, relative to both WOC and individual responses, based on the same version of the "UNU" platform which I employed.

An interesting result that should be mentioned here, is the existence of some non-statistically significant results in which the swarm performed worse than the crowd, more specifically questions 1,2,5 and 7 of the first hypothesis (H1). This could be a result of "Herding", which describes a set of phenomena based on social conformity, in which group sentiment could lead a crowd, sometimes a swarm, to stray from the truth, based on individual bias amplification. Although this effect is more prevalent in hierarchical or otherwise centralized societies (Becker, BrackBill and Centola, 2017), which swarms do not resemble, it still may pose a real drawback in systems were interaction between members is permitted (Lorenz et al., 2011). Although herding brings potential issues in swarming environments, there have been some studies to show the resilience of the slider-swarm platform to such phenomena, through a majority-minority structure (Domnauer, Willcox and Rosenberg, 2022). Another possible reason for such discrepancies in the results might have to do with the unique context of each question, as some questions due to the nature of their subject, might be considered too “vague” or “broad”, which might confuse or polarize participants. These limitations are due to the character limit imposed by the platform for the question being posed. Some examples which were communicated to me after the experiment, from the participants are the following:

- For question number 1, the word “car” is unclear on which type of vehicle it refers (professional or personal). Additionally, it might have been unclear whether the figure included or not children under 18. This could lead individuals to hold different concepts for the same question and thus lead to anomalous results and over or underestimation.
- For questions number 1 and 4, they were marked as “training questions”, so although participants had to accurately predict the answer, this might have encouraged “playing around” with the platform, rather than focusing on the accuracy of their responses. Hence, extreme or “irrational” guesses might be overrepresented in these two samples.
- For question number 7, there might be a small confusion with whether the question refers to only reported infections or all infections, including those based on predictive modeling, as is the case for the methodology employed in the cited study.

Other types of limitations relate to the online aspect of the experiment, where there is no way of knowing whether subjects looked up the answers while participating in the experiment, although it was mentioned as something prohibited, in the instructions. Another limitation pertaining to the nature of the platform, which was mentioned in the results of question 2 (H1), if a question is well known, the “2% rule” might lead the swarm further away from an initially accurate guess. Finally, a limitation referring to (H2.1), the sub-hypothesis, is that one assumption of a sign-test is between sample independence, which in the case of swarming, were group members are exposed to other members opinions does not hold, and therefore may lead to bias.

Even though the current experiment found promising results for this new swarming technology, further research is warranted, to firmly establish these effects, with the goal of employing it in a different number of fields which could benefit from it. One such field includes crisis and natural disaster management, by examining the movements of a virtual swarm, controlled by individual human operators, we can study the behavior of human crowds in evacuation scenarios, and prevent mass casualties by dealing with congestions, avoiding blockages, and reducing panic. This research could be based on papers such as Dyer et al. (2007), but performed in a virtual setting, possibly with the assistance of virtual or augmented reality software. Another crucial field in which swarming platforms and human swarming in general could prove useful is education. Lately, there have been many instances of new and innovative technologies being introduced into the modern classroom, bringing valuable insights to the teaching process (Lai and Bower, 2019). When someone considers the recent events of the covid-19 pandemic, it becomes all the clearer how invaluable technology has become to education, which is the reason why human-swarming platforms could relatively easily find application there. Swarming systems could also be applied in the classroom through another route, they could be used for collective test taking or group assignments, as they have been shown to promote group intelligence in the past (Willcox and Rosenberg, 2019). Additionally, pertaining to the more managerial side of educational institutions, swarming systems could be employed for more seamless decision making, similarly to other methods that are currently being proposed (Lekishvili and Kikutadze, 2023). Another future possible application for collective decision-making systems, such as swarming, could be in the financial sector. As I mentioned in the literature review, there have been studies on the effect of swarming on both financial and commodity index price predictions (Rosenberg, Pescetelli and Willcox, 2017), as well as crypto assets and "meme stocks" (Rosenberg et al., 2021) predictions. However, there has not been any study that I am aware of, which investigates how swarming forecasts can help improve volatility predictions in option contracts. Inspired by a recent report (Barclays, 2020), presenting evidence on how markets tend to overestimate realized volatility, human swarming systems could be tasked with making average daily, weekly, or monthly volatility forecasts and compare those with individual assessments, as well as market estimates. In this way, a new volatility risk premium could be elicited, and thus new ways of pricing option contracts could emerge, based on collective intelligent systems such as swarming. Finally, further analysis could be performed on the field, based on the factors that influence swarming systems themselves. Some of those have to do with varying swarm sizes, modifying weights for different swarm members, generating novel network architectures and communication schemes, examining concepts of group diversity on accuracy amplification, and even investigating the relation of risk profiles of group members, to swarming efficiency.

