Research Markets & the Wisdom of Crowds - A Fresh Approach to Concept Testing -

By John Kearon, Founder, CEO & Chief Juicer of BrainJuicer Ltd

With large companies committed to organic growth, innovation is once again central to their mission and the need for concept screening and testing greater than ever. Yet, these same companies are under huge budget pressure, while trying to maintain the strength of the 'consumer's voice' in the innovation process.

Unlike silicon chips, a 'Moore's Law' of ever increasing capability-to-cost in MR remains an unobtainable dream of client-side research directors the world over.

So why not mix the latest online technology with 3 fluid ounces of creativity and a healthy dose of challenging orthodoxies, and invent ways to test more concepts, with more consumers, without compromising quality, in less time and money than current approaches?

The Thinking Behind the Paper

With large companies committed to organic growth, innovation is once again central to their mission and the need for concept screening and testing greater than ever. Yet, these same companies are under huge budget pressure, while trying to maintain the strength of the 'consumer's voice' in the innovation process.

Unlike silicon chips, a 'Moore's Law' of ever increasing capability-to-cost in MR remains an unobtainable dream of client-side research directors the world over.

So why not mix the latest online technology with 3 fluid ounces of creativity and a healthy dose of challenging orthodoxies, and invent ways to test more concepts, with more consumers, without compromising quality, in less time and money than current approaches?

This paper and the experiment contained within it, is just one attempt to search for a more efficient concept testing method. The results, while not conclusive make for very interesting reading. And taken together with other examples of successful prediction markets, raise the possibility of a radically more efficient approach to concept testing that at the very least merits further testing and exploration.

This paper is going to explore the latest academic thinking on '*Research Markets*' and 'Wisdom of Crowds sampling' and their potential application to market research:

Mix challenging research orthodoxies.

1. *Could a 'Research Market' be as accurate as a monadic concept test*?
   A monadic concept testing is best but expensive, forced choice testing is more efficient but creates extremes – so is there a better alternative?
2. *Could a Crowd of Consumers be as accurate as a carefully sampled target audience*?
   Scientifically sampled target audiences are a fundamental tenet of accurate market research but is there a valid alternative?
3. *Could answering for the market be as accurate as answering for oneself?*
   Answering surveys based on your own behaviour, thoughts and interests is another fundamental tenet of market research but is there a valid alternative?

With three fluid ounces of creativity.

1. *Test concepts through a 'Research Market'.*
   Respondents are asked to buy and sell shares in the concepts presented, as a way of identifying those they think would be most and least successful in the market. The aggregate of all the buy and sell requests on each concept is what produces the overall judgement on the concept (for comparison with the monadic test results).
2. *Use a large, diverse 'crowd of consumers'.*
   Invite a random group of 500 respondents (based on willingness to participate), to take the place of the usual scientifically sampled, objective target audience.
3. *Ask the crowd to answer 'for the market'.*
   Respondents are asked to select 'concept stocks' based on their perception of those the market would reward, rather than the conventional research approach of asking for their own personal likes and dislikes.

And the latest Technology.

We utilised the Internet's ability to create a dynamic market, which our crowd could easily access to share their buy and sell opinions, through an automated interviewing system, uninfluenced by other respondents or an interviewer.

The results from the experiment, if validated, suggest *Research Markets* could be a clearer-quicker-cheaper way to test large numbers of concepts:

1. Lots of ideas	✓ Test 10-20 ideas at a time
2. Large, cheap comparable samples	✓ Among a diverse crowd of 500 consumers
3. Clear, accurate results	✓ Purchase intent matches the strongest from monadic testing but also gives greater clarity on which ideas are weak

An Exploration of Research Markets

James Surowiecki, author of *The Wisdom of Crowds*, challenges some aspects of the modern corporate decision-making model, arguing that chasing the expert is a mistake, and a costly one at that; we should stop hunting the expert and ask the crowd instead. Surowiecki's challenging assertion is that under certain circumstances, crowds make better decisions than individuals.

