

# Rethinking prediction: Were you still up when Bob called it for Kerry?

Mark Earls

Source: Market Research Society, Annual Conference, 2005

Downloaded from WARC



Discusses, in the context of the 2004 U.S. presidential election the difficulties of predicting human behaviour, why we expect too much of research data, and why polls and surveys are imperfect instruments. The author analyses why he called the result wrong. Apart from polling, considers Decision Markets (described). Most people are bad at computation, and that includes experts for whom their intuition often conflicts with statistical principles. Decisions are mainly emotion-based. There is also herd bias and 'group thinking'. Implications for researchers include: taking responsibility for how users understand the data, being honest about limits of accuracy, getting more involved with clients in how the data are used for decisions., being careful to allow for cognitive biases, express the differences between predicted outcomes in terms which make sense to the user, work out expected values, accept payment for predictive accuracy.

Mark Earls

Ogilvy Group London

## INTRODUCTION

This paper was inspired by my own rather embarrassing failure to predict the outcome of the US Presidential correctly, despite my professional and personal interest in the subject. In particular, it focuses on how I laid aside all I know about the predictive limits of opinion polls, how I ignored the signals from other reliable sources and like other, better-known practitioners 'called it for Kerry'.

However, this is not another political polling post-mortem; nor is it an antipolling or even research diatribe: I hope that my own experience provides a prism for considering some broader issues about predictive research of all sorts. Why do we users expect incredible predictive accuracy from market research data? Will survey methods ever be 100% accurate? What alternatives are emerging that might offer improvement or even challenge our current practices?

But most importantly it asks why I - 'somebody who should know better' as the Nobel Laureate Daniel Kahnemann puts it - should make such a bad call knowing all I did.

Where this paper strikes new territory is that it sees the best chance we have to improve the 'effective' accuracy of market research's predictive capabilities (in absolute terms and in the value they sustain in the eyes of our users) does not lie in some methodological arms race, despite the industry's quite proper thirst for improved performance.

Instead, it is as Valentine (2002) suggests - by turning our knowledge of humanity on ourselves and our colleagues and clients that we will make headway: our users misunderstand our data because of well-documented cognitive biases that they share with us. Cognitive biases are as much responsible for decisions based on our data as the data itself, however good the methodology. Bad decisions are made as much with good data as bad and by experts as much as non-experts. As the American Mathematician John Allen Paulos (2003) puts it:

'I was a victim (of gross deception over my disastrous investment in WorldCom) but the primary victimizer, I'm sorry to say, was not WorldCom management but myself.'

So the central challenge for those who want to improve the effective accuracy of their predictions lies beyond communicating our industry's purpose (Valentine 2002) or our data better (Donaldson and Walter 2001) but in helping the users of our data overcome the cognitive biases that lead to bad decisions.

This is largely uncharted territory but I make a small number of practical suggestions as to how we might begin to do this and invite suggestions from MRS members and other practitioners. As Valentine (2002) and others have suggested, there is a significant opportunity for practitioners in both polling and pre-testing (as well as other arms of the Market Research world) to transform themselves from mere data-providers for better decision making to people who help organizations make better decisions [1].

WERE YOU STILL UP THEN?

Midmorning on a cold grey day in early November.

It is the day after the US Presidential elections one of the most keenly observed in Europe in recent years. A strange sense of disbelief seems to pervade the bleary-eyed morning commuter masses trudging through a damp and cold Canary Wharf.

In a 10th floor office above the piazza, I slump onto my office sofa to face my colleague, David Muir. Both of us are pale and exhausted from too many late nights in recent weeks too many nights watching political TV shows, from reading the latest polls and dissections thereof. As political anoraks, we have lived and breathed the US elections for nearly 6 months.

Despite this feverish interest, the Republican victory seems to be as profound a shock to many of our friends, family and colleagues as it is to us phone calls and emails and corridor conversations this morning open with discussion of the shared shock and despair. Not only of course, at the fact of the outcome - that a much disliked [2] President could be re-elected (though this response perhaps highlights how non-Americans and East Coast residents too have failed to understand modern America). No, the bigger shock seems to be that somehow all the indicators had seemed to be pointing the other way political commentators, the TV debates, the polls, Kerry's history of strong late finishing. At one point last night, before the votes were even being counted in some states, the leading British Pollster [3] went so far as to call the election for the Democratic challenger. However, as the night wore on and the map of the USA failed to turn blue, it became clear that his pronouncement was more than premature it was plain and precisely wrong. The Republican incumbent retained the White House with a 51% share of the vote.

How was it that we (and others like us) chose to ignore the well-documented inaccuracy [4] in polls as the basis for and embrace perhaps what we wanted to see in the data?

What makes this worse is that David and I have both been studying a new and more accurate methodology for predicting election outcomes, the so-called 'Decision Markets'. The best known of these have repeatedly outperformed US opinion polls in predicting elections and one (the IEM) was predicting a 51% Bush victory as early as August of 2004. David canny Scotsman that he is has put a tenner on a spread bet some weeks before. So his pain on that grim November morning is precisely 250 less than mine. How could I have ignored this and other data sources all pointing to small Bush majority? What was I doing?

While these questions were raised by my experience of the US Presidential election, I feel the answer reveals some bigger truths about the accuracy and usefulness of research data in general for example it sheds light on a lot that lies behind the recent IPA/LBS study [5] of advertising pre-testing which has damned the practice so strongly. Could it be that the same issues of mismatch between expected and observed accuracy are operating in advertising pretesting. Perhaps the same self-delusion that polling survey users (like David and I) suffer from is also at play among advertising pre-test users. And perhaps the same solutions apply if we want to improve the 'effective accuracy' of pre-test data to improve the value of advertising pre-testing to clients and other users. To close the gap between expected and observed predictive accuracy.

Because at the moment, many users (particularly in the planning, creative and marketing functions) have come to feel that advertising and product pre-testing is like ritual sacrifice for some primitive tribe more important in the observance than in practice. More Juju than an aid to judgement [6].

If, in other words, the only reason we do pre-testing is (Broadbent 2004) 'selling advertising up the line to the boss' (and not because it predicts what will happen) this is hardly a sound basis for a business or something to take pride in.

So I believe this paper is not about elections and political polling at all. Nor is it about Robert Worcester's forecasting ability. Rather, the questions it asks have much broader ramifications for anyone involved in producing or using surveys to predict future outcomes. And the solutions this paper proposes however tentative have much wider application than election polling. They begin to scope out some new practices for market research professionals as Valentine (2002) so memorably suggested.

## PREDICTION IS HARD (BUT THAT HAS NEVER STOPPED US)

Mankind has always tried to predict the future and along the way developed some pretty weird methodologies for doing so. The Romans not only revisited the Oracles of Ancient Greece but also scoured the flight of birds and their entrails with equal enthusiasm. Renaissance French Kings found astrologists and anagramists equally (un)trustworthy. Even today the movement of celestial bodies is for millions a very popular guide as to what tomorrow has in store (and provides a large source of income for Jonathan Cainer and the exotically named Ms von Struenckel). It seems that in matters of state and our personal lives, we are all desperate to know what happens next and inventive in how we go about working out how to find out.