6. Conclusion

The goal of this paper was to explore the new version of the "UNU" human swarming platform, called "Slider swarm", and examine its efficiency when it comes to amplifying group accuracy in general knowledge questions, compared to the more classical WOC method of aggregation. What the main results showed is that although an improvement in accuracy did not hold true in all cases, the results that showed statistical significance always pointed towards a decrease in error. For the main hypothesis (H1), all three statistically significant results, showed a net decrease in the error of the mean, of 7.42 ($p < 0.001$), 2.39 ($p < 0.05$) and 4 ($p < 0.001$) percentage points, or a 44.3%, 15.53%, 48.19% relative decrease for questions number 3, 4 and 8 respectively, under swarming conditions. For the second hypothesis (H2), I also examined the effect of swarming on the accuracy of the mean error for the two matched samples. The results showed even more promising results, where once again all statistically significant effects pointed towards a decrease in mean error. This effect was now observed in the majority of questions (6/10), with varying net effect sizes: 8.42 ($p < 0.001$), 2.56 ($p < 0.05$), 4.06 ($p < 0.001$), 4.81 ($p < 0.001$), 2.63 ($p < 0.05$) and 3.5 ($p < 0.01$) percentage points, or a relative decrease of 43.92%, 16.09%, 31.13%, 42.34%, 18.95%, and 15.82%, for questions 3, 4, 5, 8, 9, 10 respectively, under swarming conditions. Lastly, for the sub-hypothesis (H2.1), the differences in individual errors where examined, and what percentage of the sample reduced or increased its error in each question. By using a one-sided matched pair sign-test, it was found that all statistically significant results showed a decrease for individual errors in the second round, with 68.75% ($p = 0.0001$), 56.25% ($p = 0.0442$), 62.5% ($p = 0.0113$), 72.92% ($p = 0.0001$) and 60.42% ($p = 0.0158$) of the sample improving its accuracy in questions 3,4, 5, 8 and 9 respectively, under swarming conditions.

Based on the findings, it was shown that in this particular experiment, swarming managed to reduce overall errors, both in the mean response as well as average errors. This suggests that the "UNU" platform "Slider swarm" might pose a legitimate tool, which can potentially amplify group intelligence, at least in general knowledge questions, relative to individual intelligence measures, as well as aggregate intelligence measures such as the mean (WOC). This could have real world applications, in both the public and private sectors, were swarming technologies could be deployed for further research on collective decision making and societal consensus convergence theories. Swarming could find real value when it comes to prediction or decision-making tasks, in various disciplines and fields, and might hold beneficial properties, for the cooperation and efficiency in human groups and societies.

7. References

1. Barclays. (2020). (rep.). Impact of Retail Options Trading (pp. 1–30). US: Barclays.

2. Basharat, H. (2019), MADS: Stata module to calculate Median Absolute Deviations. Retrieved from https://EconPapers.repec.org/RePEc:boc:bocode:s458336.

3. Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *OR*, *20*(4), 451. https://doi.org/10.2307/3008764

4. Becker, J., Brackbill, D., & Centola, D. (2017). Network Dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences*, *114*(26). https://doi.org/10.1073/pnas.1615978114

5. Befort, K., Baltaxe, D., Proffitt, C., & Durbin, D. (2018). Artificial swarm intelligence technology enables better subjective rating judgment in pilots compared to traditional data collection methods. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 2033–2036. https://doi.org/10.1177/1541931218621459

6. Beni, G., & Wang, J. (1993). Swarm intelligence in Cellular Robotic Systems. *Robots and Biological Systems: Towards a New Bionics?*, 703–712. https://doi.org/10.1007/978-3-642-58069-7_38

7. Boon, M. (2012). Predicting elections: A 'wisdom of crowds' approach. *International Journal of Market Research*, *54*(4), 465–483. https://doi.org/10.2501/ijmr-54-4-465-483

8. Camazine, S., Visscher, P. K., Finley, J., & Vetter, R. S. (1999). House-hunting by honeybee swarms: Collective decisions and individual behaviors. *Insectes Sociaux*, *46*(4), 348–360. https://doi.org/10.1007/s000400050156

9. Chafetz, J. (2005). It's the Aggregation, Stupid! [Review of The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations, by J. Surowiecki]. Yale Law & Policy Review, 23(2), 577–585. http://www.jstor.org/stable/40239647

10. Chen, H., De, P., Hu, Y. (J., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367–1403. https://doi.org/10.1093/rfs/hhu001

11. Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

12. Conradt, L., & Roper, T. J. (2003). Group decision-making in Animals. *Nature*, *421*(6919), 155–158. https://doi.org/10.1038/nature01294

13. Couzin, I. (2007). Collective minds. *Nature, 445*(7129), 715-715. https://doi.org/10.1038/445715a

14. Couzin, I. D. (2008). Collective cognition in animal groups. *Trends in Cognitive Sciences*, *13*(1), 36–43. https://doi.org/10.1016/j.tics.2008.10.002

15. Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature*, *433*(7025), 513–516. https://doi.org/10.1038/nature03236

16. de Condorcet, N. (1785). *Essai sur l'application de l'analyse à La probabilité des décisions rendues à La pluralité des Voix*. *Gallica*. Impr. royale. Retrieved February 7, 2023, from https://gallica.bnf.fr/ark:/12148/bpt6k417181.

17. Degroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121. https://doi.org/10.1080/01621459.1974.10480137

18. Domnauer, C., Willcox, G., & Rosenberg, L. (2022). Collaborative forecasting using “slider-swarms” improves probabilistic accuracy. *Lecture Notes in Networks and Systems*, 378–392. https://doi.org/10.1007/978-3-031-18461-1_26

19. Dorigo, M., & Gambardella, L. M. (1997). Ant Colony System: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, *1*(1), 53–66. https://doi.org/10.1109/4235.585892

20. Dowlen, O. (2009). Sorting out sortition: A perspective on the random selection of political officers. *Political Studies*, *57*(2), 298–315. https://doi.org/10.1111/j.1467-9248.2008.00746.x