For a crowd to be wise, Surowiecki explains, it must be *diverse*, *independent* and *faithfully aggregated*. As he puts it, "Diversity and independence are important because the best collective decisions are the product of disagreement and contest, not consensus or compromise". Faithful aggregation rather than intuitive interpretation is essential to accurately reflect the democratic sum of all opinions or the 'wisdom of the crowd'.

Surowiecki's book (already a bestseller in the States), challenges the intuitive notion that an expert or small group of experts will know best and provides a compelling argument in favour of increasingly looking to the crowd for smarter answers and more accurate decision making. In particular he raises the intriguing possibility of using '*decision markets'* or (to dramatise the inversion), what we might call '*Research Markets'*. One long-standing example is the Iowa Electronic Market, where over the last 5 years, betting on the outcome of elections has proved significantly more accurate than conventional political polling. The same 'Decision Market' mechanism is currently being applied to a number of other tasks including predicting the box office revenues of new movies, the likely success of technology innovations and even the whereabouts of terrorists and terrorist activities.

Applying this thinking to Market Research would reinforce the wisdom of using large, quantitative samples for decision-making. More challengingly, it would also question one of quantitative research's central tenets, namely the necessity of using targeted samples to accurately predict in-market performance. Finally, it raises the intriguing possibility of using '*Research Markets*' as a more accurate and efficient method of prediction. If proven, they would have radical implications for market research.

Wisdom of Crowds & Research Markets – A Synopsis

Crowds are not traditionally associated with wisdom. The case against them has been succinctly put by a number of thinkers and writers. The historian Thomas Carlyle wrote, "I do not believe in the collective wisdom of individual ignorance". For Charles MacKay, *Extraordinary Popular Delusions and the Madness of Crowds (1841)*, crowds were never wise and collective judgements were doomed to the extreme. And the fiercest critic of all, Gustave Le Bon, whose polemic *The Crowd: A Study of the Popular Mind (1895)* rejected the rise of democracy and the idea that ordinary people could be trusted to make good collective judgements. He wrote, "In crowds it is stupidity and not mother wit that is accumulated". Crowds "can never accomplish acts demanding a high degree of intelligence and are always intellectually inferior to the isolated individual".

These attitudes are quite understandable. If a group of 100 people took an IQ test, you would expect a range of individual performance (highly likely distributed in a bell curve) and the aggregate would be a mediocre number significantly below the results of the best performers. So it is rather counterintuitive to accept that a group might somehow perform better in aggregate than the best individual performers. And yet this is exactly what Surowiecki shows is possible for solving cognition, coordination and cooperation problems as long as the crowd is *diverse*, *independent* and *faithfully aggregated*. He uses an example about the TV game show, *Who Wants to be a Millionaire*, where the audience consistently outperform the 'expert friends' suggesting that a diverse, independent audience's answers to each question on an IQ test would produce a score higher than all but the brightest performers.

Another simple example and one not reliant on intelligence, is the popular fairground game of 'guess the weight of the cake'. Sociologist Kate H. Gordon replicated the game in experiments asking 200 students to rank items by weight and found the group's estimate

was 94 percent accurate, which was better than all but 5 of the individual guesses. But perhaps the best example of this comes from British scientist, Francis Galton who was actually trying to prove the opposite of what he found. Like Gustave Le Bon, Galton did not believe in the wisdom of a crowd, "the stupidity and wrong-headedness of many men and women being so great as to be scarcely creditable". He spent much of his career collecting evidence to prove that, only if power and control stayed in the hands of the select, well-bred, well educated few, could society remain healthy and strong. Surowiecki tells the story of Galton walking through the London International Exhibition in 1884:

> Galton came across a weight-judging competition. A fat ox had been selected and place on display, and members of a gathering crowd were lining up to place wagers on the weight of the ox. (or rather they were placing wagers on what the weight of the ox would be *after* it had been "slaughtered and dressed.") For sixpence, you could buy a stamped and numbered ticket, where you filled in your name, your address and your estimate. The best guesses would receive prizes.
>
> Eight hundred people tried their luck. They were a diverse lot. Many of them were butchers and farmers, who were presumably expert at judging the weight of livestock, but there were also quite a few people who had, as it were, no insider knowledge of cattle. "Many non-experts competed," Galton wrote later in the scientific journal *Nature*, "like those clerks and others who would have no expert knowledge of horses, but who bet on races guided by newspapers, friends and their own fancies." The analogy to democracy, in which people of radically different abilities and interests each get one vote, had suggested itself to Galton immediately. "The average competitor was probably well fitted for making a just estimate of the dressed weight of the ox, as an average voter is of judging the merits of most political issues on which he votes," he wrote.
>
> Galton was interested in figuring out what the "average voter" was capable of because he wanted to prove that the average voter was capable of very little. So he turned the competition into an impromptu experiment. When the contest was over and the prizes had been awarded, Galton borrowed the tickets from the organizers and ran a series of statistical tests on them. Galton arranged the guesses (which totalled 787 in all, after he had to discard 13 because they were illegible) in order from highest to lowest and graphed them to see if they would form a bell curve. Then, among other things, he added all the contestants' estimates, calculated the mean of the group's guesses. That number represented, you could say, the collective wisdom of the crowd. If the crowd were a single person, that was how much it would have guessed the ox weighed.
>
> Galton undoubtedly thought that the average guess of the group would be way off the mark. After all, mix a few very smart people with some mediocre people and a lot of dumb people, and it seems likely you'd end up with a dumb answer. But Galton was wrong. The crowd had guessed that the ox, after it had been slaughtered and dressed, would weigh 1,197 pounds. After it had been slaughtered and dressed, the ox weighed 1,198 pounds. In other words, the crowd's judgement was essentially perfect. Perhaps breeding did not mean so much after all. Galton wrote later: "The result seems more creditable to the trustworthiness of a democratic judgement than might have been expected." That was, to say the least, an understatement.
>
> What Francis Galton stumbled upon that day was the simple, but powerful, truth that is at the heart of 'Wisdom of Crowds': under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them…Even if most of the people within a group are not especially well-informed or rational, it can still reach a collectively wise decision.

The book provides compelling examples of groups being smarter than most individuals across all 3 types of problem solving:

- *Cognitive* problems with reasonably definitive answers
  - who will win, which is best, how many will it sell?

- *Coordination* problems that require groups to organise their behaviour - how to organise a company, how buyers find sellers and vice-versa
- Cooperation problems, where self-interest is at odds with the collective good – what tax to pay, how to deal with pollution, what is a reasonable salary?

Market Research is largely employed in the service of cognitive problems, in understanding the world, *as is* and predicting the world, *as it will be*, rather than with the organising mechanisms or resolution of conflicting interests. As such, the rest of this paper will focus on understanding how *Wisdom of Crowds* thinking relates to cognitive problems and how it might be applied to market research.

Having demonstrated by example, how a *diverse, independent and faithfully aggregated* crowd can be smarter than even the smartest individuals/experts, Surowiecki goes on to show how a 'market' mechanism is a surprisingly accurate and incredibly efficient means of understanding the world *as is* and *as it will be.* These markets, often referred to by economists as, 'decision markets' or 'predictive markets', enable a diverse group of independent 'investors' to each place their bets or cast their votes and allow the collective opinion of the 'market' to emerge. It is this aggregated, collective opinion which seems to be so surprisingly accurate or wise, despite the group being made up of those with a lot of knowledge and expertise and those with very little. If you think of each guess as containing two things, information and error, it seems the error contained in each individual guess is cancelled out amongst a large diverse group, leaving pure information as the emergent output. It is as if evolution has programmed us to be collectively smart. Studies of ant colonies, bee hives and pedestrian patterns all confirm the capability for smart behaviour to emerge without central control or communication from each individual's actions taken based on local knowledge and instinct.