### A more serious approach to prediction

Of course, tealeaves and divining rods are not taken seriously by most of us for the serious matters of business and government (although I do remember one PA who used to read brand tarots); instead, most of us follow the systematic approach of Western science, with observation and careful measurement. And in doing so we simplify the world.

Just as the perfectly spherical billiard balls moving on a perfectly friction-free surface which I learned about in my O-level Physics are an approximation of the real thing so in business we simplify complicated things in ways that allow us to measure them and then make educated guesses about what might happen next. However, in doing so we only have an approximate account of how things really are and thereby introduce error into our predictions. Of course, this simplification is necessary and useful it produces quite workable descriptions of how things are and allow us to deduce underlying rules from which we can predict more precisely but is not by any means the whole truth. Or the precise truth.

Unfortunately, we often forget that real world is rather less neat than these cleaned-up versions would suggest. Mandelbrot's description of the messiness of the real world (whirls and vortices in moving water) is one example of what our simplification would miss out on were we to describe how water flows. We would do well to remember this.

Predicting human responses is even more difficult

This is particularly so when we consider the unpredictability of the human animal. These curious creatures have mental, instinctive and plain physiological functions all of which makes prediction difficult. They are highly responsive to all of their environment which makes it tricky to call their physiological response to a situation let alone any form of behaviour in advance; of course, after the fact we can see tendencies and patterns and what we might call 'typical behaviour' or 'typical responses' (if we have observed this or other subjects over a long period of time). And of course, the mental capabilities can confound experimenters the 'placebo' effect is just one example of how mind and body interact in medicine and something that medical researchers have constantly to struggle with in their predictions of likely effect of different therapies [7].

We are overconfident about causality

Nisbett (2003) notes that we in the West are also very over confident in identifying causality both in general cognition and in highly specialized fields of study (such as the sciences). Clearly, in a complex world, it can be dangerous and another source of error to oversimplify in this way. Again, medical research offers a myriad of examples of this kind of this kind of thing. The recent (and now discredited) claims that the MMR vaccine 'caused' autism were not just based on bad science (small samples selected in a dubious manner and association) but also oversimplified the situation. It is possible that in some patients, given certain other conditions (e.g. imbalances in stomach flora) might respond to the combined stimulus of such vaccines and other things to affect the way that the stomach lining absorbs certain other chemicals which in turn affect the function of the brain. In certain cases, given all of these other conditions this may be the case but this is a long way from saying that the vaccine 'causes' autism. When we oversimplify the behaviour of individuals (e.g. in terms of buying a particular product or category) and look for the 'cause' of this behaviour (this ad) without the complexity of all the individual's other behaviour and the influences on it, we run the risk of introducing error again.

Predicting stable behaviour vs. singularities

And of course this is all the more important when it comes to predicting the behaviour of large numbers of people about a specific outcome such as the collective behaviour of the US electorate in an upcoming election. Much human behaviour is habitual and thus relatively stable (as Ehrenberg has repeatedly shown) and easily predictable. However, when predicting the outcome of a singularity or some attempt to change stable behaviour, things get much tougher. Consider for example the observation by Worcester et al that roughly 10% of voters make up their mind on the day.

You don't have to agree with Gustave Le Bon (1895) on the mercurial or shifting nature of crowds to recognize that is extremely difficult if not impossible to be absolutely precise both in our descriptions of current and our predictions of future behaviour. (Although as I suggest elsewhere (Earls, 2003) the econophysicists like Helbing and Ormerod are developing really interesting tools for studying and predicting human behaviour by ignoring much of what goes on between our ears by understanding the rules of interaction that govern mass behaviour and applying the statistical tools of particle physics to their models).

#### Scientism, measurement and the myth of perfectibility

Sadly, business hasn't seemed to want recognize any of these difficulties. Much of our toolkit and our ways of thinking about business are rooted in the positivist scientism of the mid 20th Century (Earls, 2002; Tasgal, 2003). We assume that the world is simple, describable and above all measurable if not now, then at some time in the near future it will be possible to explain everything down to the last atom. Computer processing power and the rise of the database seems to support this hope for absolute precision in both measurement and the resulting prediction. Hence the rise of interest in issues like micro-marketing and 6-sigma. In the positivist world, everything is measurable, everything is describable, everything is predictable and everything is perfectible or will be soon enough.

Sadly, this way of thinking has been carried over to other aspects of business and sees echoes the broader culture of the West. We imagine we can perfect ourselves through therapy or surgery, that we could avoid illness or misfortune if only we could identify and measure the causes accordingly and avoid them. We rail at inaccurate weather forecasts and late trains. We imagine that we can measure the size and shape of the UK population with great accuracy (remember the 'missing million men' of the last Census?) and predict the future distribution of the population. We imagine that opinion polls can tell us what people really think with great accuracy or that will one day be able to do so.

Of course, it is important in engineering to be precise an imprecise specification or a badly tooled spare part will reduce the performance of a manufactured product. In mass manufacturing (the origin of much of our business culture) it is essential that measurement and predictability are precise, if the same product is to be made again and again and again to the same high quality standard. Imprecision is not culturally acceptable in such operations of business. And the impossibility of improving the precision of these operations culturally unthinkable, despite the increasing evidence that senior decision-makers and financial experts alike make their decisions more on instinct than calculation.

#### Positive reinforcement of 'scientism'?

Our expectations of precision are also subject to positive reinforcement by our repeated attempts to be 'scientific' in these attempts to understand the world. Particularly science with numbers. As Tasgal (2003) and Valentine (2002) both note, numbers tend to take on a life of their own and become they thing they are supposed to be measuring or predicting, however dubious the science behind them. The more we do it and it works more or less the more we believe in it.

This is why we users expect more accuracy than we have the right to from opinion poll data. Equally why there is such a feeling of letdown when the pollsters get it partially right but precisely wrong (they did get it partially right - after all none of the polls suggested a Kerry landslide or Nader minority win). Just as when a space rocket falls out of the sky or a pharmaceutical wonder drug is revealed to have side-effects. Equally it explains why the findings of the IPA study are so disconcerting. Surely, there must be a correlation between the deployment of science and performance? We expect it because that is what we think science can and does do.

## SURVEYS STRENGTHS AND WEAKNESSES AND THE SEARCH FOR THE ULTIMATE METHODOLOGY

### Opinion polls, pre-tests and scientism

Opinion-based surveys are merely a child of this culture. They are an attempt to simplify the world in order to measure it and in the deployment for predictive purposes are open to all of the challenges above as the best practitioners like Worcester readily acknowledge.

As a practical tool for understanding how large numbers of individuals behave now and might do so in the future, opinion-based research is and will continue to be extremely useful. Both for habitual behaviour and for other kinds and singularities They are relatively quick, easy to understand, relatively cheap, provide usable data but most of all it make sense to the positivist mindset which popularized their use: if we want to find out what people do, let's ask them. And (as Wendy Gordon has pointed out) how convincing the answer can appear when a 'real' consumer gives it, whether we hear it in a viewing facility or read it off a PowerPoint chart! Far more so than the tea-leaves or chickens guts that previous generations have used to divine the future.

Moreover, one of the appeals of survey methodologies can provide a richness of data that simple observational or behavioural data struggles with. If what people think is important the opinions and attitudes they hold about the subject we are interested in then this is clearly a reasonable thing to ask punters about, too. No matter that respondents' memory is a bit faulty and they sometimes conceal their true feelings, we can get around this through finding new and better questioning methods, more insight questions or some as yet unknown analytic tool.

