

PREDICTIVE MARKETS

Utilizing the wisdom of crowds to slay the sacred cow of scientific sampling

John Kearon, Chief Juicer, BrainJuicer Group PLC, 13-14 Margaret Street, London, W1W 8RN

John Kearon is founder and CEO of BrainJuicer and a frequent speaker at market research conferences. He has presented at ESOMAR, MRS and many other conferences on their pioneering Predictive Markets techniques. His work won the ESOMAR Best Methodology Paper in 2005 and the ARF Great Mind Award in 2007.

john.kearon@brainjuicer.com

Introduction: The Potential of Predictive Markets

As the chill winds of the global economic crisis blow through the corporate world, market research departments everywhere are being challenged to do more with less, whilst increasing their impact on innovation success.

Predictive Markets is an intriguing new approach to screening early stage concept, claims and positionings which more effectively sorts wheat from chaff, helping companies say no to 'deeply average' ideas, ensuring only strong ideas are progressed, thereby reducing the time and money required to get successful innovation to market¹.

The challenging supposition behind *Predictive Markets* is that a large diverse group of people [a crowd], buying and selling notional shares in 'ideas' can predict what other people will do, just as accurately but with far greater discrimination than traditional gold standard concept testing approaches with scientifically sampled target audiences.

Predictive Markets emerged from experimental economics before being popularised in James Surowiecki's, best selling 2004 book, *Wisdom of Crowds*². Surowiecki's challenging assertion is that crowds make better decisions than experts, when they are diverse, independent and faithfully aggregated. Although he focussed on the flaws in expert opinion, the evidence he puts forward also suggests another possibility. Namely, a crowd of people operating through a notional 'Market' mechanism could be just as accurate as our best research approaches.

¹ J. Kearon (2005) Re-engineering Research to Take Advantage of The Wisdom of Crowds, ESOMAR Congress Papers 2005.

² J.Surowiecki (2004) *The Wisdom of Crowds. Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*

Now this is intriguing since *Predictive Markets* break at least three golden rules of quantitative testing and really shouldn't work. *Predictive Markets* use a random, diverse crowd rather than a targeted sample; they force a choice between many options rather than getting responses to just one concept; and they ask respondents to guess what other people will do rather than what they would do themselves. Despite breaking the rules, there is a rapidly growing body of empirical evidence in economics, political science and business that suggest they work well and are have more potential than our traditional methods. In fact our own 279 head-to-head concept testing experiments across 15 different categories in as many countries have produced the following conclusions:

1. **As accurate as Monadic testing.** *Predictive Markets* have picked out the same top quartile concept winners as gold standard monadic concept testing.
2. **More discriminating than Monadic testing.** *Predictive Markets* separate good from average ideas with far greater discrimination than conventional monadic concept testing.
3. **Identifies potential breakthrough ideas.** *Predictive Markets* ensure companies don't miss polarising ideas with low average scores that might have breakthrough potential.
4. **Comparable across markets and categories.** *Predictive Markets* achieve very similar scores across countries/categories, removing the need for normative databases at the individual country level.

Most of our experimental work has been done in concept screening but there is good evidence to show these conclusions hold true for a number of applications, including claims testing, assessing alternative positionings, promotions screening and even brand share forecasting.

More experimentation and validations are required to explore the full potential of Predictive Markets and to know where they don't work and when not to use them but they do seem to represent a tremendous opportunity to do more with less, whilst increasing market research's impact on innovation success.

Crowds are not traditionally associated with wisdom

Crowds are not traditionally associated with wisdom. The case against them has been put by a number of thinkers and writers over the years. The historian Thomas Carlyle wrote, "I do not believe in the collective wisdom of individual ignorance". And perhaps 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 normal distribution curve of performance where the average score would be significantly below that of the best performers. So it is rather counterintuitive to think a group might somehow perform better in aggregate than the best individual performers. Yet this is exactly what James Surowiecki in his book, *The Wisdom of Crowds* shows is possible so long as the crowd is *diverse, independent* and *faithfully aggregated*. In the TV game show, *Who Wants to be a Millionaire*, the audience consistently outperform the 'expert friends' suggesting a diverse, independent audience's aggregated answers to an IQ test would produce a score significantly higher than the average score of independent test takers and might well be close to the brightest performers. It seems we can be collectively wise.

Surowiecki's challenging assertion is that crowds make better decisions than experts, when they are diverse, independent and faithfully aggregated. A great example comes from British scientist, Francis Galton who was actually trying to prove the opposite of what he found. 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". In fact 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 placed 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.

**collective
guess
=
1197 lbs**



**actual
weight
=
1198 lbs**

Figure 1. Galton's surprise findings

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. Sociologist Kate H. Gordon has replicated these findings with the modern version of the Galton's ox, namely, 'guess the weight of the cake' or how many jelly beans are there in the jar'. 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.

Predictive Markets it seems is a classic example of 'We' research³, enabling a diverse group to place their bets or cast their votes to predict what other people will do.

Perhaps the most compelling and longest running example of *Prediction Markets*, is the Iowa Electronic Markets (IEM) founded in 1988 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. Over more than 700 elections, the IEM has been more accurate than the most accurate poll, 75% of the time which is a dramatic finding, considering polling is the gold standard approach to election prediction⁴.

Surowiecki cites several other examples of successful *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).

³ M.Earls and J.Kearon (2009) Me-to-We Research: From asking unreliable witnesses about themselves to asking people what they notice, believe & predict about others, ESOMAR Congress paper 2009.

⁴ Berg and Rietz (2005-01-00), The Iowa Electronic Market: Lessons Learned and Answers Yearned

There are some 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 *Predictive Markets* as a source of information and almost 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.”

Applying the Wisdom of Crowds to Market Research

In his book, Surowiecki showed how Predictive Markets have been utilised for successful prediction in many different worlds and other published papers showed similar success, so the question for us became, why not market research? The approach may break many of our golden rules of quantitative research and challenge the need for scientific sampling but the empirical evidence was certainly strong enough to suggest experimenting to see if it worked if market research.

The first documented research application was our own Predictive Markets paper in 2005, which won ESOMAR's Best Methodology award that year. What the experiment showed was a large group of 500 people (not the target audience), buying and selling notional shares in ideas, can predict the success of new concepts just as accurately as a scientifically sampled audience but with far greater discrimination than conventional concept testing.