21. Dyer, J. R. G., Ioannou, C. C., Morrell, L. J., Croft, D. P., Couzin, I. D., Waters, D. A., & Krause, J. (2008). Consensus decision making in human crowds. *Animal Behaviour*, *75*(2), 461–470. https://doi.org/10.1016/j.anbehav.2007.05.010

22. French, J. R. (1956). A formal theory of social power. *Psychological Review*, *63*(3), 181–194. https://doi.org/10.1037/h0046123

23. Galton, F. (1907). Vox Populi. *Nature, 75*(1949), 450–451. https://doi.org/10.1038/075450a0

24. Gardner, M. (1970). Mathematical games. *Scientific American*, *223*(4), 120–123. https://doi.org/10.1038/scientificamerican1070-120

25. Green, D. G., Leishman, T. G., & Sadedin, S. (2007). The emergence of social consensus in boolean Networks. *2007 IEEE Symposium on Artificial Life*. https://doi.org/10.1109/alife.2007.367823

26. Huth, A., & Wissel, C. (1992). The simulation of the movement of Fish Schools. *Journal of Theoretical Biology*, *156*(3), 365–385. https://doi.org/10.1016/s0022-5193(05)80681-2

27. Kennedy, J. (1997). The particle swarm: Social adaptation of knowledge. *Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC '97)*. https://doi.org/10.1109/icec.1997.592326

28. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*. https://doi.org/10.1109/icnn.1995.488968

29. Kirschenbaum, M., & Palmer, D. W. (2015). Perceptualization of particle swarm optimization. *2015 Swarm/Human Blended Intelligence Workshop (SHBI)*. https://doi.org/10.1109/shbi.2015.7321681

30. Krause, J., Ruxton, G. D., & Krause, S. (2009). Swarm intelligence in animals and humans. *Trends in Ecology & Evolution*, *25*(1), 28–34. https://doi.org/10.1016/j.tree.2009.06.016

31. Lai, J. W. M., & Bower, M. (2019). How is the use of technology in education evaluated? A systematic review. *Computers & Education*, *133*, 27–42. https://doi.org/10.1016/j.compedu.2019.01.010

32. Lehmann, E. L. & Dahj, M. (2006). Comparison with Student's t-test. In *Nonparametrics: Statistical methods based on ranks* (1st ed., pp. 76–81). essay, Springer.

33. Lekishvili, T., & Kikutadze, V. (2023). Decision-making process transformation in post-covid-19 world in Higher Educational Institutions. *Digital Management in Covid-19 Pandemic and Post-Pandemic Times*, 169–177. https://doi.org/10.1007/978-3-031-20148-6_16

34. Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

35. Li, Y.-X., Lukeman, R., & Edelstein-Keshet, L. (2008). Minimal mechanisms for school formation in self-propelled particles. *Physica D: Nonlinear Phenomena*, *237*(5), 699–720. https://doi.org/10.1016/j.physd.2007.10.009

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

37. Maas, Y. (2019). (thesis). *Outlier detection in non-Gaussian distributions*. Retrieved from https://repository.tudelft.nl.

38. Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology*, *107*(2), 276–299. https://doi.org/10.1037/a0036677

39. Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Kellis, M., Collins, J. J., & Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. *Nature Methods*, *9*(8), 796–804. https://doi.org/10.1038/nmeth.2016

40. Marshall, J. A., Bogacz, R., Dornhaus, A., Planqué, R., Kovacs, T., & Franks, N. R. (2009). On optimal decision-making in brains and social insect colonies. *Journal of The Royal Society Interface*, *6*(40), 1065–1074. https://doi.org/10.1098/rsif.2008.0511

41. Matache, M. T., & Heidel, J. (2005). Asynchronous random boolean network model based on elementary cellular automata rule 126. *Physical Review E*, *71*(2). https://doi.org/10.1103/physreve.71.026232

42. Metcalf, L., Askay, D. A., & Rosenberg, L. B. (2019). Keeping humans in the loop: Pooling knowledge through artificial swarm intelligence to improve business decision making. *California Management Review*, *61*(4), 84–109. https://doi.org/10.1177/0008125619862256

43. Millard, W. B. (2011). The Wisdom of Crowds, the madness of crowds: Rethinking peer review in the web era. *Annals of Emergency Medicine*, *57*(1). https://doi.org/10.1016/j.annemergmed.2010.11.012

44. Moreno-Gámez, S. (2022). How bacteria navigate varying environments. *Science*, *378*(6622), 845–845. https://doi.org/10.1126/science.adf4444

45. Palan, S., Huber, J., & Senninger, L. (2019). Aggregation mechanisms for crowd predictions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3391052

46. Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '87*. https://doi.org/10.1145/37401.37406

47. Ritchie, H., & Roser, M. (2019, September 20). *Age structure*. Our World in Data. Retrieved February 8, 2023, from https://ourworldindata.org/age-structure#citation.

48. Robinson, B. (2022, February 1). Remote Work Is Here to Stay and Will Increase Into 2023, Experts Say. *Forbes*. Retrieved from https://www.forbes.com/sites/bryanrobinson/2022/02/01/remote-work-is-here-to-stay-and-will-increase-into-2023-experts-say/?sh=60da2d8020a6.