To illustrate the potential of markets to be very smart, Surowiecki explores a fascinating study by professors Michael T. Maloney and J. Harold Mulherin of the stock market's reaction to the *Challenger* space shuttle disaster. At 11:39 AM on January 28, 1986 the Challenger space shuttle exploded. Eight minutes later the first story appeared on the Dow Jones News Wire and within minutes investors started dumping stocks in the four major contractors: Rockwell who built the engines, Lockheed who managed ground support, Martin Marietta who manufactured the fuel tanks and Morton Thiokol who built the solid-fuel booster rocket. By 12:01 PM, just 21 minutes later, the stock of the first three was down six, five and three percent respectively but so many investors were trying to sell Thiokol stock with so few buyers, trading was suspended almost immediately. By the time it started trading one hour later, it was down six percent and by the end of the day was down 12 percent compared to only a three percent drop on the day for the others.

It took an exhaustive Presidential Commission 6 months to decide that the O-ring seals on the booster rockets made by Thiokol were to blame but the stock market came to the same collective conclusion just 30 minutes after the explosion. How did the market get it right? Maybe it was pure luck but Maloney and Mulherin searched without success for rational explanations like insider dealing. It seems savvy insiders did not cause the first-day drop, it was all those investors – most of them relatively uninformed – who refused to buy the Thiokol stock vis-à-vis the other manufacturers stock which remained reasonably buoyant. They concluded their study by quoting economist Maureen O'Hara who said, "While markets appear to work in practice, we are not sure how they work in theory."

Surowiecki goes on to show how these market mechanisms have been successfully utilised to predict the world *as it will be* in a number of different scenarios. The original and longest running is the Iowa Electronic Markets (IEM) founded in 1988 by the Business College at the University of Iowa:

> The IEM features a host of markets designed to predict the outcomes of elections – presidential, congressional, gubernational, and foreign. Open to anyone who wants to participate, the IEM allows people to buy and sell futures "contracts" based on how they think a given candidate will do in an upcoming election. While the IEM offers many different types of contracts, two are most common. One is designed to predict

> the winner of an election. In the case of the California recall in 2003, for instance, you could have bought an "Arnold Schwarzenegger to win" contract, which would have paid you $1 when Schwarzenegger won. Had he lost, you would have received nothing. The price you pay for this kind of contract reflects the market's judgement of a candidate's chances of victory. If a candidate's contract costs 50 cents, it means, roughly speaking, that the market thinks he has a 50 percent chance of winning. If it costs 80 cents, he has an 80 percent chance of winning, and so on.
>
> The other major kind of IEM contract is set up to predict what percentage of the final popular vote a candidate will get. In this case, the payoffs are determined by the vote percentage: if you'd bought a George W. Bush contract in 2000, you would have received 48 cents (he got 48 percent of the vote) when the election was over.
>
> …So how has the IEM done? Well, a study of the IEM's performance in forty-nine different elections between 1988 and 2000 found that the election-eve prices in the IEM were, on average, off by just 1.37 percent in presidential elections, 3.43 percent in other U.S. elections and 2.12 percent in foreign elections…The IEM has generally outperformed the major national polls, and has been more accurate than them months in advance of the actual election. Over the course of the presidential elections between 1988 and 2000, for instance, 596 different polls were released. Three-fourths of the time, the IEM's market prices on the day each of those polls were released was more accurate. Polls tend to be very volatile, with vote shares swinging wildly up and down. But IEM forecasts, though ever-changing, are considerably less volatile, and tend to change dramatically only in response to new information. That makes them more reliable as forecasts.