In this first 2004 experiment, we tested the thinking by evaluating 10 new product concepts in a crowd-based Predictive-Market of 500 hundred diverse respondents and the same concepts in 10 separate classic monadic concept test with a scientifically sampled audience. Effectively we were testing the following;

1. Could a diverse crowd be as accurate as a carefully sampled target audience?
2. Could buying & selling ideas prove as accurate as a classic monadic concept test?
3. Could judging on behalf of the market be as accurate as answering for oneself?

The results were compelling. The *Predictive Market* selected the same top five ideas, all of which were top quartile performing concepts in the monadic test with almost no significant difference in purchase intention scores [Figure 1]. But the *Predictive Market* also proved much more discriminating giving significantly lower scores for four out of the remaining five concepts and giving clear guidance for which ideas to drop.

Figure 2 is the head-to-head results for the 10 household ideas

<u>New Product Concepts</u>	<u>Monadic Test</u> with matched samples of 100 in the target market (Top 2 Box Purchase Intent)	<u>Predictive Market</u> With diverse group of 500 people (Top 2 Box Purchase Intent)	<u>Significant Differences</u>
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 / Predictive Market = 507

Figure 2. 2005 Original Monadic v. Predictive Markets test results

At the end of the 2005 paper we set ourselves the challenge of repeating the experiment at least 10 times with different categories and varying numbers of ideas to see if the results were consistent whether we could learn more about the nature & perhaps limitations of Predictive Markets. The subsequent experiments supported the results from the first and were written up in a 2007 ESOMAR Congress paper⁵.

Since 2007, we have continued experimenting, torture testing the approach and adding further validations. From all the data gathered over the last 5 years, there are now 279 head-to-head experiments versus gold standard monadic or sequential monadic concept testing methods, across 15 different categories and as many countries, from which we have drawn the following conclusions for *Predictive Markets*:

Conclusion 1: As accurate as gold standard monadic concept testing

Across over 180 new product ideas, Predictive Markets have picked out almost exactly the same top quartile concept winners as gold standard monadic concept testing (0.91 correlation). With sequential monadic the comparison has been a 0.81 correlation, which may say more about the accuracy of sequential testing rather than Predictive Markets. (Figure 3)

Monadic	Ideas x	No. of Green Light Concepts
Snacks	15	5 out of 5
Cleaners	10	5 out of 5
Air Care	5	2 out of 2
Drinks	8	1.5 out of 2#
Waters	11	1 out of 2#
	x48	9.1 out of 10
Sequential		
Telecoms	42	8 out of 12
Chilled Food	11	3 out of 3
Cleaners	29	6.5 out of 9
Gum	7	4 out of 4
Lighting	20	9 out of 10
Air Care	28	5 out of 7#
	x137	8.1 out of 10

Figure 3. Number of top quartile concepts the same in the head-to-head tests

⁵ J.Kearon (2007) Predictive Markets: Is the Crowd Consistently Wise?, ESOMAR Congress paper 2007

In other tests of accuracy, the Predictive Markets approach has proved successful at the results of public events, including the TV talent show X Factor, where the crowd predicted Leona Lewis' success from the first week and early enough to beat the odds at the bookies (Figure 4). In this case, the bookies are the equivalent of the traditional monadic approach and the crowd's edge over the bookies represents a real opportunity to use this technique to better predict public events, including elections – but that will have to be the subject of more experiments and another paper.

<u>Weeks</u>	<u>£10 to Win</u>	<u>Odds</u>	<u>£ Winnings</u>
1	£10 Leona / £10 Kerry	7-1 / 5-1	£60
2	£10 Leona	3-1	£40
3	£10 Leona / £10 Ben	3-1 / 6-4	£30
4	£10 Ben	6-4	-
5	£10 Leona	2-1	£30
6	£10 Leona	10-11	£11
7	£10 Leona	1-2	£15
8	£10 Leona	2-7	£12
9	£10 Leona	1-4	£12
10	Leona to win	No odds	-
TOTAL	(£120) bet		£210 won

Figure 4. Crowd Prediction each week for X Factor talent show

The crowd was also wise in predicting the success of the 17 films released for Christmas 2007, achieving a 0.81 correlation with the first week box office sales. (Figure 5) There were two films which significantly reduced the correlation; Alvin and the Chipmunks which was the surprise hit of the season. Interestingly this film was the most highly polarizing with the crowd, with a significant number selecting it as potentially the most successful but an equal number selling all their shares in the film because they thought it would be a stinker (Figure 11). This sort of polarization, explained in more detail in conclusion 3 below, is the basis of *Predictive Markets'* capability to identify potentially breakthrough ideas – as Alvin and the Chipmunks was in this case. The one film the *Predictive Market* missed was Alien vs. Predator which did far better than the crowd predicted but the rest were highly correlated with box office sales.