49. Rosenberg, L. B. (2015). Human swarms, a real-time method for collective intelligence. *07/20/2015-07/24/2015*. https://doi.org/10.7551/978-0-262-33027-5-ch117

50. Rosenberg, L., & Baltaxe, D. (2016). Setting group priorities — swarms vs votes. *2016 Swarm/Human Blended Intelligence Workshop (SHBI)*. https://doi.org/10.1109/shbi.2016.7780279

51. Rosenberg, L., & Pescetelli, N. (2017). Amplifying prediction accuracy using Swarm A.I. *2017 Intelligent Systems Conference (IntelliSys)*. https://doi.org/10.1109/intellisys.2017.8324329

52. Rosenberg, L., & Willcox, G. (2018). Artificial Swarm Intelligence vs Vegas betting markets. *2018 11th International Conference on Developments in ESystems Engineering (DeSE)*. https://doi.org/10.1109/dese.2018.00014

53. Rosenberg, L., & Willcox, G. (2018). Artificial swarms find Social Optima. *2018 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. https://doi.org/10.1109/cogsima.2018.8423987

54. Rosenberg, L., Baltaxe, D., & Pescetelli, N. (2016). Crowds vs swarms, a comparison of intelligence. *2016 Swarm/Human Blended Intelligence Workshop (SHBI)*. https://doi.org/10.1109/shbi.2016.7780278

55. Rosenberg, L., Lungren, M., Halabi, S., Willcox, G., Baltaxe, D., & Lyons, M. (2018). Artificial swarm intelligence employed to amplify diagnostic accuracy in Radiology. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. https://doi.org/10.1109/iemcon.2018.8614883

56. Rosenberg, L., Pescetelli, N., & Willcox, G. (2017). Artificial swarm intelligence amplifies accuracy when predicting financial markets. *2017 IEEE 8th Annual Ubiquitous*

Computing, Electronics and Mobile Communication Conference (UEMCON). https://doi.org/10.1109/uemcon.2017.8248984

57. Rosenberg, L., Willcox, G., Askay, D., Metcalf, L., & Harris, E. (2018). Amplifying the social intelligence of teams through human swarming. *2018 First International Conference on Artificial Intelligence for Industries (AI4I)*. https://doi.org/10.1109/ai4i.2018.8665698

58. Rosenberg, L., Willcox, G., Palosuo, M., & Mani, G. (2021). Forecasting of volatile assets using artificial swarm intelligence. *2021 4th International Conference on Artificial Intelligence for Industries (AI4I)*. https://doi.org/10.1109/ai4i51902.2021.00015

59. Rosenberg, L.B., Hornbostel, C., & Pescetelli, N. (2018). Swarm AI to Detect Deceit in Facial Expressions.

60. Rowe, M. B. (1974). Wait-time and rewards as instructional variables, their influence on language, logic, and fate control: Part One-wait-time. *Journal of Research in Science Teaching*, *11*(2), 81–94. https://doi.org/10.1002/tea.3660110202

61. Schweitzer, F., & Hołyst, J. A. (2000). Modelling collective opinion formation by means of active Brownian particles. *The European Physical Journal B*, *15*(4), 723–732. https://doi.org/10.1007/s100510051177

62. Seeley, T. D., & Buhrman, S. C. (1999). Group decision making in swarms of honeybees. *Behavioral Ecology and Sociobiology*, *45*(1), 19–31. https://doi.org/10.1007/s002650050536

63. Seeley, T. D., & Buhrman, S. C. (2001). Nest-site selection in honeybees: How well do swarms implement the "best-of- N " decision rule? *Behavioral Ecology and Sociobiology*, *49*(5), 416–427. https://doi.org/10.1007/s002650000299

64. Shah, R., Astuto Arouche Nunes, B., Gleason, T., Fletcher, W., Banaga, J., Sweetwood, K., Ye, A., Patel, R., McGill, K., Link, T., Crane, J., Pedoia, V., & Majumdar, S. (2022). Utilizing a digital swarm intelligence platform to improve consensus among radiologists and exploring its applications. *Journal of Digital Imaging*. https://doi.org/10.1007/s10278-022-00662-3

65. Stahl, R.J. (1994). Using "Think-Time" and "Wait-Time" Skillfully in the Classroom. (ED370885). ERIC. https://eric.ed.gov/?id=ED370885

66. Sumpter, D. J. T., Krause, J., James, R., Couzin, I. D., & Ward, A. J. W. (2008). Consensus decision making by fish. *Current Biology*, *18*(22), 1773–1777. https://doi.org/10.1016/j.cub.2008.09.064

67. 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.

68. Toner, J., & Tu, Y. (1998). Flocks, herds, and schools: A quantitative theory of flocking. *Physical Review E*, *58*(4), 4828–4858. https://doi.org/10.1103/physreve.58.4828

69. Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "reading the mind in the eyes" test: Systematic Review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*, *18*(4), 326–354. https://doi.org/10.1080/13546805.2012.721728

70. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., & Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, *75*(6), 1226–1229. https://doi.org/10.1103/physrevlett.75.1226

71. von Neumann, J., & Burks, A. W. (1966). *Theory of self-reproducing automata*.