It's important to note that the academic papers on decision markets in general and the success of the IEM in particular, offer a different explanation to Surowiecki for the accuracy. Robert Forsythe, Forest Nelson, George R Neumann and Jack Wright, in *Anatomy of an Experimental Political Stock Market* attribute the success of the IEM not to the collective Wisdom of the Crowd but to a small minority of rational and foresighted investors – *marginal investors* – who kept the market smart by buying and selling whenever they felt prices were starting to deviate from their true value. Whilst acknowledging the intuitive appeal of *marginal investors* in retaining our faith in 'a few smart people', Surowiecki rejects the explanation, "No trader – even in the stock market, where some investors control enormous amounts of money – has enough capital to outweigh the aggregated buying and selling power of all other investors. That means that the decisions taken by *marginal investors* will be overridden immediately if the 'crowd' disagrees with them. This is obviously true of the IEM, where the stakes people can use to buy and sell contracts are limited to $500. As a result, if the aggregate judgement of most IEM traders was really irrational and inaccurate, the effect of their inaccurate trading would overwhelm the attempts of smart investors to resist it." He acknowledges some traders may well be smarter than others, just as Forsythe et. Al. put forward in their paper but reminds us that in any collective cognition problem solving, there will be a small minority that do better than the group as a whole but their judgement alone does not make the market smart. It is the aggregated judgement of all the traders that makes the market smart.

Surowiecki cites several other examples of successful 'decision/prediction markets', such as the Hollywood Stock Exchange (HSX) which predicts box office returns for new films as well who will win the Oscars. It scored a notable success in 2000, the year that The Wall Street Journal actively canvassed the members and predicted five out of the six top awards. However the HSX performed even better, picking six out of six, as well as picking 35 of the 40 nominees. The self-evident virtues of these decision markets have spawned others such as, NewsFutures,TradeSports, Innovation Futures and the ill-fated U.S. government sponsored DARPA market to predict what, where and when the next terrorist strike may occur (the market was closed due the adverse publicity surrounding the perceive gain from betting on such events).

There are some smaller companies that now mirror this crowd-driven market behaviour – Betfair.com and Sporting Index, both involved in sports betting, stocks/shares. However, to date, there are few if any examples of large companies adopting 'decision markets' as a source of information and no published examples of Research Agencies exploring their potential. This is despite them being tailor-made for corporate decision making. As Surowiecki says, "Corporate strategy is all about collecting information from many different sources, evaluating the probabilities of potential outcomes, and making decisions in the face of an uncertain future." But as he also says referring to seeming corporate resistance, "...just because collective intelligence is real it doesn't mean that it will be put to good use."

A Research Markets Experiment

The IEM electoral prediction market is one of the only published examples of the potential of using *Wisdom of Crowds* thinking for Market Research. And, as highlighted in the previous section, over 596 polls between 1988 and 2000, it proved more accurate than our quantitative polling methods three-fourths of the time. These results alone suggest the Market Research community would be wise to expend efforts to validate or repudiate the claims and understand whether and where it might be usable.

To this end, we wanted to design an experiment to test the *Research Markets* and *Wisdom of Crowds* ideas in the context of market research. Specifically, we wanted to test the following hypotheses, each of which is at odds with current market research thinking (conventional wisdom is summarised underneath each):

Three hypotheses to test:

1. ***A diverse crowd of people will be as accurate as a targeted sample*** (in conventional MR wisdom a representative sample of the target audience is the only accurate way to reflect in-market reality)
2. ***A forced choice 'Research Market' will be as accurate as a monadic test*** (in conventional MR wisdom a forced choice test will produce a more extreme and less accurate result than a monadic test) + asking the group to predict others' behaviour will be less accurate than predicting one's own behaviour)
3. ***Asking people to speak for the market will be as accurate as a personal answer*** (in conventional MR wisdom the only accurate answer is one based your own behaviour, opinions and interests)

Devising a test for these hypotheses to detect the world *as is* would probably have been easier, since it would be possible to validate against in-market reality. However, we decided to try and test these hypotheses in the more difficult context of the world *as it will be*, or more specifically, in the arena of testing new product concepts. If the hypotheses proved true in predicting the future, this would be even more significant for the MR community.