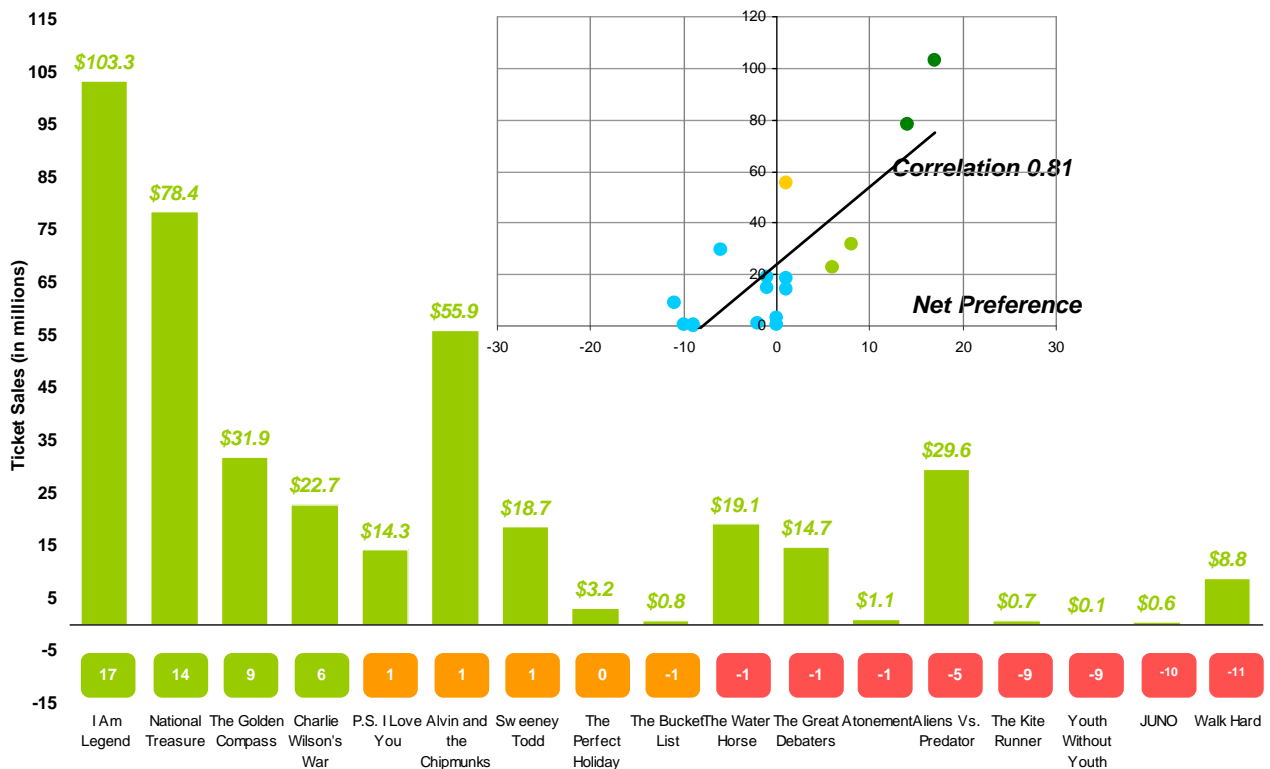


Figure 5. Predictive Markets on 17 films released pre-Xmas 2007

In an extreme test of *Predictive Markets* accuracy, a large pharmaceutical company wanted to see if the approach could possibly work with prescription drug concepts. So we tested 6 concepts previously tested with specialist rheumatologists and picked out the same 3 top quartile performers in the same order (Figure 6). This was repeated twice more with concepts tested previously with cardiologists and diabetes specialists with the same results. The company estimated that because of the very high cost of interviewing medical specialists, using the *Predictive Markets* approach would save them 90% of their current concept testing costs.

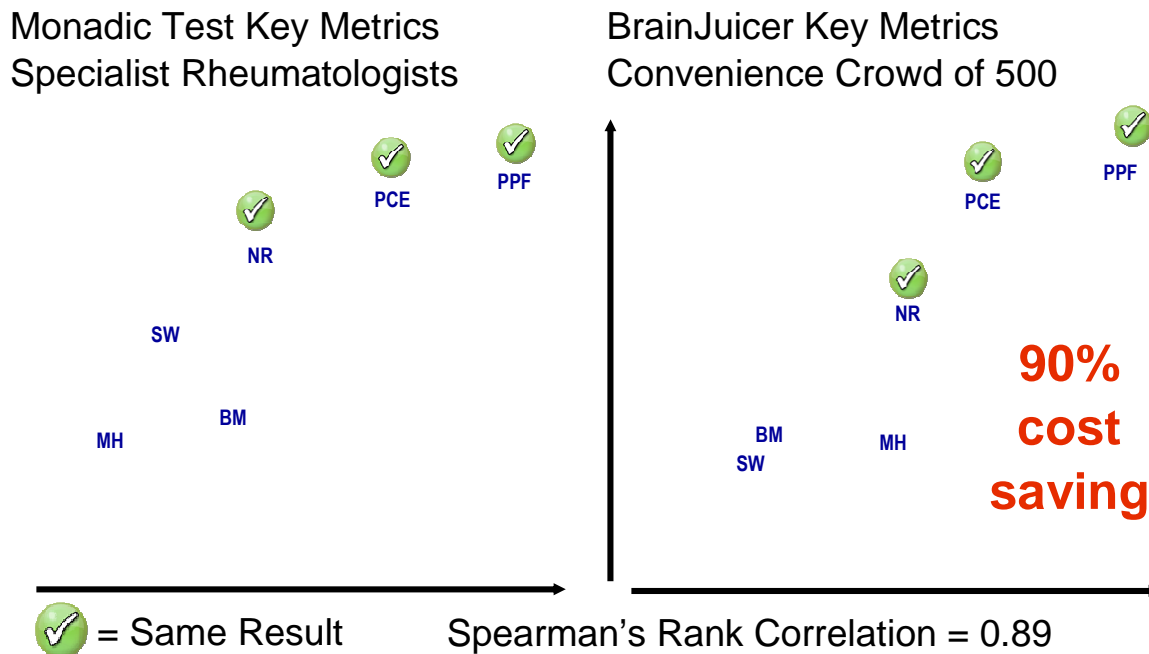


Figure 6. Head-to-head test of prescription drug concepts

Conclusion 2: More discriminating than Monadic testing

Predictive Markets separate good from average ideas with far greater discrimination than conventional monadic concept testing. Anyone who's done a lot of concept testing knows, the biggest problem with classic monadic tests is that scores tend to drop to the mean. You put in six reasonable concepts and get six back with average scores and little or no significant difference between them. This makes it hard to decide which to drop and which to progress and often average concepts have more resources invested in them, long after they should have been shelved.

The forced choice nature of a *Predictive Market* where people have to decide which is best and worst concept i.e. which one they would choose to double their shares in and which one they'd choose to sell all their shares in. This creates far greater discrimination between ideas and makes it much easier to say no to ideas that will always be 'deeply average' and not worth investing in any further.

Figure 7 shows a typical output and the clear spread of ideas with recommendations for which to progress, which to shelve and which to give further thought to before deciding to drop or progress.