We ask the questions, count up the heads, head off for some over claim or downright lying and - bingo we have the positivist and scientific truth about how things are and how things will be.

### Political polling and the papers

While no doubt some of the election post mortems (as in the US in 2000 and the UK in 1992) will blame the media for misusing political polling data, political polling has long been a servant of the news media and not some pure science. Who else (apart from the parties) would pay for political polling research?

The first recorded use of opinion polling was in Harrisburg PA in 1824 on behalf of the local paper. Since then, statistical and technological advances have enabled the development of much larger surveys with greater accuracy than the Harrisburg original (thanks to the efforts of George Gallup et al) and more and more has been learnt about the methodological pitfalls (sampling, question formulation and analysis tools like weighting). They have become so advanced that no American politician of any substance can do without the expert psephologists at least none would dare. Notwithstanding the rise and rise of qualitative research in politics (as in business), the pollsters are more important than ever in politics.

Let's be clear - polls are not perfect

Curiously, while we know more and more about the weaknesses of opinion surveys and their causes, few of us dare to discuss them [8]. At least, not openly. Perhaps, this is again a function of the culture of scientism we all work within which of us wants to hedge and qualify our findings and recommendations when faced with an audience hungry for scientific? That said, on the US State Department Website, the leading American pollster, Zogby (2004) is open about some of the pitfalls at least.

He cites things like sample size (curiously 1000 is somehow acceptable when 500 is not how does this compare with sample sizes for product and advertising pre-tests?), when and how the sample is drawn (and is honest



about the issue of rapidly falling response rates) and the difficulties in achieving a properly representative sample of such a huge universe of electors. It is interesting to conjecture what Gallup would have made of these technical matters; he regularly achieved samples of 20,000 plus.

Naturally, Zogby (2004) notes the pressure from the organizations commissioning such surveys for speed over accuracy as a major source of error in political polling he cites Florida in 2000 as a good case of the haste of news organizations to be first with results as a driver of methodological imprecision. Frankovic (2001) made much the same point closer to the time. Only in passing does he refer to things that have concerned practitioners in advertising research things like asking the right question or analyzing the responses of the right people (e.g. Hall and Heath's insistence that we get our models of how advertising works right before we choose our measures or approach to analysis of our chosen measures).

### Other challenges

However, Zogby fails to mention some of the more fundamental challenges to survey (and any ask-answer no, 'qualies' can't escape) methodology, which a number of us now believe suggest that there are deeper and inherent weaknesses in opinion-based surveys and ask-answer research approaches generally.

Weaknesses such as the fact that humans are notoriously unreliable narrators of their own lives (therapists and clinicians are still confounded why we read so much into respondents accounts of what they do or might do in the future -see Earls (2001) & (2003). Our ability to recall our own experience is poor (Heath has pointed out a number of times the misleading nature of 'explicit memory' and the key role that 'implicit' memory plays in influencing behaviour) and our opinion is as likely to follow our behaviour as precede it (e.g. Ehrenberg various).

And of course he is silent on the 'herd theory' of human behaviour (Earls 2003). That is, that human behaviour is largely (and unconsciously) influenced by the (perceived and actual) behaviour and perceived opinion of others and not determined by the individual himself or herself. In other words, we do what we do largely because of the influence of others, even in the most secret of locations such as the polling booth. This is perhaps why a number of countries have imposed strict limits on publishing polling results during the course of elections it would be a way for the opinions of others (however inaccurately described) to be seen by lots of individuals.

By asking lots of individuals how they would vote or whether they would be more inclined to buy the advertised brand and then aggregating their responses up, we are missing out the key mechanic of human behaviour other people. It is possible to compensate for this to a certain extent through a smart selection of measures which simulate this (e.g. Reichheld's (2003) net recommendation scale considers whether advocacy will be generated) or more radically through Godes's (2003) online methodology (which suggests that the volume and distribution of word of mouth in chat rooms and on bulletin boards is a good indicator of the likely future audience for new TV shows). But these are still largely only theoretically better approaches and as yet unproven.

Whatever your particular view of survey methodologies whether you are a vendor or a user, whether you buy the herd-theory (even for some situations) or not it is difficult to avoid concluding that the whole approach is riddled with methodological issues which lead to inaccuracy and this is what leads us to constantly under perform in the precision of our predictions. Under perform against the unfairly high expectations of those commissioning and using the data, that is.

### Surprisingly good polling

In fact, it is a credit to the skill and expertise of the vendors and practitioners that surveys are able to be as precise as they are and repeatedly so. Remember the US polls were mostly right but precisely wrong. And for

the users of data this precision is important, both too all of us in this culture of scientism and particularly to those in business, where important decisions are made on the basis of survey data. It is easy to call clear electoral contests correctly but closer ones are harder to call. Equally, it is relatively easy to distinguish a great ad from the stinkers; most effective advertising is somewhere in between. Which is where Broadbent's (2004) criticism of advertising pre-testing is focused: bad or imprecise research leads to bad decisions. And that is costly to both advertisers (who commission the pretests) and to the agencies (whose work is neutered or rejected).

But it is because we operate in this same culture that has created the gap between expected precision and actual precision that we within the industry are engaged in a continuous 'methodological arms race': we believe in the measurability of everything and the perfectibility of survey methods. We strive to find the ultimate question, the ultimate pre-test approach, the ultimate sampling or analytic tool. The ultimate methodology.

Of course, it is good that we do continue to try to improve, but the quest for the ultimate methodology is an illusion, both in political polling and advertising pre-testing, because survey research has any number of inherent weaknesses. And as we see, the real opportunity lies elsewhere.

## ALTERNATIVES TO POLLING

Back to the morning of November 3rd.

So I know the weakness and relative inaccuracy of even the best polling data (particularly exit polls which caused such a stir on election day itself). Like most expert users, I wasn't by any means a one-club golfer.

As you would have expected, I have studied other reports based on a number of different approaches from high journalism to academic studies. One that sticks in my memory was by James E. Campbell (2004) of the University of Buffalo which used a thorough analytics-led historical model to based on the three

'...fundamentals that are in place before campaigns begin the public's opinion about the in-party, the general state of the election-year economy and incumbency'.

Campbell comments on the volatility of opinion-polls particularly during the run-up to an election but insists that having a firm grasp of the 'fundamentals' allows a much clearer long term view. (Incidentally, he correctly calls the election narrowly for Bush. Darn!).

Of course, I chose to ignore the weight and breadth of this kind of analytic approach to prediction preferring the stories of the Kerry's late finishing and the late surge in the polls to the considered and multi-discipline approach I would have recommended myself, were I advising a client.

But there was another more accurate source of predictions, which I somehow screened out literally to my own cost. That is the so-called 'Decision Markets' a by-product of experimental economics originally designed to study investor behaviour. What distinguishes this approach from most practice in market research is that rather than ask those whose behaviour we are trying to predict, Decision Markets take a step back and ask outsiders (often non-participants) to predict the outcome. And to do so competitively through the trading of assets linked to a specific outcome. A bit like betting.