72. Vul, E., & Pashler, H. (2008). Measuring the crowd within. *Psychological Science*, *19*(7), 645–647. https://doi.org/10.1111/j.1467-9280.2008.02136.x

73. Waldron, J. (1995). The wisdom of the multitude. *Political Theory*, *23*(4), 563–584. https://doi.org/10.1177/0090591795023004001

74. Willcox, G., & Rosenberg, L. (2019). Short paper: SWARM INTELLIGENCE AMPLIFIES THE IQ of collaborating teams. *2019 Second International Conference on Artificial Intelligence for Industries (AI4I)*. https://doi.org/10.1109/ai4i46381.2019.00036

75. Willcox, G., Rosenberg, L., & Schumann, H. (2019). Sales forecasting, polls vs swarms. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3390043

76. Willcox, G., Rosenberg, L., Donovan, R., & Schumann, H. (2019). Dense neural network used to amplify the forecasting accuracy of real-time human swarms. *2019 11th International Conference on Computational Intelligence and Communication Networks (CICN)*. https://doi.org/10.1109/cicn.2019.8902352

77. Wolfram, S. (1983). Statistical mechanics of Cellular Automata. *Reviews of Modern Physics*, *55*(3), 601–644. https://doi.org/10.1103/revmodphys.55.601

78. Wolfram, S. (1984). Universality and complexity in cellular automata. *Physica D: Nonlinear Phenomena*, *10*(1-2), 1–35. https://doi.org/10.1016/0167-2789(84)90245-8

79. Zarnowitz, V. (1984). The accuracy of individual and group forecasts from Business Outlook Surveys. *Journal of Forecasting*, *3*(1), 11–26. https://doi.org/10.1002/for.3980030103

8. Appendix

Emails

Thesis experiment invitation

CM **Charilaos Magoulas <616762cm@student.eur.nl>**
9/21/2022 12:33 PM

To: Charilaos Magoulas

Welcome to my thesis experiment, my name is Charilaos Magoulas and I will be your moderator. Your interest in this experiment is greatly appreciated.

In the link below you will find all relevant information about the experiment (date and time) as well as instructions; please read them carefully so that you understand the process, it is only a 5-minute read. I highly recommend watching the videos included in the instructions as they are very helpful as a visual aid.

Link for document: https://docs.google.com/document/d/1fciyTgG1u8R7OS8y9Bb68-l1KctUtappDgh5cXUzoE0/edit?usp=sharing

Instructions for Charilaos' thesis experiment

Invitation/Context/Information Greetings, my name is Charilaos Magoulas and I am a behavioral economics postgraduate student in Erasmus Universiteit Rotterdam. This is an invitation for you to participate in my thesis' experiment on the topic "Wisdo...

docs.google.com

You will receive two reminder emails before the experiment takes place so keep an eye out for them.

Hope everything is clear and if you have any questions, please do not hesitate to contact me on the following email address: 616762cm@student.eur.nl

Email 1: invitation, and link for experimental instructions.

Reminder thesis experiment

CM **Charilaos Magoulas <616762cm@student.eur.nl>**
9/23/2022 7:51 PM

To: Charilaos Magoulas

Dear participant,

Thank you for agreeing to be a part of this research opportunity.

We will provide you with the link to the experiment on **25/09/2022**, 1 hour before the start of the swarming activity. Please be on the lookout for the e-mail with the link and instructions for joining. The session will start at **20:00 (GMT+3)**, Eastern European Summer Time, but please make sure to join 10 minutes earlier to ensure your Internet connection. The Swarm session will take approximately 30 minutes.

As a reminder, to participate in this study, all you need is a good Internet connection and a Chrome, Safari, or Firefox browser on a desktop or laptop computer (mobile devices are not supported).

If you still haven't done so, please take a quick look at the instructions provided in this document: https://docs.google.com/document/d/1fciyTgG1u8R7OS8y9Bb68-l1KctUtappDgh5cXUzoE0/edit?usp=sharing

Have fun and enjoy the research project, we really appreciate your help and look forward to your participation!

Charilaos Magoulas

Email 2: reminder, sent two days before experiment.

Final reminder experiment

Charilaos Magoulas <616762cm@student.eur.nl>
9/25/2022 6:02 PM

To: Charilaos Magoulas

Dear participant,

We are ready for you to begin the Swarm session! Please follow the instructions below. The total session will take approximately 30 minutes.

To participate in this study, all you need is a good Internet connection and a Chrome, Safari, or Firefox browser on a desktop or laptop computer (mobile devices are not supported).

Your participation is voluntary. You may stop participating at any time by closing the browser window.

The swarm will begin at **20:00 (GMT+3), approximately 1 hour from now.**

Please join 10 minutes early to ensure your connection is stable.

If you want to take a last look at the instructions, you can do so here: https://docs.google.com/document/d/1fciyTgG1u8R7OS8y9Bb68-l1KctUtappDgh5cXUzoE0/edit?usp=sharing

By clicking on the link below, you agree that:

- You are at least 18 years old
- You agree to participate in this online research

Link to join: https://charilaos.swarm.ai/j/935-968

This is the entry code in case the platform asks you for it: **935-968**

Please be patient as it may take some time to join the session, otherwise you can try joining through a different browser

When you enter the session, the moderator will provide important information and participation instructions in the chat space on the left so keep an eye on it

See you soon!

Charilaos Magoulas

Email 3: last reminder, sent one hour before experiment.

Debriefing, correct answers and small survey

Charilaos Magoulas <616762cm@student.eur.nl>
2:08 PM

To: Charilaos Magoulas

Debriefing and answers.pdf
151.17 KB

Hello everyone, once again I would like to thank you for participating in my experiment.
In this email you will find one document and one link. The pdf document includes debriefing information (purpose of the experiment, hypothesis tested etc.) as well as the correct answers to the questions asked, which a lot of people were curious to know.

The link includes an extremely short survey, through which I aim to gather some basic demographic questions such as age and gender. Although my intention was not to collect any, I was advised by my thesis supervisor it would be better to do so. Once again, all data gathered will be completely anonymous.