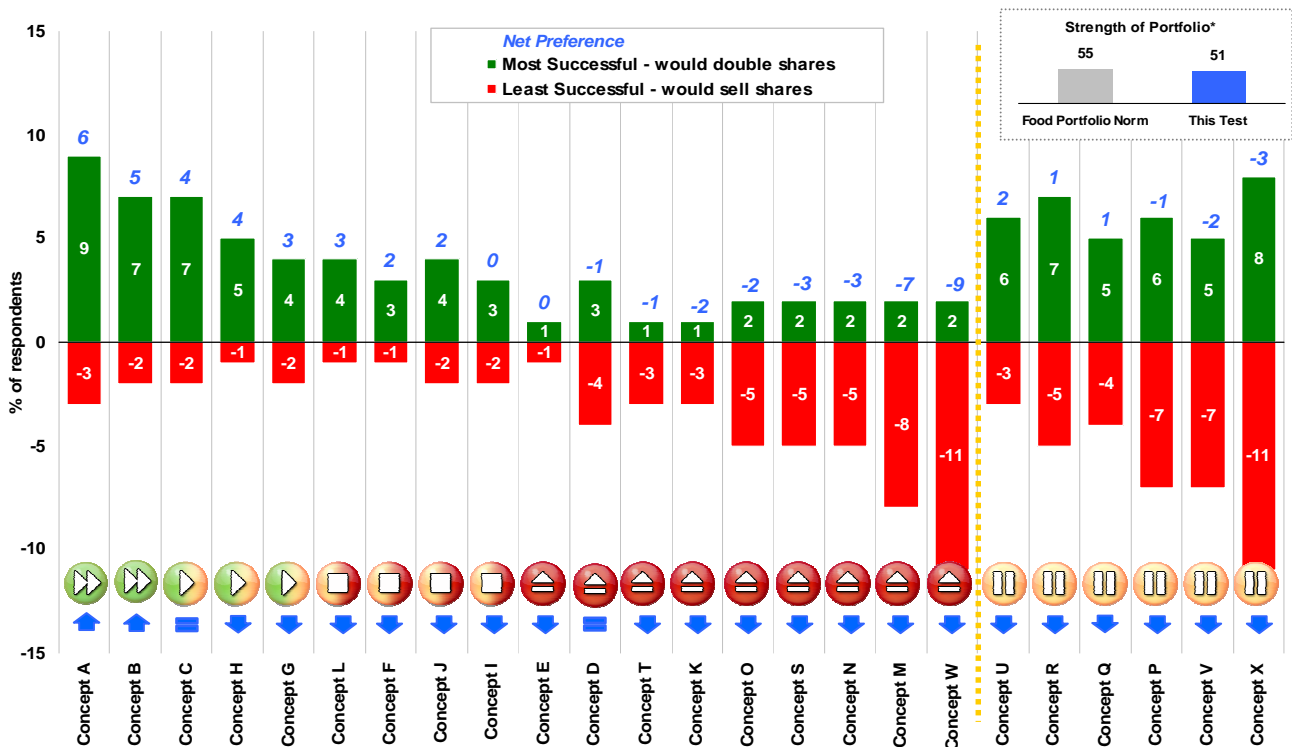


Figure 7. Typical Predictive Markets output of Net Preference i.e. buyers minus sellers

Figure 8 shows a comparative study on seven children's food products, a category with notoriously high normative scores and where the tendency for children is to rate everything highly as in this case where there was no significant difference between the seven concepts. By comparison, the Predictive Market clearly discriminated between the three concepts worth pursuing and the four to drop.

Figure 9 shows the significantly greater range of scores and discrimination achieved by Predictive Markets across a large number of tests.

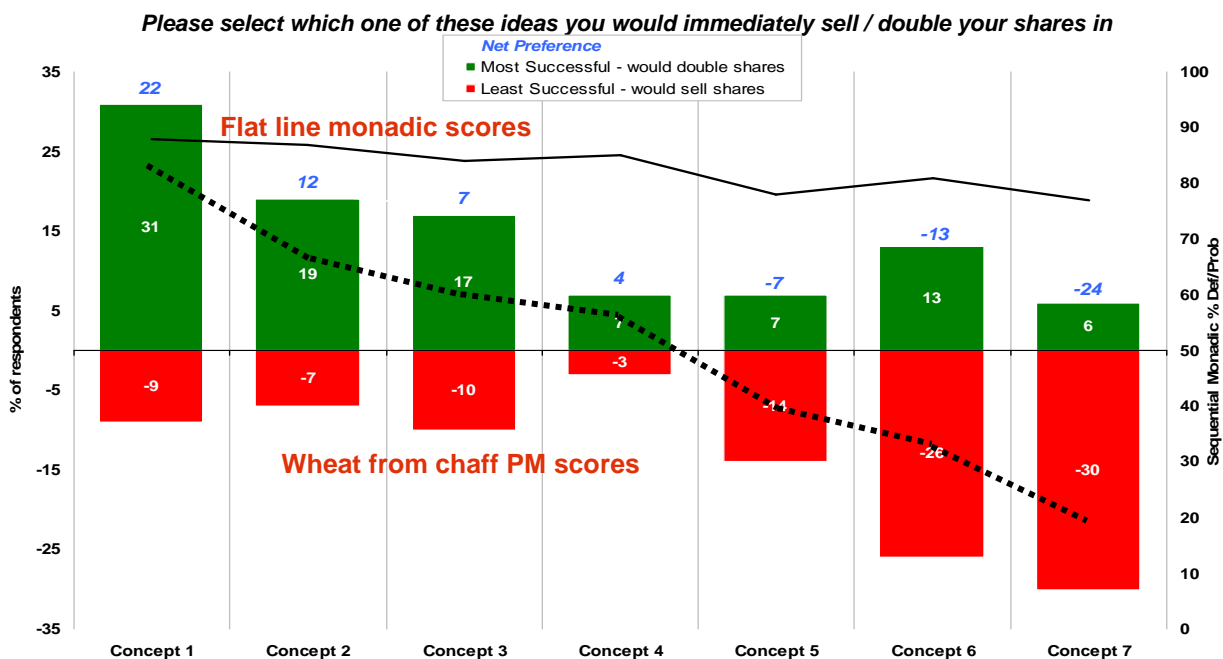


Figure 8 & 9. Far greater discrimination between Predictive and Monadic test results

<i>Monadic</i>	<i>Ideas Tested</i>	<i>Predictive Markets Net Preference Range</i>	<i>Monadics PI Range</i>
<i>Air Fresheners</i>	5	19.6	4.8
<i>Savoury snacks</i>	15	26.0	6.1
<i>Cleaners</i>	10	26.9	10.7
Average		24.2	7.2
<i>Sequential</i>			
<i>Telecoms UK</i>	14	12.8	7.8
<i>Telecoms France</i>	14	14.2	8.9
<i>Telecoms Spain</i>	14	20.3	18.4
<i>Chilled Food</i>	11	14.3	14.0
<i>Window cleaners</i>	17	27.6	18.1
<i>Chewing gum</i>	7	20.8	4.1
<i>Cereals</i>	11	17.2	2.0
Average		18.2	7.3

The question that arises from experienced market researchers is, ‘what happens in extremis if all the ideas tested are poor, because the forced choice approach of insisting on respondents doubling their shares in one concept would over-reward that concept on the Net Preference chart in Figure 7 or 8. This is a good question and an accurate observation, so to counteract this potential for over or under reading the potential of concepts, we introduced a ‘probably buy or probably sell’ question. The first thing respondents have to do for each concept being tested is to select whether they would ‘probably buy’ or ‘probably sell’ shares in the idea. Not how many shares, just a binary decision on whether they believe other people would like or dislike the idea. The percentage of people choosing to ‘probably buy’ turns out to be the equivalent of the top two box purchase intention score (see Figure 10). The ‘probably buy’ score turns out to be normative and indicative of whether the idea is a genuinely top quartile or not. If all the ideas tested were poor, then none of them would have a ‘probably buy’ score above the norm.

