

## Innovation Pilot: The wisdom of (very small but well trained) crowds

### Overview

We are piloting a Mobile App that will use alternate techniques and methodologies to improve the forecasting of future events. Rather than 'Big Data' our contexts are 'Small Data' where forecasts are based on minimal data points and the subjective views of individuals. The solution draws on 'Wisdom of the Crowd' and 'Calibration Training' concepts. An initial pilot has taken place with 30 respondents, this pilot will broaden the scope and include some real forecasting activities.

### Wisdom of the Crowd

In 1906 Francis Galton undertook his famous analysis of a 'Guess the weight' competition of a Prize Bull at a Country Show. The average guess of 787 contestants was 1,197lbs - the bull actually weighed 1,198 lbs.



If we relax our requirements from 'Precise' to 'Approximately Correct' then, a much, much smaller number of observations is required. For instance; if we wanted a range that included the crowds average guess with 90% confidence then we would need only the first five responses [1,113/1,240/1,320/1,236/1,231]<sup>1</sup>. For many business decisions 'approximately correct' satisfies our needs and is beyond what is normally available.

### Forecasting Calibration Training

Behavioural science (and experience) tells us that people are over confident in their abilities to provide estimates under conditions of uncertainty. Ranges are too narrow, there is a reluctance to move away from initial estimates ('Anchoring') and people are influenced by non-relevant facts.



Researchers have found that providing feedback on professional forecasts (meteorologists & medics) greatly improves performance. Furthermore, feedback in a field outside the predictors vocation, including general knowledge, also improves performance.

## About the Forecaster App

The premise of the App is that Calibration Training improves the forecasting skills of individuals; the collective intelligence of groups of trained individuals can then be used to generate forecasts.

Part 1 - Calibration Training – this involves sets of general knowledge questions that respondents are unlikely to know the correct answer to. They are given feedback on their estimating performance through the questions sets as well as tips for better forecasting. The training is intended to increase competence and identify any high performing individuals. This is of benefit generally and also to the crowd intelligence component of the App.

---

<sup>1</sup> Based on a spreadsheet simulation using Galton's original results

The App is designed to be engaging and 'game like' to encourage completion (points scoring, leader boards). As a mobile app the training can take place during 'down time' for instance when travelling