Link for short survey (computer and mobile device compatible): https://erasmusuniversity.eu.qualtrics.com/jfe/form/SV_6olt8uluxe6ymVM

Best,
Charilaos

Email 4: final email, includes debriefing, answers to questions and link for demographic survey.

Instructions document provided to participants (first email):

<u>Instructions</u>

Greetings, my name is Charilaos Magoulas, and I am a behavioral economics postgraduate student in Erasmus Universiteit Rotterdam. This is an invitation for you to participate in my thesis' experiment on the topic "Wisdom of Crowds", with the main goal of observing group accuracy in decision making. This experiment will be conducted through the "Swarm" platform and will take roughly <u>around 30 minutes, the date and time is Sunday 25th of September, 20:00 GMT+3 (Eastern European Summer Time)</u>. You will receive two reminder emails, the first one a couple of days before and a second one 1 hour before the experiment, which will include the link for the session.

For a brief summary of the platform, see information at the end of the document. *

To participate you will need:

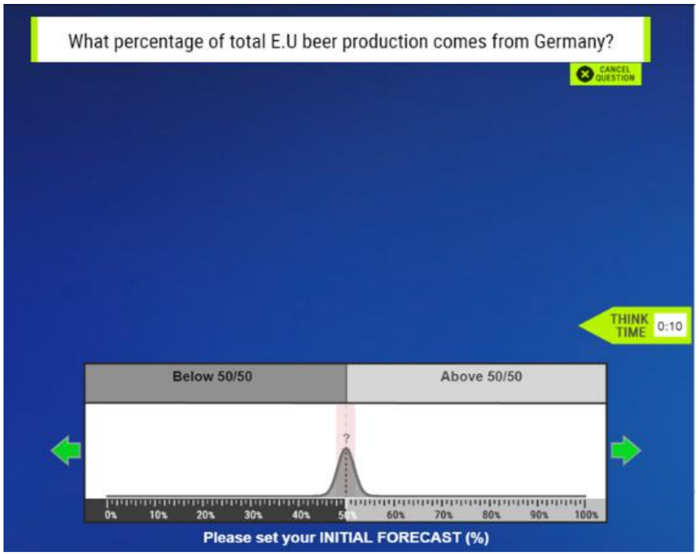
- Good Internet connection
- Up to date web browser (e.g., Chrome, Safari, Firefox)
- Desktop / laptop computer (**Mobile devices are not supported for this experiment**)
- To be at least 18 years old

<u>Instructions for the experiment (PLEASE READ CAREFULLY)</u>

During this experiment you will be asked a series of <u>8 different questions</u>, for each of those questions there will be <u>2 rounds/guesses</u>.

1. <u>Think time</u>.

First of all, there is a period of "THINK TIME", as indicated by the green arrow on the right side (see image below). During this time, please read the question carefully on top of the panel.



2. First round, Individual Deliberation.

During this first round (right after Think Time is over), you will be asked to input your first answer. All answers are in percentages, you can adjust your answer by clicking on the small green arrows located on either side of the panel. There is a timer on the bottom right side of the panel, indicating the amount of time you have left to set your guess. You do not need to "lock in" your answer, as it is automatically registered when the timer ends.

You can also click on this link to watch a video example (the one labeled "first gif"), full screen is available.

When a round starts, you'll notice a "DEADBAND" (a shaded region) around your distribution (in both rounds), with a "?" appearing on top of your distribution. This is the 2% region you need to leave in order for your answer to be counted. When you do leave, your answer/distribution enlarges and changes from GRAY to GREEN.

3. Second round, Group Deliberation.

In this final step, you will be asked to provide your final guess for the same question, again using the little green arrows on the sides. This will be done immediately after the first-round ends, so keep that in mind.

As you can observe in the second video (click and then scroll down), the panel now hosts a second window on top of the first one. In this second window, you will notice several blue distribution curves moving around (independently from you) and sometimes overlapping with each other to form much larger and taller ones. These are the inputs of other players participating in the experiment at the same time as you. The more players choose a certain answer the wider and taller a blue distribution will become.

By dragging your answer on the bottom panel (again using the green arrows), you will notice your movement reflected on the upper "blue panel" (green dotted line labeled "You"), this is so you can see where your estimation lies respectively to other players and the overall mean. Once again, in order for your estimation to be counted you will need to move until your distribution turns green in the bottom window.

Just keep in mind that the larger a distribution gets, it means the more players agree. This is not an experiment on crowd influence, rather, its aim is studying accuracy in crowds as opposed to individuals.

For the duration of the experiment, there will be a chat on the left side (as shown in the first video), this is where the moderator conveys instructions for participants.

If it still sounds complicated, do not worry, there will be two training questions before the experiment starts, so participants can familiarize themselves with the platform and the nature of the task.

Data handling

The data gathered will only include answers to the questions, as well as usernames which will be automatically and randomly created and will hold no affiliation to the participant, this data will be stored anonymously using password protection. No personal data will be collected whatsoever, and the answers given to the questions cannot in any way be linked back to individual participants that took part in the experiment. The experiment is run under the ethical guidelines of Erasmus Universiteit Rotterdam. The experimental data collected will be shared with the parent company of the platform Unanimous A.I for research purposes.

If you have any questions about the experiment, the platform, the data, or anything really, feel free to contact me at 616762cm@student.eur.nl

Hope everything is clear and that you enjoy yourself in this fun and revolutionary little experiment!