## Bets, Bookies and the Power of Markets

Betting is a public competition of cognition an attempt to guess a verifiable fact - albeit rather more difficult than Galton's guess the weight of a pig competition described by Surowiecki (2004). More difficult in that the cognition is about a future outcome rather than about some measurable quality of a tangible object.



The competition has the basic challenge of who can best predict what will happen? Which horse will win? Who will score first? What will the final score be? In gambling markets, we bet against each other (but also against the bookie who sets the prices and takes a margin if he gets the prices right) to determine whose prediction of a particular outcome is most accurate.

Unfortunately few individuals can be really good at these acts of cognition at any one time, no matter what our specific expertise or inside knowledge we have. As a result most individual bets are inaccurate predictions of the outcome. More importantly, almost none of us can sustain it over time (this is why bookies tend to be better off than their punters!).

Yet somehow the net-effect of our competitive action seems to produce strikingly good predictions. Surowiecki (2003) cites Hoerl and Fallin (1974) who compared betting market predictions and actual results in 312 horse races at one US track in 1970 and found a remarkable correlation between what the market predicted and the outcome. Other authors (e.g. Pennock et al 2001) make much use of the 'game' metaphor. The literature is full of examples such as these perhaps we should not be surprised by the obsession with sports betting that so many academic economists seem to share. However much time they have spent at racecourses or bookmakers, it appears to have helped the development of economists' understanding of how markets work to predict outcomes.

How can the collective be smarter than the smartest individual?

For most of us, it is counterintuitive that we can be smarter collectively than the smartest individual in any given group and to be so consistently. Surely, one individual (the 'expert') must be better than others? Maybe a small number of experts' or 'better guessers' affect the market's guess?

Not so, suggests the literature. The power of the collective to guess better than the best individual lies in a curious feature of markets (and one overlooked until recently). Markets are not just good at aggregating prices but information, particularly the tacit knowledge and information that is not easily articulated by anyone individual. Markets are able to take advantage of all the implicit knowledge that exists in and between the participants in the market. This is the engine of their predictive capability.

Note that this is not because, as Le Bon or Freud [9] or HG Wells (who envisaged a 'World Brain' some 80 years before the internet) might put it - because the crowd have some kind of 'collective brain' or Collective consciousness a number of writers in the field also fall into this slip of the pen. No, a collective brain is just as illusory as the sense that individual punters at a bookies or racetrack have of feeling lucky. Or that wearing their lucky pants or any other superstition might impact on their acts of prediction. There are no supernatural powers at work here. There is no Lady Luck. Nor is there anything like a 'collective mind'; it just seems like there is.

The Iowa Electronic Markets

So how do Decision Markets actually work to predict elections (or any other kind of tangible future outcome)?

The best-known ones - the Iowa Electronic Markets [10] - are an online market facility which allows a wide range of individuals to trade assets whose ultimate value depends on the outcome of some future event (e.g. an election but other assets are also traded). They are real-money futures markets (although other markets are just as efficient using virtual money). They operate 24 hours a day, using a continuous double-auction trading mechanism. Traders invest their own funds up to \$500 per trader but typically much less and conduct their own research and make their own trades. Election markets have ranged in size from a few dozen to more than 500. So just like a bookies, then? Apart from the flat caps.

The striking thing the literature suggests about these markets is their ability to predict election outcomes with extreme accuracy the university claims a prediction error of only 1.37 percentage points across the 47 election markets run before this Autumn.

More specifically, the markets have been shown to be very good at short-run prediction of election outcomes (see Berg et al, 1997) and are considerably more accurate than polls (Berg et al, 2003) in long run forecasting both across elections and across long periods of time preceding elections (i.e. not just for eve of election forecasts).

This much I knew before a vote was cast in the 2004 US elections. I'd even written a presentation about the markets and their amazing capabilities to forecast all kinds of things (including sales, resource requirements and drug licences) with greater precision than traditional research tools. I'd suggested they might even challenge our dependence on expensive pre-tests for advertising and product testing [11]. David had researched the software needed to run them. We'd both checked IEM prices on a weekly basis and more often during the televised debates (wherever I found myself in Europe). I'd followed the Hollywood Stock Exchange and its success in predicting 35 of the 40 Oscar nominations for 2004. And ruined dinner parties and bored friends about that, too.

So why was it I ignored them? And what does that tell us about the way to improve the 'effective accuracy' of traditional survey research how to overcome the inherent methodological weaknesses discussed above?

'PEOPLE THEY AIN'T NO GOOD (AT SUMS)'

Is this just a case of innumeracy?

One possibility we must consider particularly with quantitative data sources - is that my misjudgement is down to something as simple as my own (in)numeracy. (OK, so I only have two maths O-levels!). People are generally not very good at mental computations and very often make very obvious mistakes.

For example, if I tell you that Jesus Christ was betrayed by 8.5% of his disciples, this generates a different impression than suggesting that it was just one disciple that did it. Numbers often hide more than they reveal: if I tell you that 3 cups of coffee a day increases the likelihood of you having a heart attack by 50% this seems a big risk (most of us will avoid Starbucks for a while). In order to make sense of the risk we need to put the percentage into perspective by understanding the real numbers they represent. I.e. by first identifying the real number of people (say in your age band) who we would expect to have heart attacks each year (let's say this is 20 per 10,000 head of population). Clearly this is still a rare event for people of your age. Increasing the 20 by 50% still represents a tiny albeit ever so slightly larger number than we would expect for the population as a whole.

Equally it is well documented that we are pretty poor at basic probability calculations most people assume that if (as is the case as I write) the England cricket captain Michael has only won 7 of the 23 tosses he has been part of, somehow the probability of his winning the next toss is something other than 50%. I'm sure that when he returns to the dressing room next time he loses, his colleagues will think so. Is this kind of error what lies behind my false prediction of the Bush-Kerry contest?

A number of contemporary commentators have drawn attention to the widespread innumeracy in modern society, particularly in the UK and USA. Lotteries for example have been described as a 'tax on the innumerate'. Mathematicians (e.g. Paulos, 1988) and employers organizations are as one in decrying the levels of numeracy of those entering the workforce in many cases the educational system is blamed and counter-measures such as the compulsory mathematics foundation courses for those entering tertiary education are tested. This is kind of

thing that education ministers and professional educational committees like getting to grips with improving performance levels of this kind of skill is tangible, measurable and apparently easy to do.

Whether or not we can improve levels of numeracy in the users of our data over the longer term, trying to do so may be an illusory goal. It seems that the central issue in developed markets is not levels of numeracy but rather (as everywhere else) the way we humans are designed.

Experts are as bad as civilians in computation

I found it reassuring to discover how well documented it is that experts and the highly numerate repeatedly make the same mistakes as non-experts and those with less skill in statistics. For example Tversky and Kahnemann (1971) detail studies amongst psychologists and psychology students that show that experts are victims of repeated calculation errors:

'Remarkably, the intuitive judgments of these experts did not conform to statistical principles with which they were thoroughly familiar. In particular, their intuitive statistical inferences and their estimates of statistical power showed a striking lack of sensitivity to the effects of sample size. We were impressed by the persistence of discrepancies between statistical intuition and statistical knowledge which we observed both in ourselves and in our colleagues. We were also impressed by the fact that significant research decisions, such as the choice of sample size for an experiment, are routinely guided by the flawed intuitions of people who know better.' [12]

The literature of decision-making is full of such examples. Gladwell (2005) builds heavily on the work of Oskamp, which showed how more information actually impedes the diagnosis of psychologists to argue for intuition and gut feel in decision-making.