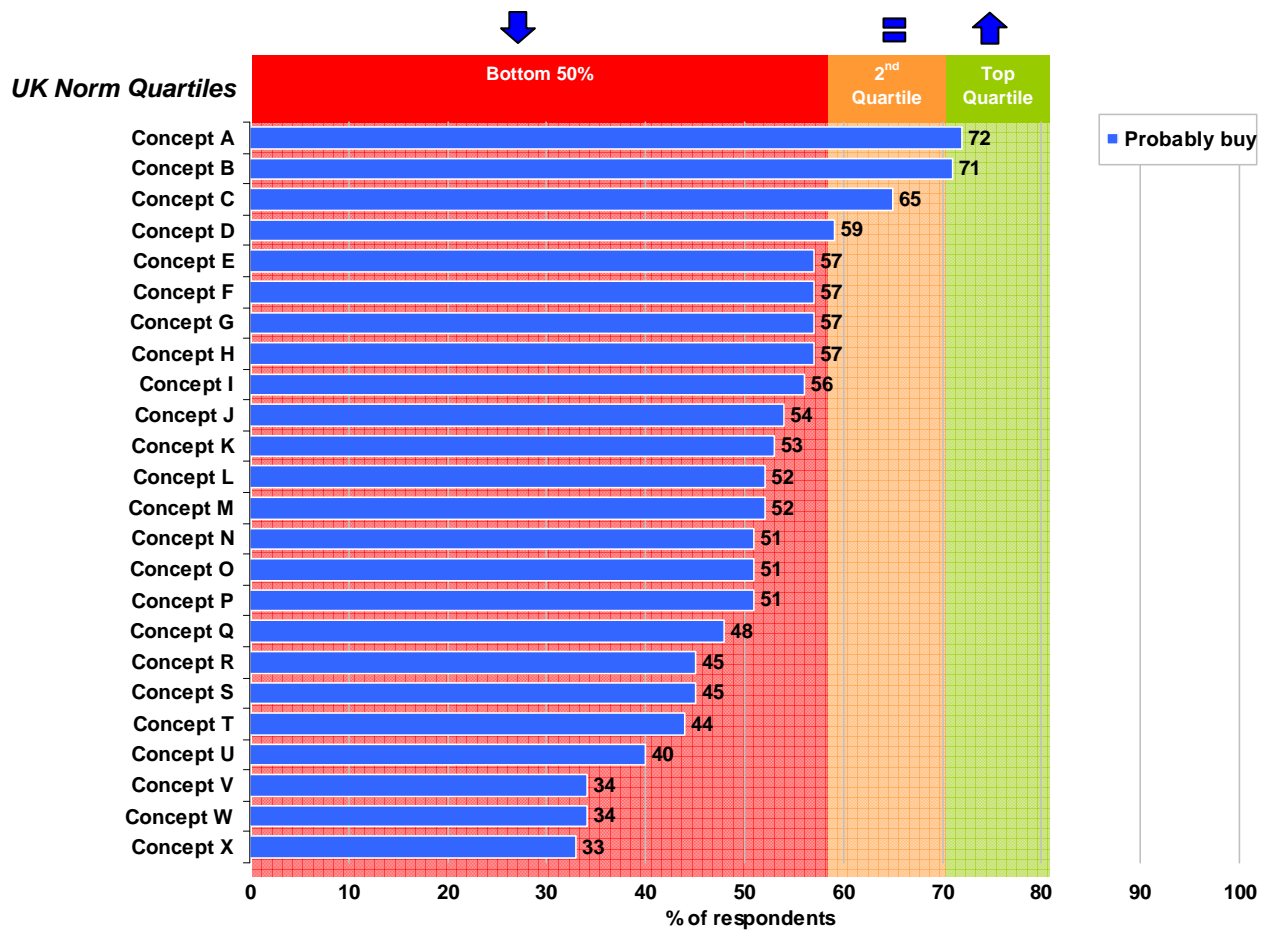


Figure 10. The % of respondents choosing to 'probably buy shares' in the concept

The 'probably buy' score also turns out to be highly stable. Figure 11 shows the 'probably buy' score for the same concept retested among a higher scoring set of concepts and again among a lower scoring set. Reassuringly, the 'probably buy' score remains almost identical despite the surrounding concepts changing in quality, providing confidence it is a normable score and indicative of a concept's overall potential.