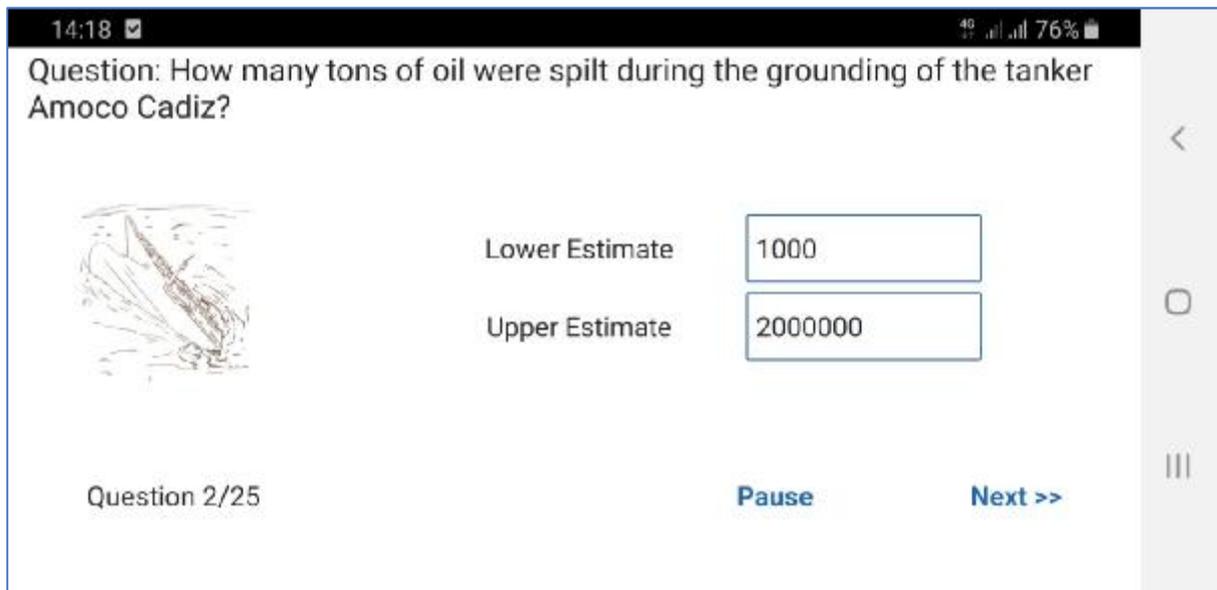


Diagram 1 - App Screen Shot 1 – Ranging Question

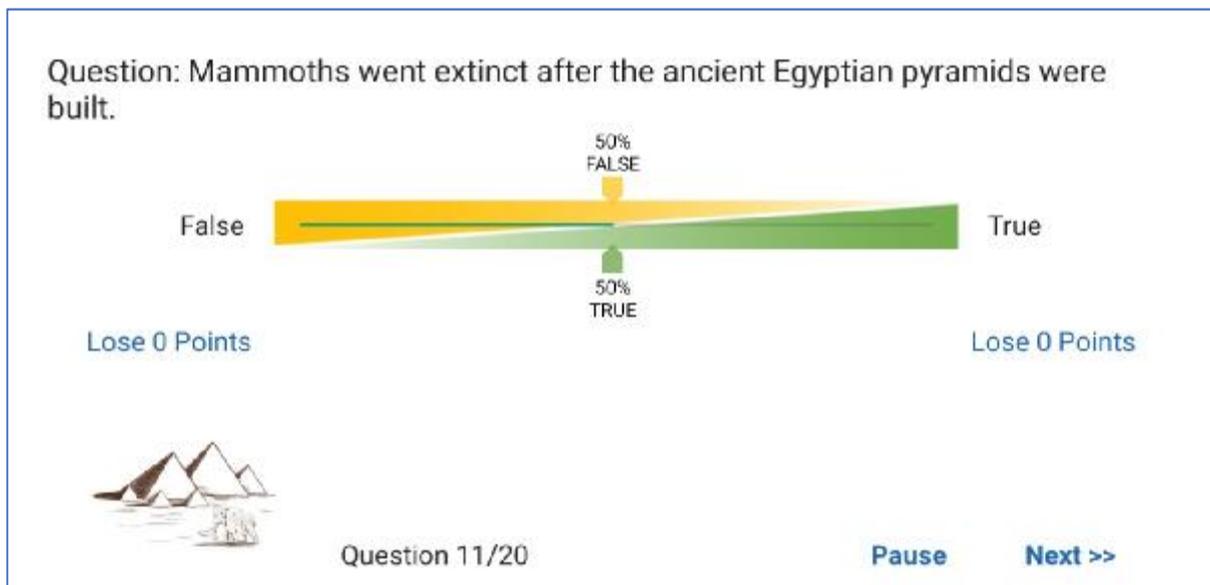


Diagram 2 - App Screen Shot 2 – True/False Question

Part 2 – Crowd Intelligence - When trainees have completed the training, they are invited to provide inputs to ‘intelligence’ type questions likely to be of general interest or requested by their organisation. The results are anonymous and will be available in real time from the receipt of 5 responses (or a smaller number of high performing forecasters), and updated and refined as more inputs are received, or as respondents update their responses in line with new information (for the pilot these will be undertaken manually or by email to allow us to develop the analytics and interface).

Intelligence questions for the Pilot might be along the lines of:

*How long do you think it will be before the first truly self-driving car will be available for public sale in the UK?*

*Will competitor xxx launch a new product this year?*

We are open to suggestions from piloteers on what these questions should be either for their organisation or generally.

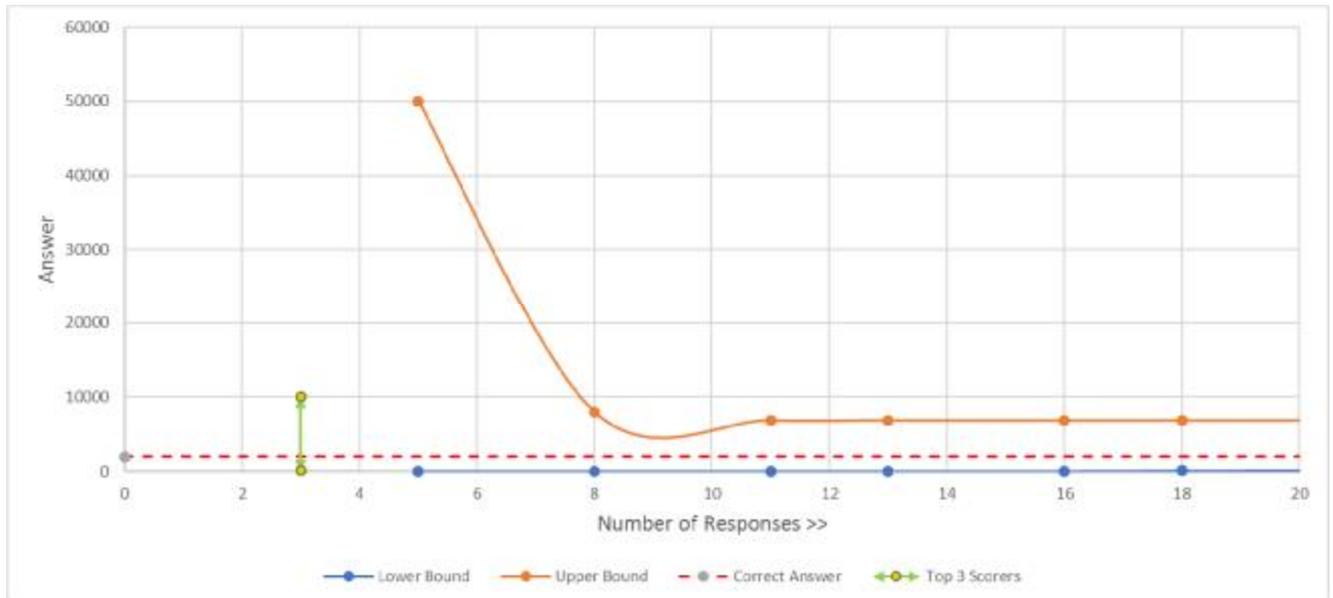
## Key Findings from initial pilot

The initial pilot was a group of 30 people. The analysis confirmed the validity of the forecasting analytics – although the individuals professed to having no knowledge of the answers, the group provided sound 90% confidences (9 out of 10 range questions were correct). The pilot also identified 3 individuals who performed particularly strongly.

Scoring for the range types question consists of a precision rating (how tight the range was) and whether the answer was correct. The graph below compares precision versus correctness:



The three high performers fairly consistently got the measure of the group response. The graph below shows the response for estimates for weight of a large object. The 3 high performers outperformed the next 5 respondents and collectively the group did not outperform the 3.



To take part please contact:



**Stephen Cresswell**

Director, Into Risk Limited

e: [stephen.cresswell@intorisk.com](mailto:stephen.cresswell@intorisk.com)

t: 01932 500 088

m: 07810 544 073