Medical professionals are aware of the problems in getting patients to make good decisions about treatment but McNeill et al (1982) describe how medical practitioners are just as prone to miscalculation as patients and their family. By framing the probabilities of the outcomes of two different therapies (surgery and radiation) in different terms (i.e. by using the emotive terms of 'mortality' and 'survival'), both professionals and patients miscalculated and chose what we might otherwise have expected the 'rational' decision-maker to reject. Thanks to people like Kahnemann and Tversky, the old model of the rational decision-maker has been abandoned in academic economics at least. In the area of applied Social Psychology, it is now widely accepted that decision-making is rarely rational in any recognizable sense of the word.

Thinking for a moment about our own business: it is striking how firmly we cling to the assumption of rationality in the participants in our business ourselves, our clients - and the nature of the dialogue we have with them. We like to think that we present objective and exhaustive data without bias to enable objective and bias-free decisions. We like to pretend that we are scientists, free from human error. But this like so much surrounding our culture's scientism is sadly an illusion.

Thinking about my own errors again, perhaps it is not innumeracy that led people like me and other users of polling data astray. All of us expert or civilian are just as bad. It appears that we do not think with the calculators in our head at all, but our 'hearts'. With our emotions first and foremost.

Emotional decision making

You would be right to think that this is not a new insight. In the history of mankind (both real and imagined) many have highlighted the truth of this in their descriptions of human decision making, from Cicero in *de Oratore*:

'For men decide more problems by hate or love or lust or rage or sorrow or joy or hope or fear or illusion or some other inward emotion than by rationality or authority or any legal standard or judicial precedent or statute.'

To Mr Spock of Starfleet command (talking to McCoy)

'Desires and passions pervert human judgments, and thus people make errors to which more logical life-forms are not prone'

But the lack of novelty does not mean that the insight is unimportant. Recent MRS conference papers have made much of neuroscience in order to better understand for example how advertising works on consumers or how to measure its effect on them (e.g. Heath various). And at the heart of these papers is this insight that we are emotional rather than rational decision makers. Hence the repeated citations of Damasio's Phineas Gage. And it is true that these papers are beginning to make their contribution to what we call the methodological arms race they show the ability to improving our ability to measure and predict performance (e.g. Heath, 2003).

So much is easy to accept of consumers ('civilians') but its truth (and many of the implications) are equally if not more important when it comes to considering the users of the data we produce. We are more Kirk than Spock. We too are prone to 'errors to which more logical life-forms are not prone'.

Emotional Decision-Making, Heuristics and errors of judgement

It is generally accepted in the literature [13] that rationality (and especially computation) only plays a secondary role in our cognition and decision-making. Kahnemann and Tversky's 2 system model is one such account based on experimental data. (See [Figure 1](#))

Instead, in what Kahnemann (2002) calls 'natural' cognition - we respond emotionally to observed situations with 'heuristics' (that is cognitive rules of thumb which are either learned or hardwired see for example Nicholson, 1999). And we only rarely engage the more considered thinking skills of System 2. And even then System 2 thinking is largely governed by the inputs of the emotional nature of System 1.

The curious feature of these rules of thumb is that their value seems to be in the cognition they permit rather than their own accuracy: they enable us an enormous amount of practical reasoning in a very efficient manner in terms of our mental processing power. However, they are not necessarily precise or indeed consistent or exclusive (they are often imprecise and an individual can and does use several conflicting heuristics).

This account explains a number of the cognitive biases observed in the literature. The 'laziness' of our cognition is best illustrated by what is called 'accessibility bias'. For example, few of us will immediately recognize the numbers 1776 and 37x48 represent the same amount. The former is much more 'accessible' than the latter. Equally the way a set of blocks (Higgins, 1996; cited in Kahnemann, 2002) is laid out to make computation easier or harder will lead us to greater or lesser errors in estimate of the height of a tower that can be built. Also the way we frame a situation in our description of it can change the way it is understood as the surgery vs. chemotherapy example cited above suggests.

We tend to be better at spotting change (again, Nicholson, 1999) has a good explanation of this) rather than things that stay the same (is this why we respond so strongly to minute changes in sales or tracking data and resist Ehrenberg's call to reduce the rounding of sales or share data to at most 1 decimal point?).

One of the shared features of many of these heuristics is similarity (we might articulate the question our brains ask of a situation or phenomenon as 'what is this similar to that I have already seen?'). This leads to a bias

known as 'substitution' we see things through the lens of things we already know. For example, the language of American Politics is riddled with 'another Watergate' or 'the next Kennedy'.

We tend to see what our heuristics tell us we should see and ignore that which they do not (known as Confirmation Bias). And we are more concerned about and fight harder for what we might lose rather than what we might gain Paulos' (2003) account of his self-deluding investment in WorldCom stock is a classic narrative of how we throw good money after bad in our attempts to save something we already have.

Group biases

And, of course, all of these biases are amplified by our imitative social nature. 'Do what the others are doing' seems to be a very strong heuristic and leads to small groups being particularly open to bias Milgrim's (1992) famous street corner sky watching experiment is a useful indicator of the power of this (debrief audiences take note!).

This is particularly the case in small groups - 'group think' (mentioned in both the Butler Report and the Senate committee report into the build up to the Iraq War). Surowiecki (2003) describes the phenomenon in his description of how the Columbia Shuttle Mission Management Team failed to investigate or act on the damage caused to the vehicle at launch by loose foam fragments because the committee chairman pronounced.

'And I really don't think there is much we can do so it's not really a factor during the flight because there is not much we do about it'

Equally, in the area of investor behaviour the 'herding' that is caused by this kind of cognitive mechanism is likely to lead to impoverishment rather than riches. Paulos (2003) is clear about why the 'irrational exuberance' which Alan Greenspan first described arises:

'Humans are a social species, which means we're all connected to each other, some in more ways than others. This is especially so in financial matters. Every investor responds not only to relatively objective economic considerations, but also in varying degrees to the pronouncement of national and world leaders (not least of those Mr Greenspan), consumer confidence, analysts ratings (bah), general and business media reports and their associated spin, investment newsletters, the behaviour of funds and large institutions, the sentiments of friends colleagues and of course the much derided brother-in-law.'

Robert Menschel (2002), a Senior Director of the Goldman Sachs Group, warns against it as an investment strategy

'the tension between the individual and the crowd predates all of us and will continue long after we are all gone, but the individuals who resist the pressure of their panicky peers the ones who act on their internal judgements, not on the mores of the mob will be the ones who succeed in the end, always'

Conclusion: bias, damn bias

In conclusion then, the way our (emotionally-wired) brains work means that each of us is open to a number of biases all of the time and often without us realizing that this is the case. We will only use the purely rational and computational capabilities when we have to and even then we will do so in the light of the emotional elements of our cognition.

This explains this is why mathematically skilled experts make such dramatic misinterpretations of data and make

such errors of calculation. People like me. And you. The people we present our data to. Our colleagues and clients.