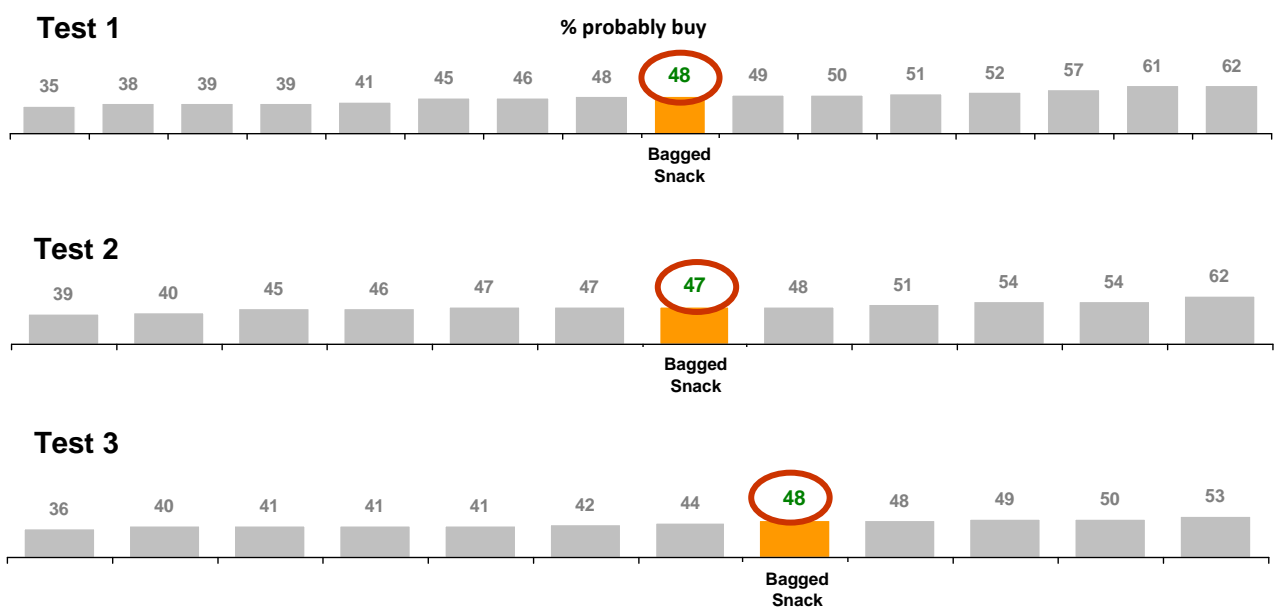


Figure 11. The same concept retested among better and worse scoring concepts

Conclusion 3: Identifies potential breakthrough ideas

Predictive Markets ensure companies don't miss polarising ideas with low average scores that might have breakthrough potential. Another problem with classic concept testing is the danger of controversial ideas getting a low average score and being dismissed without further consideration, even though some people like it very much and it might just have potential as a breakthrough idea. Ideas that change people's behaviour, like an iPod and the Sony Walkman before it, have the potential to change and challenge a whole category but they will always be controversial and almost never test cleanly as a top quartile concept. The Walkman famously bombed in research but went on to decades of commercial success and there are many other ideas like Bailey's Irish Cream (the world's biggest selling liqueur) and the first super premium crisps in the UK (Phileas Fogg) that failed in research but which on closer inspection showed enthusiasm among a smaller group which at least indicated its polarising potential.

In *Predictive Markets*, respondents doubling and selling notional shares in the concepts helps identify any highly polarising ideas by highlighting those which get a significant amount of attention, even though it is equally negative as positive. These ideas can be thought of as, 'pause for thoughts' and their polarising nature should encourage a little more investigation before choosing whether to discard or perhaps pursue as a potentially breakthrough idea?

The example of the Christmas release films in Figure 12 highlighted Alvin and the Chipmunks as a highly polarised and in fact proved to be the surprise box office hit of that season.

Alvin and the Chipmunks
 Starring: David Cross, Jason Lee, Cameron Richardson, Jane Lynch, Ross Bagdasarian

Alvin and the Chipmunks, a global phenomenon to generations of fans, become a live action/CGI motion picture event with a contemporary comic sensibility. Songwriter Dave Seville (Jason Lee, My Name Is Earl) transforms singing chipmunks Alvin, Simon and Theodore into pop sensations – while the out-of-control trio lays waste to Dave's home, wreaks havoc with his career, and turns Dave's once-orderly life upside-down.

Rated: PG
 Release Date: December 14th

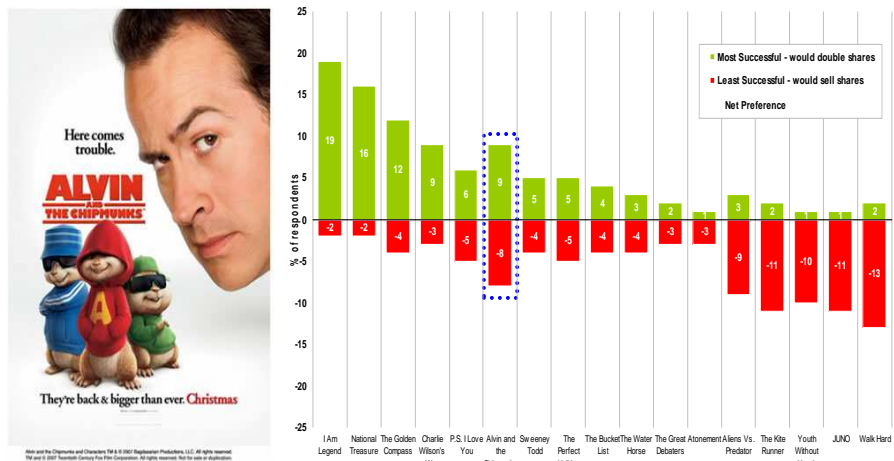


Figure 12. Typical polarizing idea with the potential for breakthrough success

Conclusion 4: Comparable across markets

Predictive Markets achieve very similar scores across countries/categories, removing the need for normative databases at the individual country level. In classic monadic concept testing, scores vary significantly by country and category. In fact as Figure 13 shows, the concept norm in South America is 45 percentage points higher than Europe and North America. By contrast, the equivalent difference across the same concepts and markets for *Predictive Markets* is just 12 percentage points, significantly reducing

the danger of cultural effects in the scores and largely removing the need for large databases of normative data for every country. It would seem that the combination of answering for others and being forced to buy and sell an idea largely removes the large cultural differences usually seen between how concepts are rated. We have seen the same pattern when it comes to different categories which also vary in classic monadic by as much as 25% between children’s products like chewing gum and domestic appliances like lighting. The equivalent difference for *Predictive Markets* across the same categories is just 8%.

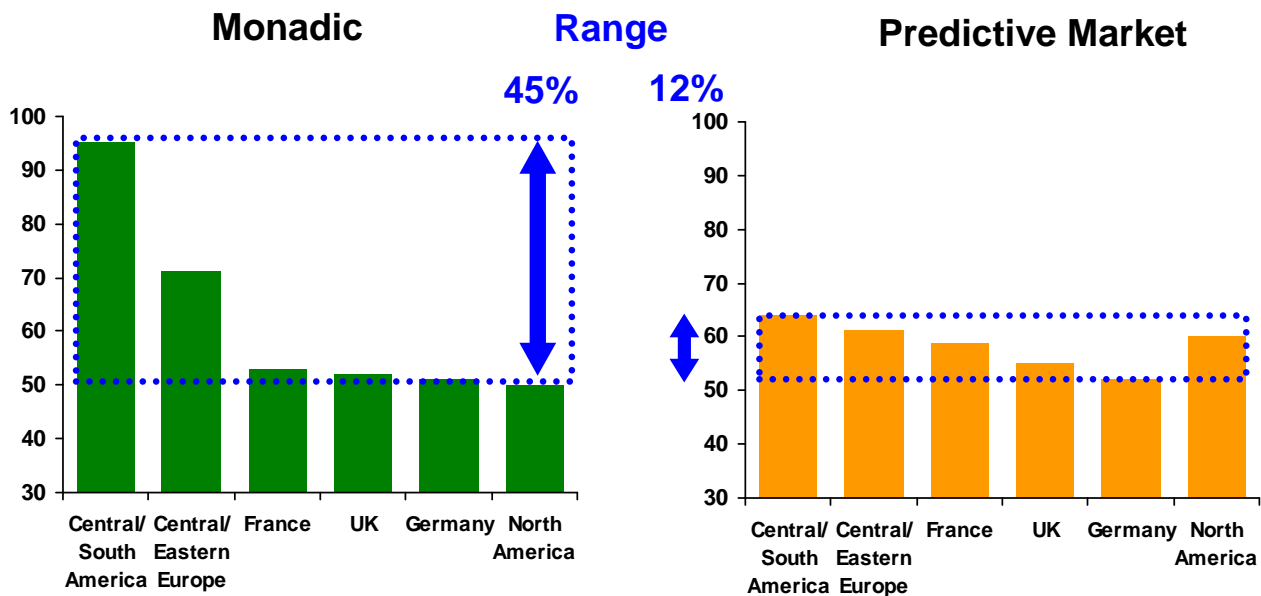


Figure 13. The difference between normative country scores between Monadic tests and PMs

How do Predictive Markets Work?