The difficulty with testing the hypotheses on new product concepts is that by necessity we're asking about ideas that are not yet launched and therefore have no way of validating the in-market accuracy of the results – at least not for a long time to come. We decided to use the closest substitute to in-market accuracy, which in current MR thinking is a monadic concept test among a representative sample of the target audience.

The other difficulty with the 'Research Market' was in deciding the best market mechanism to use and how technically to best execute the experiment. Surowiecki himself says, "A decision market is an elegant and well-designed method for capturing the collective wisdom. But the truth is that the specific method that one uses probably doesn't matter very much...The real key, it turns out, is not so much perfecting a particular method, but satisfying the three conditions of *diversity, independence and faithful aggregation*. So after much debate and a few pragmatic constraints we decided on the following design:

- *An online 'Research Market' of 10 new product ideas (in the same category)*:
    - Using a diverse, non-representative panel of 500 people
    - Each asked to independently bet on which of the 10 ideas they believe will be most and least successful in the market place. For the ideas they chose as

best and worst, they answer standard concept test questions from the point of view of the market rather than themselves; purchase intent, ratings, likes and dislikes, reasons why
    - The results from the ideas chosen as most and least successful were aggregated to produce the final results i.e. the scores of those selecting as their favourite were aggregated with those selecting as their least favourite – to mimic the actions of buyers and sellers in a market

vs.

- *An online monadic concept test for each one of the same ten new product ideas*
    - Using matched, representative samples of 100 respondents in the target audience
    - Each independently asked the standard concept test questions for the new product in question; purchase intent, ratings, likes and dislikes, reasons why
    - The results from each monadic test were compared for the final analysis

The Results of the Experiment

The results, while not conclusive, are certainly very interesting.

An important thing to note is that we ran the 'Research Market' twice, once with a betting mechanism and then again without. The reason for this was a suspicion on our part that the betting mechanism we used had distorted the market, exaggerating the results in a way we had not anticipated. Sure enough, removing this mechanism from the 'Research Market' made a significant difference in that the results were less extreme and produced an even stronger match with the monadic results for the favoured ideas.

Before we look at the results, let me try and explain why we believe the betting mechanism used distorted the results. Since we had no in-market validation of each idea (something that would be impossible within the time frame of the experiment), we decided to create a proxy in order to have a basis on which to reward 'investors' for selecting the 'best and worst' concept stocks. Each person was given a £1 thank you for taking part but they were told they could win an additional £2 if the ideas selected by them as 'best' and 'worst' were the ideas selected by the majority as 'Best' and 'Worst'. The distortion came, we believe, from people 'second-guessing' which idea other people would choose rather than from trying to predict which idea they genuinely felt would be most successful in the market. In other words it made it a game with it's own internal market logic, rather than one linked directly to what will actually happen in the market.

Research Market 1 Results: - Diverse crowd of 500 people "betting" on the winning concept

That said, let's look at the results of 'Research Market 1' (Fig. 1) compared to the monadic test results. There were many measures on which the results could be compared but for the sake of simplicity and clarity we have used *Top Two Box Purchase Intention (P.I.)* as perhaps the single most common measure in concept tests.

The first thing to note is that the monadic test rated five of the 10 ideas as above average P.I. and the '*Research Market'* selected the same five ideas as above average. For two of the five there was no significant difference in the actual scores recorded i.e. the '*Research Market'* got it essentially perfect for two of the top five.