## WON'T GET FOOLED AGAIN?

So maybe it is time for me to submit my own errors to this analysis. Any or all of the following could well provide the answers to my embarrassing call for Kerry:

- I willingly ignored my own expert knowledge (and previous experience) because I wanted Kerry to win (or rather Bush to lose)
- I didn't do as much computation as I should have preferring to follow patterns of change rather than absolutes. I allowed my emotional interest to lead my computation. In particular, the volatility of the opinion poll findings seemed to attract my attention more strongly than the very stable IEM prediction of the outcome (bor-ing!). And I ignored my statistical training and failed to consider the error in poll data to any great extent (this got worse as the election got closer).
- I suspect I was guilty of substitution bias as instead of answering the question 'who will win in November?' I repeatedly answered a different question ('who do you want to win/lose in November?')
- I was a victim of my own confirmational bias I saw the signs I wanted to see and ignored those I didn't.
- Despite my knowledge that Kerry started behind and would have to really perform brilliantly to have any chance, I suspect I was 'anchored' in my memory of Clinton's late surge (a successful 'Comeback Kid' from a previous race)
- And I suspect that I was also subject to a kind of group think most Europeans (certainly the ones I meet) seemed to want anyone but Bush to win. Did we talk ourselves into a mass delusion? Perhaps.

What do I conclude from understanding my own sources of error in calling the Bush-Kerry contest? Another example of how even the supposed experts are misled not by even the best of data sources but by themselves. As our colleagues and clients can be. And are. On a regular basis. And without realizing it.

## SUMMARY SO FAR AND THE WAY FORWARD

We have established that we and the users of our data have unreasonably high expectations of the accuracy of survey data. Largely, this comes from the culture within which we operate with its roots in mid 20th Century positivism. In addition, our failure to admit or discuss the possibility of inaccuracy has reinforced these expectations. (See [Figure 2](#))

Also, that while survey approaches provide good and practical data for our users, as predictive tools they are inherently imperfect and imperfectible: they are based on approximations of a highly complex creature, have a number of sources of statistical error (such as sampling etc) but are based on the unreliable testimony about the respondents own lives.

Other sources seem to offer greater theoretical and practical accuracy in predicting human behaviour in particular, the Decision Market approach seems to offer enormous scope in prediction of a range of activities for which survey approaches are currently both an input and often the key tool (such as product testing or elections).

However, these too are open to the cognitive biases to which all humans experts by no means excluded are



subject and which radically reduce the effective accuracy of the data itself. (Yes, Bob that means you and me).

So where do we go from here?

Clearly, a continuation of the 'methodological arms race' the quest to constantly improve survey research methodologies is a good thing. Better questions, better data collection, better analysis all these things will continue to improve the predictive quality of market research data. However, the threat of Decision Markets notwithstanding, the bigger opportunity lies not in the data or the methodologies or merely as Donaldson (2001) suggests in better ways of presenting our data

'You still need to produce top quality consumer insight, based on robust techniques. But you need to broadcast the findings in a way that engages, informs and excites the audience in a way that drives the whole organization to actually do something with the knowledge that you have spent so much time and money getting'

It is my belief that the greatest opportunity for us and the greatest need is for us to take responsibility for how our data is understood and used (rather than merely transmitted). In doing so, we must embrace the biases that our users unwittingly bring to the data the biases that distort and devalue what we have worked so hard to bring to them.

Compare and contrast this excerpt from the Advance Programme for 2002 Conference (cited in Valentine 2002):

'The papers at Research 2002 will concentrate on the need for researchers to provide the insight and the analysis, and how, by using innovative methodologies and tools, researchers may persuasively communicate the information and relate it to client issues thereby firmly establishing the MR function at the end of the businessThe programme aims to demonstrate that MR is an integral part of and not just a supplier for the whole marketing process'

What follows are some initial suggestions as to how practitioners might move beyond the methodological arms race and the transmission of information and data. This is by no means an exhaustive list; nor is the suggestions by all means realistic or even palatable to all current practitioners. It would certainly benefit from consideration and shaping by practitioners. That said, taken as a whole, it suggests a different model of engagement a different role for Market Research practitioners in general but particularly suitable for those whose data is used to predict the future.

## SUGGESTED ACTIONS

### Take responsibility

The first and most perhaps most important step is to decide to take responsibility for how our users understand our data. This will be difficult and may well involve new skills (see below) and a different model of engagement. It is not essential. NB it will still be possible to be a vendor of data pure and simple but for those vendors who want to become trusted advisors of clients (and not the juju men of some forgotten ritual as the Broadbent camp might have it) this provides a real opportunity.

Particularly so, as the other marketing disciplines are finding it harder and harder to sustain this advisory role. Essentially, such a change would mean transforming research agencies and planners from mere data providers to organizations who help clients make smarter decisions (either on their own or in collaboration with other key partners such as ad agencies or other consultants).

## Honesty about accuracy

Key to this is an unashamed honesty about the limits of accuracy possible with different data sources. Maybe even to own the agenda about accuracy and the impossibility of achieving perfect data (given the discussion above). This also seems something that the MRS could lead the marketing profession as a whole to engage with. Perhaps under the overall banner of 'smarter decision-making' (to encompass both technical and other issues discussed below).

Another approach (not incompatible) is to focus on how we present uncertainty. I suspect that the kind of health warnings used by the early qual agencies [14] would be unpopular with many vendors but would certainly something that would provide tangible support Valentine's 'significant acts' - for a campaign for smarter decision-making. I am sure that the account planning community would almost certainly be supportive of such a move in relevant sectors.

Equally, an MRS quality standard for prediction would be of great use in certain sectors if we are so confident in our predictive skills, let's open them up to auditing by a 3rd party, funded perhaps by levy of vendors.

On a more practical note, it is clear that users neither understand the sources of error in survey research nor are they able to properly use the statistical measures such as confidence limits or the implications for the quoted number of such limits. One way to address this might be to start to develop ways of expressing the range of numbers suggested by the confidence level rather than just including '+4.5%' in a tiny footnote (particularly if as suggested below we can translate this into terms the users understand the value of). However, we choose to do this, in the days of PowerPoint presentations, it must be possible to find a way of represent the range of possible scores on any particular measure to illustrate this.

## A new engagement model

It is clear that the engagement model would have to change, too if we are to embrace the cognitive biases identified above: vendor: user interactions should no longer be about transmitting data from one party to another but a means to make smarter decisions (using the data provided, of course). It is curious that the debrief has remained unchanged for so long (from overheads to laptops is not a long way); perhaps we could learn to let go of these 'data' transmission sessions. Clearly, there will be a need for retraining if not recruitment. A different skill set would be needed to accomplish this, but that may in turn provide some fresh blood for the MR world.

## Knowing and avoiding the bias in everyday practice

Engaging successfully in this way will also mean that great care will be required in all interactions with users to avoid the elicitation of cognitive biases.

It is all too easy to frame a research report or debrief with the value judgements about the subject in hand. While we all like to hear good news, the opposite is true of bad news and announcing it as such at the beginning of a debrief or report is not good practice. How often does the 'Columbia Shuttle' effect take hold of a debrief today?

Equally, many standard statistical practices like comparing to norms or 52 week highs would seem to be good practice but can all too often become strong and unhelpful anchors. (What is wrong after all with an average ad? If the effectiveness of ads follows some kind of normal distribution, then an average ad would be like most ad, averagely effective). Equally, as discussed above, the numbers can often hide more than they reveal.

