Repeatability and Accuracy
of
Swarms and Surveys

Overview:

The repeatability of a surveying method, defined as the frequency of agreement among repeated surveys, is commonly interpreted as a measure of accuracy. To investigate the relationship between repeatability and accuracy, we collected a large number of surveys and swarms that answered a widely used subjective judgement test. We find that, contrary to common assumption, repeatability in surveys does not translate to accuracy as surveys can be repeatable and simultaneously inaccurate. We also find that swarms are more accurate and similarly repeatable compared to surveys of the same size.

Procedure:

We performed this study with data collected from the Mind's Eye test: a widely used set of 35 questions designed to measure social intelligence. The test was given to 283 individuals, followed by 66 swarms comprised of those individuals. To quantify the repeatability, we generated 10,000 surveys of a set size by randomly selecting a crowd of individuals from the tested population with replacement, found the most popular, i.e. "crowd", answer to each question, and calculated the frequency of the most popular answer choice averaged over all questions. This process was repeated for surveys up to size of 50 individuals. In addition to surveys, an analogous procedure was performed to find repeatability of swarms. We also calculated the accuracy of all surveys and swarms on this test.

Results:

As shown in figure 1, a survey's repeatability approaches 100% as the survey size grows. Meanwhile, as depicted in figure 2, survey accuracy does not approach 100%. In other words, an increase of repeatability does not result in the same increase of accuracy. **Surveys can become more repeatable without becoming more accurate.** This is explained by the fact that surveying larger numbers of people yields the <u>same</u> answer more often, but not necessarily the <u>correct</u> answer**.** As observed in figure 1 and 2, **swarms are more accurate than surveys,** despite virtually identical repeatability. Additionally, a crowd of 7 swarms, or 27 people, was 1.6% more accurate while 2.8% less repeatable than the full survey of 283 people. In other words, a crowd of **7 swarms outperforms a poll more than 10 times its size.**
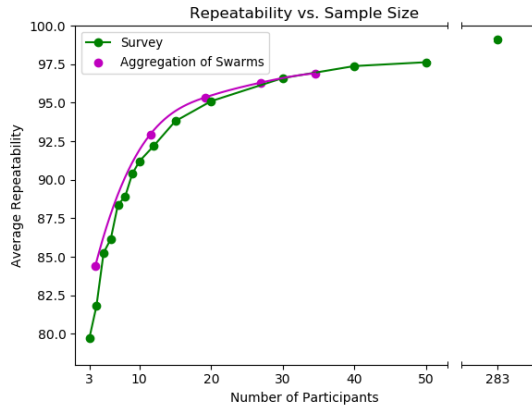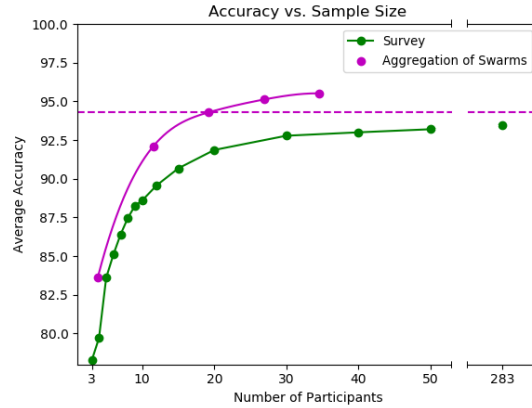
Figure 1: Repeatability vs sample size

Figure 2: Accuracy vs sample size

Case Study:

To investigate this important distinction between repeatability and accuracy more closely, we examined an individual question as a case study. On this question, repeatability and accuracy were calculated for a survey of 283 individuals, a survey of 4 individuals, a single swarm, and a crowd of 3 swarms. As depicted in figure 3, a large survey was far more repeatable **(98.6%)** than the swarm **(62%)**. However, because the majority of individuals answered incorrectly, a survey of 283 was correct only **1.4%** of the time. Meanwhile, the *majority* of swarms **(62%),** comprised of the *same* individuals, were correct.
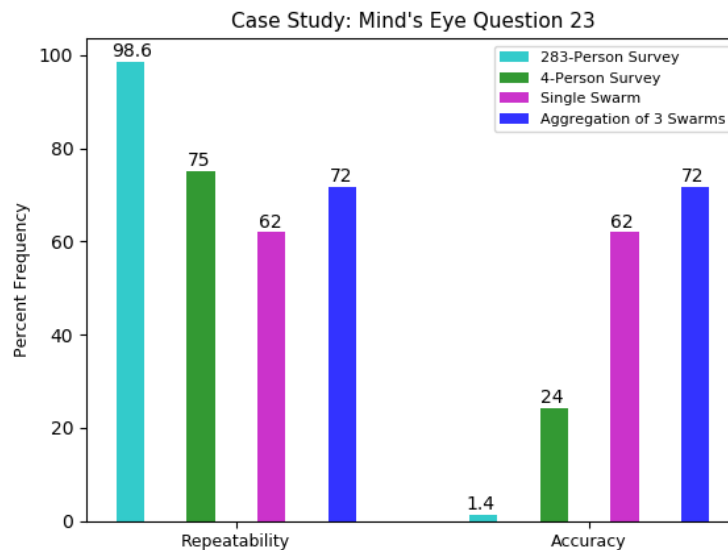

Figure 3: Repeatability and Accuracy on Case Study

We learn from this case study that repeatability is a function of sample size, which does not necessarily indicate accuracy. Additionally, connecting individuals in swarms enables more frequent convergence on the correct answer, *even* if the correct answer is chosen by the minority of constituents when surveyed individually.