Although this sounds promising, how does it compare with what we would expect of a forced choice concept test? It is well known that forced choice tests, produce significantly more extreme (polarised) results than a monadic test. And sure enough, three of the top five ideas have a significantly higher PI than the monadic results and three of the bottom five have a significantly lower PI than the monadic test results. Therefore, the results seem to be very much in line with what we would expect from a forced choice test. However, we need to remember these results came from a 'crowd' rather than a representative sample of the target audience, providing some evidence of the 'crowd' effect as suggested by Surowiecki.

Fig. 1, Top Two Box Purchase Intention Results – Research Market 1*			
New Product Concepts	**Monadic Test** with matched samples of 100 in the target market (Top 2 Box Purchase Intent)	**Research Market 1*** With diverse group of 500 people - *betting (Top 2 Box Purchase Intent)	**Significant Differences**
A	85	86	
B	83	82	
C	81	90	+
D	78	91	**
E	74	69	
UK Norms (top 2 box)	67	67	
F	64	24	***
G	64	22	***
H	54	53	
I	49	39	***
J	43	17	
Respondent Base Sizes: Monadic = 100 per cell / Research Market *Betting = 502			

So how did 'Research Market 1' perform against our *Three Wisdom of Crowds* hypotheses:

1. *A diverse crowd of people will be as accurate as a targeted sample*
   ✓ = this does seem to be the case. The results were at least in line with a forced choice concept test with a representative sample of the target audience
2. *A forced choice 'Research Market' will be as accurate as a monadic test*
   × = although it did correctly identify the top five ideas and the exact score for two of those top five, it's performance overall was closer to the results we would expect from a forced choice concept test
3. *Asking people to speak for the market will be as accurate as a personal answer*
   × = as per 2 above

Research Market 2 Results - With diverse crowd of 500 people – no betting

Now let's look at what happened when we removed the betting mechanism that we suspected was encouraging participants to second-guess other people's answers rather than predicting which would be most and least successful in the market. (Fig.2)

Fig.2, Top Two Box Purchase Intention Results – Experiment 1*			
New Product Concepts	**Monadic Test** with matched samples of 100 in the target market (Top 2 Box Purchase Intent)	**Research Market 2*** With diverse group of 500 people - *no betting (Top 2 Box Purchase Intent)	**Significant Differences**
A	85	85	
B	83	76	+
C	81	80	
D	78	86	
E	74	70	
UK Norms (top 2 box)	67	67	67
F	64	28	***
G	64	28	***
H	54	35	*
I	49	45	
J	43	16	***
Respondent Base Sizes: Monadic = 100 per cell / Research Market *Without Betting = 507			

This time, for four out of the top five ideas, there was *no* significant difference between the 'Research Market' and the monadic test results and the difference for the fifth idea was only just significant. The 'Research Market' got it almost perfectly right. Yet this should not have happened - remember, this is a crowd of people rather than a representative sample of the target market and a forced choice test which is known to produce more extreme results.

It is worth pausing to note that unlike the almost perfect match for the top five, the P.I. for four of the bottom five ideas was significantly lower than the monadic test. Does this suggest the 'Research Market' is less accurate than a monadic concept test or is this more extreme 'punishing' of less good ideas an added benefit, in that it helps to sort 'wheat from the chaff' and focus the company's resources on ideas worth pursuing. If you think about the time and money most companies expend trying and failing to turn marginal ideas into winners, then a research mechanism which provides greater clarity on the fittest for survival could be of real advantage.

So were the almost perfect results for the top five ideas just a fluke? Possibly, but the odds are massively stacked against the crowd selecting and the matching the P.I. for all five of the top ideas by chance. It would seem that just as Surowiecki predicts and the existing decision markets discussed earlier have shown, the market mechanism in this case was working as a pretty good surrogate for monadic testing.