Key to overcoming cognitive biases would be the ability to facilitate useful discussions and debates, to elicit the agendas and biases of those in the room and help them avoid the small group biases discussed above. One

suggestion is that in what we now know as 'debrief' meetings the researcher should actively elicit both the agendas/expectations from the audience of decision makers prior to showing data and also check for understanding during and after the presentation. In some ways, this might be seen as an evolution of Mike Hall's Framework model which is based on first working out how users think an ad will work and then evaluating first and foremost against those dimensions relevant to that model. Again not something many practitioners will feel comfortable about today but essential to move on.

Predictive or not?

Finally, it is time for vendors to get off the fence about the predictive nature of their data this is a particular issue for pre-test vendors, for political pollsters are. As Broadbent notes (2004), too many assume predictive powers without being willing to be held accountable for them.

Either our research work is predictive or it is not; if it is, then why not make it easier for the users to understand the difference between different outcomes by expressing the predicted outcome in terms, which make sense to the user in sales or economic value rather than awareness. As Miller and Muir (2004) argue, part of the reason for suspicion of marketing professionals is that we do not talk the same language as business. Why do we not make the move towards them?

Compare expected values not AI scores

A particular means to do this would also encapsulate the troublesome element of probability the financial notion of Expected Value. That is, instead of taking just the predicted outcome to assign probabilities to the possible outcomes and provide a sum of these together with an indication of the variance [15]. Not only would this provide a better basis for decision-making but would also signal our intent to move closer to the rest of the business - many financial professionals are used to dealing with this kind of estimate.

Payment by predictive accuracy

And to signal this predictive conviction what about the suggestion (first raised at the APG sponsored debate at a recent MRS conference) of payment by accuracy? This surely would be the ultimate proof of our predictive powers. Sadly, I suspect that this is the least likely of my suggestions to be taken forward particularly by ad pre-test vendors, but probably the most important. If we really have the ability to make practical predictions about which ads are going to be effective, then this should be part of the proof to clients.

Acknowledgments

Thanks for suggestions and thoughts to Roddy Glen, David Muir, Adam Lilley, Roger Goodchild of Nick Southgate, Gerry Earls, Peter Wells of Nilewide and my long-suffering PA Linda Witt.

References

Baskin M (2003) How to Use Research to Develop Campaign Ideas Admap April 2003 Issue 438.

Berg JE, Forsythe R and Rietz TA (1997) What Makes Markets Predict Well? Evidence from the Iowa Electronic Markets in W Albers W Guth P Hammerstein B Moldovanu and E van Damme (eds) (1997) Understanding strategic interaction: Essays in honour of Reinhard Selten New York Springer 444-463.

Berg JE, Forsythe R, Nelson F, Rietz TA (2003) Results from a dozen years of Elections futures markets. Online at University of Iowa <http://www.biz.uiowa.edu/iem>.

Berg JE Nelson F and TA Rietz (2003) Accuracy and Forecast Standard Error of Prediction Markets Working Draft Online at University of Iowa <http://www.biz.uiowa.edu/iem>.

Broadbent T (2004) Why pretesting is obsolete Admap October Issue 454

Burden S (2003) Ad Creatives' Relationship with Ad Research Admap April 2003 Issue 438.

Campbell JE 2004 The fundamentals in US Presidential Elections: Public Opinion, The Economy and the Incumbency in the 2004 Presidential election, presented at the Elections, Public Opinion and Parties (EPOP) conference at the University of Oxford, September 2004.

Donaldson S and Walter P (2001) Seeing is believing. Consumers do not live in reports and charts so why is consumer insight so often presented that way MRS Conference papers

Earls M (2001) The Death of Marketing: what now for market research? MRS Conference papers: (2002) Welcome to the Creative Age J Wiley and Son London; (2003) Advertising to the herd: how understanding our true nature challenges the ways we think about advertising and market research MRS Conference papers.

Ehrenberg ASC (1997) How do consumers come to buy a brand? Admap March 20-24. (1998) Repeat Buying Charles Griffin and Company.

Frankovic K (2001) Fixing the US Election Polling Problems Means Fixing Election Problems Newsletter of World Association of Public Opinion Research Q.1 2001.

Freud S (1921) Massenpsychologie und Ich-Analyse Leipzig Vienna and Zurich, Internationaler Psychoanalytischer Verlag included in Freud S (2004) Mass Psychology and Other Writings Penguin Modern Classics tr. JA Underwood.

Gladwell (2005) Blink: the power of thinking without thinking Allen Lane Godes D with Mayzlin D (2003) Using online conversations to measure word of mouth communication online <http://dor.hbs.edu>.

Hall M(1992) Using advertising frameworks Admap March 1992 Hanson R(2003) Shall We Vote on Values, But Bet on Beliefs Online <http://hanson.gmu.edu>.

Heath R (2002) The Hidden Power of Advertising how low involvement processing influences the way we choose brands Admap Monograph no 7 (2003) Measuring the hidden power of emotive advertising MRS Conference Papers.

Higgins ET (1996) Knowledge activation: Accessibility, applicability, and salience D in Higgins ET and Kruglanski A (eds) Social Psychology: Handbook of basic principles New York Guilford Press.

Hoerl AE and Fallin HK (1974) Reliability of Subjective Evaluations in a High

Incentive Situation Journal of the Royal Statistical Society 137 227-30.

Kahnemann, Daniel (2002) Maps of Bounded Rationality: A Perspective on Intuitive Judgement and Choice Nobel Prize lecture online at [www.nobelprize.org/economics/laureates/2002/kahnemann-lecture.pdf](http://www.nobelprize.org/economics/laureates/2002/kahnemann-lecture.pdf).

Kittlitz K (2003 )Experiences with the Foresight Exchange at [www.extropy.online](http://www.extropy.online).

Kull S, Miller D et al (2004) Global Public Opinion on the US Presidential Election and US Foreign Policy Program on International Policy Attitudes, Centre on Policy Attitudes and the Centre for International and

Security Studies at the University of Maryland, Online.

Kurkov A (2002) *Zakon uliki Folio Kharkov* published in UK (2004) by Harvill Press transl. George Bird as *Penguin Lost*.

Le Bon G (1895) *The Madness of Crowds* London.

Matthews R (2002) Ask the oracle *New Scientist* No. 2335 22nd March 2002

McNeill BJ , Pauker SG, Cox HC and Tversky A (1982) On the elicitation of preferences for alternative therapies *New England Journal of Medicine*, 306 1259-1262.

R Menschel (2002) *Mobs and Mayhem: a modern look at the madness of crowds* J Wiley and Sons.

Milgrim S (1992) *The Individual in a Social World* NY McGraw Hill

Miller J Muir D (2004) *The Business of Brands* J Wiley and Son.

N Nicholson (1998) How Hardwired is Human Behaviour? *HBR* July/August.

Nisbett R (2003) *The Geography of Thought* The Free Press NY.

Paulos JA (1988) *Innumeracy: Mathematical Illiteracy and its Consequences* Basic Books NY. (2003) *A Mathematician Plays the Market* Basic Books NY.

Pennock DM , Lawrence S, Nielsen FA Giles CL (2001) Extracting Collective Probabilistic Forecasts from Web Games Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001) pp. 174-183 ACM NY.