*Further reading on Swarm platform (not essential, read if interested)

Swarm AI® technology, developed by Unanimous AI, employs a unique combination of real-time human input and AI algorithms that are modeled after swarms in nature. Swarm Intelligence is the reason why birds flock, bees swarm, and fish school – they are smarter together than alone. Nature shows us that by forming closed-loop systems, groups can produce insights that greatly exceed the abilities of any individual member. While humans have not evolved this ability naturally, Swarm A.I technology enables this artificially, allowing groups to amplify their intelligence by forming real-time swarms.

You can learn more about Swarm Intelligence here:

- *Collaborative Forecasting using "Slider-Swarms" Improves Probabilistic Accuracy (most recent Published Research Paper)*
- *New Hope for Humans in an A.I. World (TED Talk)*
- *Artificial Intelligence Turns $20 into $11,000 in Kentucky Derby Bet (Newsweek)*
- *7 Predictions on The Next Era of Digital Retail (Forbes)*
- *Human–machine partnership with artificial intelligence for chest radiograph diagnosis (Nature, Digital Medicine)*
- *Artificial Intelligence Shows Potential to Gauge Voter Sentiment (Wall Street Journal, November 2020)*

Swarms have generated insights for the Washington Post, TechCrunch, TIME, Forbes and a host of Fortune 500 companies. More published research using Swarm® is available at https://unanimous.ai/publications/

Debriefing and correct answers document (fourth email):

Swarming experiment

Hello everyone and once again thank you for taking the time to participate in my thesis experiment, I hope you enjoyed it as much as I did. I have managed to gather really useful information from the session, which I would be glad to share with you once I am done reviewing the data. If you are interested in the outcome of the analysis, let me know at the email address provided below.

Under the topic "The Wisdom of Crowds", the main purpose of this study was to harness collective intelligence by observing any potential improvement in group accuracy under two conditions, when individuals answer independently of one another and when there is dependency among the group members.

Through this experiment, my expectation is to find an overall improvement of the mean accuracy in the second round, where participants were exposed to everyone else's opinion. This is the main idea behind the "Swarm" platform, utilizing crowd dynamics to aggregate and combine individual opinions in a way that will produce more efficient estimates.

Related studies have been done in the past providing support for dynamic and swarming systems, you can find some of them in the following link if further interested. https://unanimous.ai/publications/

If you have any questions about this study, feel free to contact me at the following email address: 616762cm@eur.nl

Answers to the questions

Below you will find the correct answers to the questions posed in the experiment along with the sources from which they were extracted (percentages are rounded to the nearest whole number).

1. What percentage of the world's population is left-handed?
   - **10%**
2. What percentage of the E.U population owns a car?
   - **56%**
3. What percentage of the planet is covered by water?
   - **71%**
4. What percentage of the world's population was under British Empire rule?
   - **23%**

5. What percentage of global wealth is owned by the richest 1%?
    - **46%**
6. What percentage of E.U energy consumption comes from renewable sources?
    - **22%**
7. What percentage of the global population is of Christian faith?
    - **31%**
8. What percentage of global population is estimated to have had covid-19?
    - **44%**
9. What percentage of the global population are tobacco users?
    - **22%**
10. What percentage of the world's population lives in urban areas?
    - **56%**

And here are their sources matched by number.

1. *Papadatou-Pastou, M., Ntolka, E., Schmitz, J., Martin, M., Munafo, M. R., Ocklenburg, S., & Paracchini, S. (2019, April 23). Human handedness: A meta-analysis. https://doi.org/10.1037/bul0000229*

2. *European Automobile Manufacturers' Association. (2022, April 8). Motorisation rates in the EU, by country and vehicle type. ACEA. Retrieved from https://www.acea.auto/figure/motorisation-rates-in-the-eu-by-country-and-vehicle-type/*

3. *United States Geological Survey. (2019, November 13). How much water is there on earth? completed. |U.S. Geological Survey. Retrieved from https://www.usgs.gov/special-topics/water-science-school/science/how-much-water-there-earth#:~:text=About%2071%20percent%20of%20the,percent%20of%20all%20Earth's%20water*

4. *Maddison, A. (2001), The World Economy: A Millennial Perspective, Development Centre Studies, OECD Publishing, Paris, https://doi.org/10.1787/9789264189980-en.*

5. *Shorrocks, A., Davies, J., & Lluberas, R. (2022). (rep.). Global Wealth Report 2022. Credit Suisse Research Institute. Retrieved from https://www.credit-suisse.com/about-us/en/reports-research/global-wealth-report.html*

6. *European Environment Agency. (2022, March 4). Share of energy consumption from renewable sources in Europe. Site. Retrieved from https://www.eea.europa.eu/ims/share-of-energy-consumption-from#:~:text=With%20a%2022.1%25%20share%20of,according%20to%20data%20from%20Eurostat*

7. *Hackett, C., & McClendon, D. (17AD). (rep.). Christians remain world's largest religious group, but they are declining in Europe. Pew Research Center. Retrieved from*

https://www.pewresearch.org/fact-tank/2017/04/05/christians-remain-worlds-largest-religious-group-but-they-are-declining-in-europe/#:~:text=Christians%20remained%20the%20largest%20religious,Pew%20Research%20Center%20demographic%20analysis.