So how did 'Research Market 2' perform against our *Three Wisdom of Crowds* hypotheses:

1. *A diverse crowd of people will be as accurate as a targeted sample*
   ✓ = the 'Research Market' results were at least in line with a monadic concept test with a representative sample of the target audience. The principles behind sampling and the 'error' represented by the crowd should make this impossible. And yet just as Surowiecki predicts the errors in the sample are cancelled out, allowing an accurate answer to emerge
2. *A forced choice 'Research Market' will be as accurate as a monadic test*
   ✓ = this time the market was an almost perfect match with the top five ideas from the monadic tests. Again, this is an extremely surprising result, given what is known about forced choice approaches but perhaps the same 'cancelling out of error' mechanism is at work, once again allowing an accurate answer to emerge?
3. *Asking people to speak for the market will be as accurate as a personal answer*
   ✓ = as pre 2 above.

Given these results, it's worth remembering the words Galton wrote of his fairgoers, "The result seems more creditable to the trustworthiness of a democratic judgement than might have been expected." And as Surowiecki said, "That is, to say the least, an understatement."

It seems the mechanisms used in '*Research Markets*' can vary but as long as it is *diverse*, *independent* and *faithfully aggregated*, then the crowd can indeed be wise.

The Possible Implications for Market Research

Obviously more experiments are needed to validate the predictive potential of '*Research Markets*' but these results, together with the existing evidence from decision markets like, the IEM and HSX, would certainly merit further investigation.

If these results can be consistently replicated, there are clear cost, speed and capability advantages to using '*Research Markets*' over traditional MR methods, both for understanding the world *as is* and predicting the world *as it will be*.

If validated, there are some obvious clearer-faster-cheaper benefits of '*Research Markets*':

1. *Accurate but even clearer result*
   Matched PI with the monadic test but even greater clarity on the weakest ideas
2. *Quicker to set up and complete*
   Respondents are much easier to recruit, the market is quick and easy to set up,

fieldwork takes just a few days and the results are easily generated and analysed.

3. *Significantly cheaper to run*
   Smaller sample sizes, when testing larger numbers of ideas, easier to administer and analyse and all results from all tests are more easily comparable.

Questions Still to be Answered

If there is merit in this approach, there are a number of that still need answering:

1. Can this results of this experiment be replicated and the Research Market approach validated?
2. How many ideas can the respondents pick between, before the accuracy of prediction declines?
3. What's the minimum number of respondents required for an accurate result?
4. What's the minimum amount of knowledge required for the crowd to be accurate?
5. Is the market mechanism accurate enough to predict sales volume over time?

Summary

In a world where every research department is being asked to do more with less, without compromise and to do it faster than ever before, '*Research Markets*' would be a very appealing alternative if the approach could be validated.

Perhaps the last word should go to James Surowiecki, speaking of one of the most severe critic of the stupidity of groups, "Gustave Le Bon had things exactly backward. If you put together a big enough and diverse enough group of people and ask them to 'make decisions affecting matters of general interest,' that group's decisions will, over time, be 'intellectually [superior] to the isolated individual,' no matter how smart or well-informed he is."

Time and more experimentation will hopefully judge the wisdom of his words.

Appendices:

References:

1. Forsythe, Robert / Nelson, Forest / Neuman, George R. and Wight, Jack (1992), Anatomy of an Experimental Political Stock Market http://www.biz.uiowa.edu/iem/archive/forecasting.pdf
2. Johnson, Steven (2001), Emergence
3. Malony, Michael T. and Mulherin, J. Harold, (1988) The Stock Price Reaction to the Challenger Crash: Information Disclosure in an Efficient Market http://ssrn.com/abstract=141971
4. Surowiecki, James (2004), The Wisdom of Crowds

Author

John Kearon, **Chief Juicer;** Founder, CEO and Chief Juicer of BrainJuicer; Europe's fastest growing research agency. John was recently named 'Entrepreneur of the Year' and his company BrainJuicer won, 'The Most Innovative Use of IT' and 'Service Business of the Year' 2004. John has formerly worked in senior marketing roles at Unilever, was Planning Director at Publicis ad agency and previously founded Innovation agency, Brand Genetics Ltd.