Reichheld FF (2003) The One Number You Need to Grow *HBR* Dec 2003.

Surowiecki J (2004) *The Wisdom of Crowds* Little Brown.

Tasgal A (2003) *The Science of Brands* MRS Conference papers.

Tversky A & Kahnemann D (1971) Belief in the law of small numbers *Psychological Bulletin* 76 105-110.

Wardle J (2003) Ad Development: Using Qualitative Research *Admap* April 2003 Issue 438.

Worcester R (2005) Correspondence with Author plus a wide range of other publications.

Valentine V (2002) Repositioning Research: a new MR language model MRS Conference papers.

Zogby J (2004) The Polls, the Pundits and the Elections of 2004 on US State Department Website <http://usinfo.state.gov>.

---

[1] NB Rather than discuss particular vendors and their methodologies or applications of survey-based skills to specific marketing needs (e.g. advertising pre-testing), we have chosen to use the example of the recent US Presidential elections and the author's own cognitive biases as a focus for our discussion. This allows us to raise important questions for all practitioners and users without hopefully giving the impression that we are somehow anti-survey research or specific methodologies (such as political opinion polling). This is also the reason why we do not reference much of the excellent work by Worcester et al on the specific application of survey based research to politics. Equally, while the paper concerns itself with quantitative surveys, many of the

conclusions are applicable to qualitative practice and the suggestions of leading qualitative practitioners' thinking (Wardle, 2003; Burden, 2003; Baskin, 2003).

[2] For example see Kull et al (2004).

[3] It should be noted that this was a late reversal of a 4-month call for Bush. The reversal however late was within a hairs-breadth of being correct. If only 68,000 people who voted for Bush had instead voted for Kerry, the White House would now be under Democrat control.

[4] Expert practitioners such as Worcester (2005) have 'consistently argued that opinion polls don't, won't, cannot and never will 'predict' anything' but this seems to have fallen on deaf ears among users.

[5] See Broadbent (2004) for a challenging discussion of the practice of pretesting advertising which arises from this study.

[6] In Andrey Kurkov's 'Penguin Lost' (2002) little boys in black and white become essential fixtures at Russian mafia funerals but no-one seems to know why. They are surreal and unrecognized echoes of Misha the pet penguin who gives his name to Kurkov's satirical novels. For a while, everyone has to have them.

[7] Interestingly, when Western researchers first tried to conduct drug trials in China, the local counterparts referred to the placebo effect (such a bad thing to western researchers as it can mask or mislead about the actual pharmaceutical effect of a drug under test) as 'the right kind of medicine' you don't actually have to give the patient any medicine and they improve.

[8] Worcester aside.

[9] See for example the one text which deals explicitly with Mass Psychology Massenpsychologie und Ich-Analyse (1921).

[10] From a range of sources including Surowiecki (2003), Berg et al (1997 & 2003), Kittlitz (2003) and Matthews (2002) etc <http://www.biz.uiowa.edu/iem>.

[11] Decision Markets are a potential threat to MR businesses While I promised earlier in this paper not to be negative about polling or other survey methods, I am certain that the Decision Market approach provides a significant threat to large parts of the MR business they are cheap (the software is a one-off cost and administration can be cheaply and easily outsourced), they are able to deal with a wide range of predictive challenges (See R Matthews (2002), R Hanson (2003)) from sales forecasting to investment options in pharmaceutical development and they are more accurate than survey-based models. But perhaps the most appealing aspect to business decision-makers is that they derive their efficacy from the way they harness the knowledge that already exists in the client company.

By this I am not suggesting that Decision Markets will wipe out the survey business (indeed opinion-polls seem to be one of the information sources that Decision market traders use in trading) or that they will do so over night, but they do offer significant advantages in prediction over survey based approaches and thus an unusual source of external challenge to our industry. Perhaps some of the larger players in our industry will find ways of embracing this methodology to block new entrants perhaps they are already experimenting with an advertising or product decisionmarket product! Whatever one's view of the likely impact of Decision Markets on the survey business, it is undoubtedly the case that they are a spur to change in the survey business. Particularly over the slightly embarrassing gap between expected and actual precision of the more traditional tools for decision-making.



[12] Kahnemann (2002).

[13] See Kahnemann (2002) for an extensive review of the literature.

[14] See Earls (2003) 'These findings are not predictive and are our own interpretations based on a limited number of interviews with a (small) sample of probably unreliable individuals who may or may not be representative of the population at large and whose interaction may or may not reflect how other networks of individuals may react in the future' (intended for qualitative debriefs).

[15] Paulos (2003) tells the story of trying to decide whether to park in a car park or on the street. The car park will cost you \$10 or \$14 depending on whether you stay for an hour. Let's assume that it is 3 times more likely that you stay for less than an hour than you stay longer. This means the probability of each event is 75% and 25% respectively. You could alternatively take your chance of parking illegally on the street and have reason to believe (say from previous experience) that 1 time in 5 you get a ticket for \$30, 1 in 20 you get towed away (this will cost you \$100) but 3 out of 4 times you will get away with it.

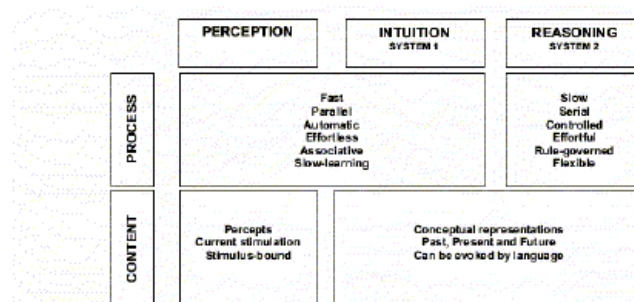
Which should you do?

The Expected Value of an outcome is calculated by multiply each potential outcome by the probability and then summing all the possible outcomes together. So for the parking lot outcome this is  $(10 \times .75) + (14 \times .25)$  or \$13. By contrast, the Expected Value to you of the street option is less \$11  $((30 \times .20) + (100 \times .05) + (0 \times .75))$ .

Now this seems a better bet (\$11 is less than \$13). However, before you chose this option remember that 1/4 of the time you're going to be in for between \$30 and \$100. How good does parking on the street look now?

## NOTES & EXHIBITS

FIGURE 1



Source: Kahnemann (2002)

FIGURE 2

Finding	Comments
High expectations of accuracy	- "Scientism" - Positively reinforced by us
Surveys inherently imperfect	- technical (e.g. sampling) - methodological (ask-answer/herd)
New challenges – more accurate	Decision-markets/Econophysics
All data misunderstood	Cognitive biases in users - both good and bad methodologies

**Source: Kahnemann (2002)**

---

© Copyright Market Research Society 2005

Market Research Society

15 Northburgh Street, London, UK, EC1V 0JR

Tel: +44 (0)20 7490 4911, Fax: +44 (0)20 7490 0608

[www.warc.com](http://www.warc.com)

All rights reserved including database rights. This electronic file is for the personal use of authorised users based at the subscribing company's office location. It may not be reproduced, posted on intranets, extranets or the internet, e-mailed, archived or shared electronically either within the purchaser's organisation or externally without express written permission from Warc.

**WARC**  
AN ASCENTIAL COMPANY