8. *Barber, R. M., Sorensen, R. J., Pigott, D. M., Bisignano, C., Carter, A., Amlag, J. O., Collins, J. K., Abbafati, C., Adolph, C., Allorant, A., Aravkin, A. Y., Bang-Jensen, B. L., Castro, E., Chakrabarti, S., Cogen, R. M., Combs, E., Comfort, H., Cooperrider, K., Dai, X., … Murray, C. J. (2022). Estimating global, regional, and national daily and cumulative infections with SARS-COV-2 through Nov 14, 2021: A statistical analysis. The Lancet, 399(10344), 2351–2380.* https://doi.org/10.1016/s0140-6736(22)00484-6

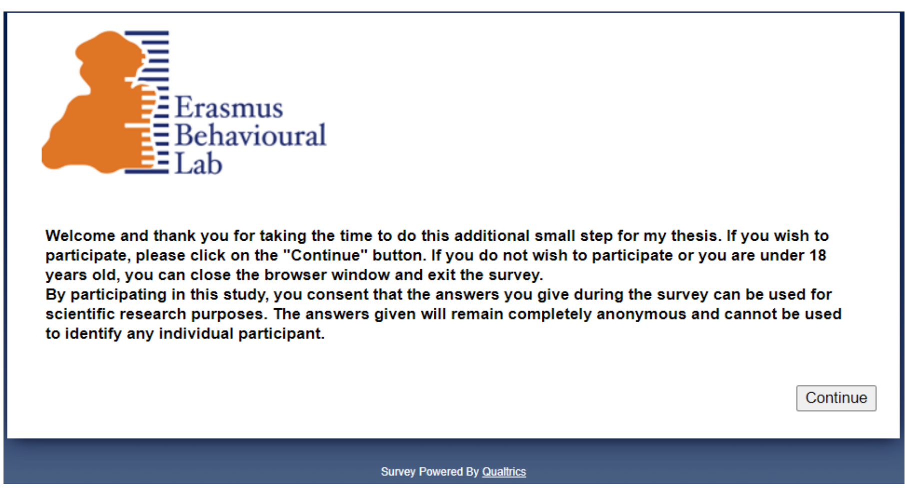
9. *World Health Organization. (2022, May 24). Tobacco. World Health Organization. Retrieved from* https://www.who.int/news-room/fact-sheets/detail/tobacco#:~:text=Over%2080%25%20of%20the%20world's,7.8%25%20of%20the%20world's%20women

10. *United Nations Conference on Trade and Development. (n.d.). Total and urban population. UNCTAD Handbook of Statistics 2021. Retrieved from* https://hbs.unctad.org/total-and-urban-population/

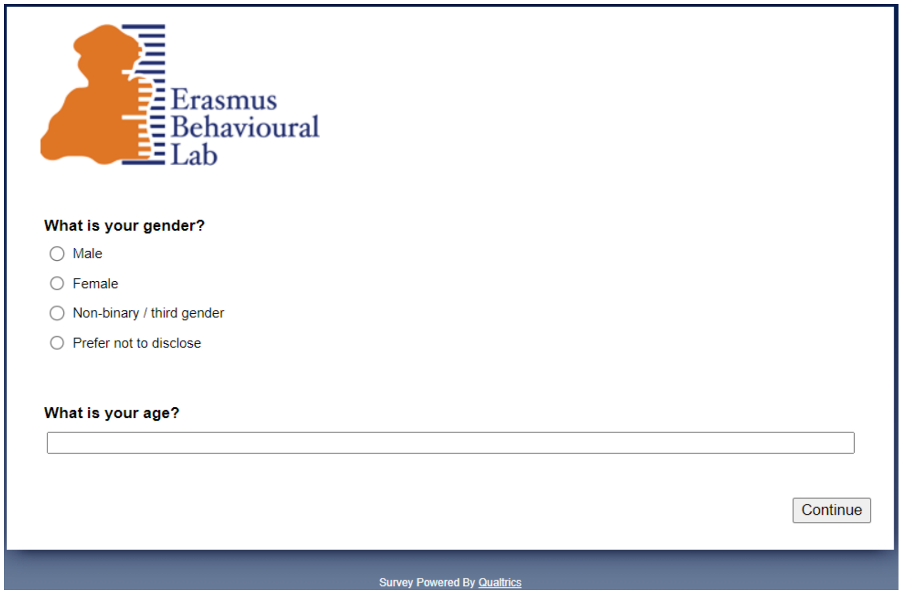
Thank you for your time and cooperation,

Charilaos Magoulas

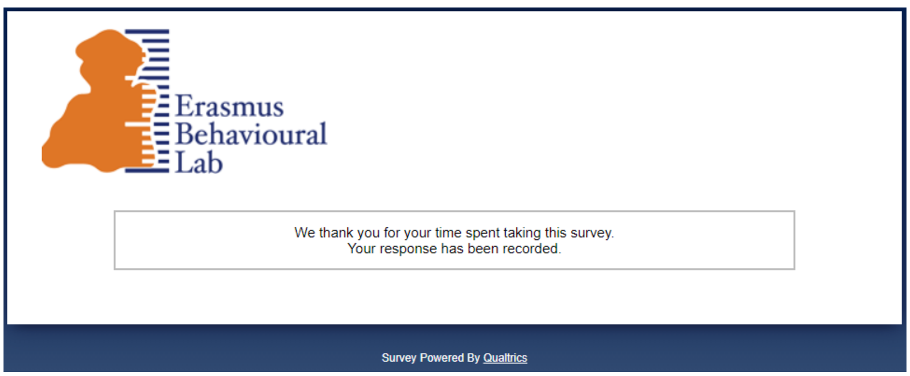
Qualtrics survey for demographic characteristics (fourth email):



Qualtrics survey (1): Consent Slide.



Qualtrics survey (2): Data collection slide.



Qualtrics survey (3): Thank you slide.