The natural question that arises is how and why do Predictive Markets work? The answer unfortunately is far from straightforward and not entirely agreed on by academics. One explanation is as social animals we’ve evolved the ability to be extremely adept at judging what other people are likely to do. In fact there is good evidence from the psychological literature to show that we may be better at judging what other people will do than we are at judging what we’ll do ourselves, since our arm’s length view of others avoids the little delusions we carry around about ourselves. This certainly explains the occasions in quantitative research when we see a marked difference between what people say they’ll do and what they then actually do.

A fuller and more mathematical explanation of how Predictive Markets and the wisdom of crowds works is by Scott E. Page, Professor and Associate Director of the Centre for Complex Systems at the University of Michigan. His book, ‘The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies, lays out his underlying theory. He starts from basic arguments about heuristics – mental tools or rules of thumb for dealing with complex situations – and perspectives, basic mappings of reality. A heuristic tells individuals how to search for a solution to specific challenges. Page⁶ argues that individuals will build from their perspectives and interpretations towards quite different predictive models and aggregating many

⁶ Page, Scott E. (2007) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*

the different predictions, cancels the individual errors and produces a collectively wise result – just as many hundreds of small mirrors looking at the stars from different locations are the way that astronomers now pinpoint more accurately the positions of astral bodies. Page recreates the conditions under which Surowiecki showed that groups of diverse individuals will be able to outperform groups of ‘experts.’ He uses some simple agent-based modelling to support this claim, and outlines the underlying logic of a “mathematical proof that provides sufficient conditions for the result.” The reason why this holds is because experts tend to be quite like each other, so that they tend to converge on the same solution, while heterogeneous agents have a better chance of coming across other, superior solutions. In terms of market research, this is the diverse crowd instead of the expertise of the target audience used in classic monadic testing. In Page’s words, ‘when solving problems, diversity may matter as much, or even more than, individual ability’.

In a number of our *Predictive Market* experiments, we’ve increased the size of the crowd taking part from 500 to 700 or even 900, to be able to see how the results differ by sub groups and even target audiences within the crowd. As the example in Figure 14 shows, there is a reassuringly high correlation between any sub group and the overall result but the closer the sub group gets to the target audience for the ideas, the lower the correlation gets. The findings support Professor Page’s work that groups of experts produce narrower predictions which tend to be less accurate than those of a more diverse group. As such, it is important to run *Predictive Markets* with a diverse crowd which can include a proportion of those in the target audience but should not be run only with the target audience, since the results become less accurate. We saw the same result in another experiment involving a *Predictive Market* for 18 new product concepts for Marmite; the polarising yeast spread, which advertises itself as ‘you either love it or hate it’. The crowd split roughly 50:50 into lovers and haters and together their prediction of top quartile performers was highly correlated with a monadic comparison. However, the results from Marmite lovers only proved less accurate than the crowd as a whole, just as we had found in the Figure 14 example.

Naturally enough, there are companies experimenting with live trading markets as an approach to Predictive Markets and assume their greater sophistication means greater accuracy in prediction. But this has not been our experience. Our evidence suggests trading markets are vulnerable to exaggerating herding effects where the crowd’s starting diversity narrows very quickly as players see which ideas are being punished and rewarded by others and follow suit. The evidence suggests it is more consistently accurate, to avoid trading markets and use a blind, democratic voting mechanism, where each person makes their prediction blind to the prediction of others in the crowd.

27 Claims	Total Gen Pop	Open Non Category Users	Category Users	Moms Open Health Seekers	Moms with Kids Category Users
A	74	70	81	80	84
B	71	73	75	75	78
C	73	74	75	81	87
D	72	72	82	77	83
E	59	55	69	63	69
F	59	57	65	61	63
G	66	65	69	71	64
H	68	67	76	74	75
I	70	70	72	68	71
J	62	57	70	68	76
K	58	55	64	67	59
L	56	52	67	56	62
M	60	60	61	67	66
N	57	54	66	51	59
O	63	63	65	75	60
P	54	50	63	57	70
Q	63	62	66	61	57
R	55	56	60	59	59
S	53	49	61	56	62
T	55	54	65	53	56
U	52	49	58	58	71
V	53	53	60	62	63
W	56	58	59	62	56
X	37	39	39	40	37
Y	51	52	56	59	56
Z	41	40	43	41	39
AA	21	25	18	20	14
Correlation	NA	0.98	0.97	0.95	0.91

Figure 14. Sub-group analysis of correlation with overall results

Other Applications for Predictive Markets

The focus of BrainJuicer’s experiments and the bulk of the evidence shared in this paper demonstrate the potential of *Predictive Markets* for claims, names and concept screening. Not surprisingly given its similar nature, the approach has also proved its worth in screening promotional, brand activation ideas, as well as assessing different positioning options for brand restages and/or new brand launches. It is worth noting that companies, who use *Predictive Markets* for screening, still use a classic monadic approach with their target audience for validating later stage concepts but find PMs a better way to screen for those that perform well in later stage testing.

It is still early days, but our experiments to date suggest *Predictive Markets* also have strong potential for demand forecasting. The accuracy of the crowd’s prediction of Leona Lewis as the X-Factor winner from week 1 enabled them to beat the bookies and make money from their collective wisdom (Figure 4). The crowd’s

prediction was also highly correlated with the box office success of 17 films released pre-Xmas 2007 (Figure 5). But we decided to test how accurate a *Predictive Market* could be in predicting the actual brand share of a new luxury chilled dessert being launched by Unilever. As you'll see in figure 15, the crowd's prediction was on or just below a 2.0% value share and the actual share achieved was 1.75%. A parallel Microtest volumetric test predicted also predicted a 2% value share. Unfortunately for the brand, it needed a 3% value share to maintain listings and was removed from the market after 18 months but it did demonstrate the technique's strong potential for demand forecasting. We are undertaking more demand forecasting experiments, including some with Roche to predict sales of newly launched prescription drugs and we will report the results in a subsequent paper.

The other application showing promise, is to use *Predictive Markets* for early stage screening of communications. In two experiments, a large multinational had received below average scores on a copy test of an animatic and needed to improve the messaging. A range of potential messages were screened in a Predictive Market and the most compelling message was used to revise the animatic and retested. The revised animatic scored in the top 5% of all adverts the multinational had ever copy tested. This was repeated for a second brand campaign to similar effect.

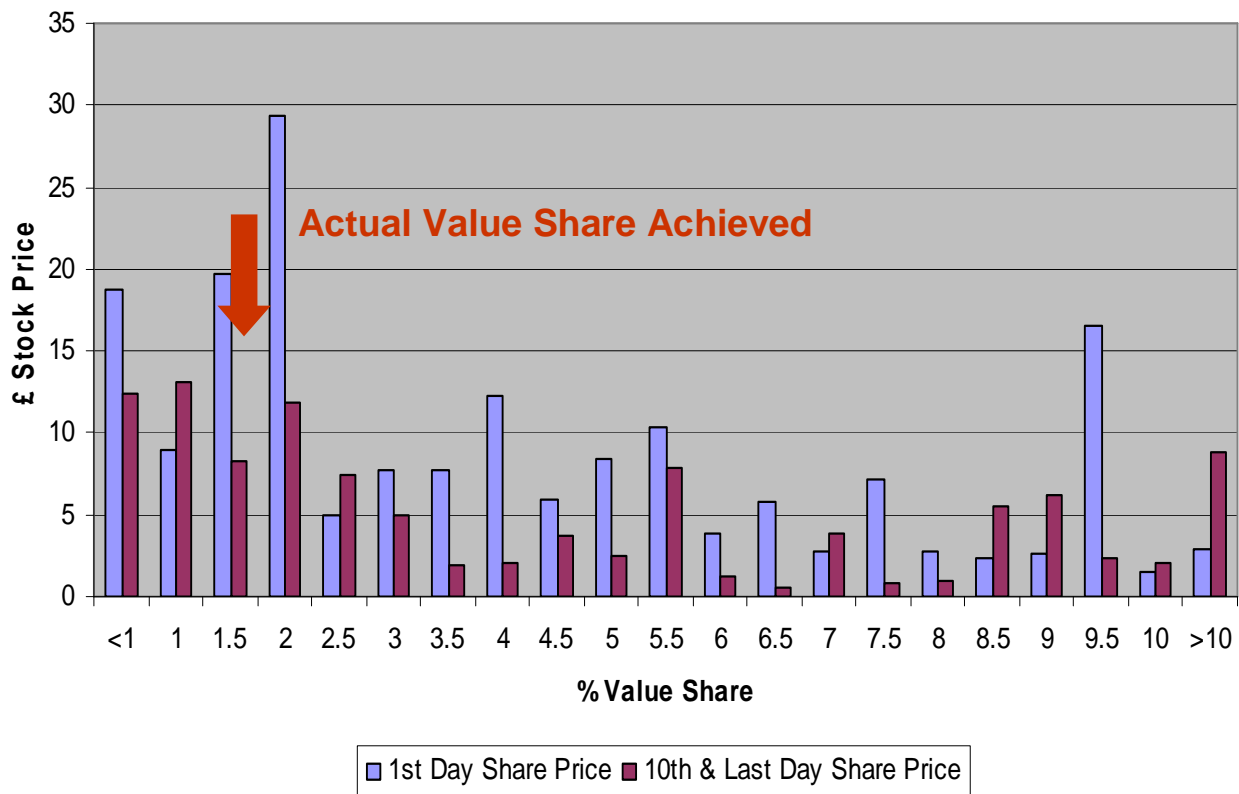


Figure 15. Actual Value Share Achieved vs. Predictive Market

Conclusions

The ability to spot winning ideas, help companies 'say no' to deeply average ideas and identify potential breakthrough ideas, makes *Predictive Markets* an exciting new approach to early stage screening of claims, names, concepts, positioning's and promotions.

But *Predictive Markets* are also controversial. They break almost every golden rule of classic monadic concept testing, including replacing the scientifically sampled audience with a diverse crowd answering for the market rather than for themselves.

Predictive Markets' potential and controversy makes it vital that practitioners invest the time and resources required to validate the approach and to learn on what and when it works. Our evidence from 279 head-to-head concept testing experiments across 15 different categories in as many countries, strongly suggest the following conclusions:

- 1. As accurate as Monadic testing.** *Predictive Markets* have picked out the same top quartile concept winners as gold standard monadic concept testing.
- 2. More discriminating than Monadic testing.** *Predictive Markets* separate good from average ideas with far greater discrimination than conventional monadic concept testing.
- 3. Identifies potential breakthrough ideas.** *Predictive Markets* ensure companies don't miss polarising ideas with low average scores that might have breakthrough potential.
- 4. Comparable across markets and categories.** *Predictive Markets* achieve very similar scores across countries/categories, removing the need for huge normative databases.
- 5. Potential for use in demand forecasting.** *Predictive Markets* have shown early promise in predicting value shares of new concepts.

More experimentation and validations are required to explore the full potential of *Predictive Markets* and to know where they don't work and when not to use them but they do seem to represent a tremendous opportunity to do more with less, whilst increasing market research's impact on innovation success.

Bibliography

Berg and Rietz (2005-01-00), The Iowa Electronic Market: Lessons Learned and Answers Yearned

Christiansen, Jed (2006), Prediction Markets: Practical Experiments in Small Markets and Behaviours Observed

Earls, Mark and Kearon, John (2009) Me-to-We Research: From asking unreliable witnesses about themselves to asking people what they notice, believe & predict about others, ESOMAR Congress paper 2009.

Forsythe, Robert / Nelson, Forest / Neuman, George R. and Wight, Jack (1992), Anatomy of an Experimental Political Stock Market

Kearon, John (2005), Re-Engineering Quantitative Research to Access the Wisdom of Crowds, ESOMAR Congress

Kearon, John (2006), Predictive Markets; A Fresh Approach to Concept Testing, MRS Conference

Kearon, John (2007), Predictive Markets: Is the Crowd Consistently Wise? ESOMAR Congress

Manki, Charles (2004-01-00) Interpreting the Predictions of Prediction Markets

Page, Scott E. (2007) The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies

Surowiecki, James (2004), The Wisdom of Crowds

Wolfers and Zitzewitz (2004-05-00), Prediction Markets