



Statistics and the art of legitimation:

The strange case of 'n Ach'

Adrian C. Brock

University College Dublin



Adrian C. Brock

Abstract

This article examines the role of significance testing in psychology, using the well-known work of David C. McClelland on achievement motivation as a case study. McClelland amassed some impressive statistics in support of his views, though they appear much less impressive when it is understood how they were achieved. A number of methodological decisions had to be made along the way and there appear to have been a series of fortunate coincidences whereby the methodological decision that McClelland made always led to significant results, whereas other decisions that he might have made and which might seem more reasonable, would not have led to significant results. His statistics seem to have been little more than a way of conferring scientific legitimacy on theories to which he was already committed. Using this and other examples, it is suggested that this is typical of how psychology proceeds. Significance testing is better understood as a social practice whose main function is to confer scientific legitimacy on pre-existing beliefs.

Persuasion and legitimation

All human beings who are not hermits engage in the business of persuading others from time to time, whether it be persuading one's employer of the justice of a salary increase or persuading one's children to clean their rooms. Some are involved in persuasion on a more professional scale. This might include politicians, journalists, lawyers, priests and advertisers. We might also add academics, including scientists, to the list. All of us intuitively know that some arguments are more likely to persuade a certain audience than others. Quoting from the Bible may impress a group of Christians but it is less likely to impress a group of atheists. Similarly, quoting Margaret Thatcher may impress a group of conservatives but it is unlikely to do the same with a group of liberals. The authority that is likely to work differs from case to case.

Anyone with more than a passing acquaintance of psychology will be aware that one of, if not *the*, most important forms of legitimation in the field is statistics; and, in particular, a certain type of statistics known as the significance test. This is not true of every aspect of psychology, of course, but it applies to most branches of the subject. The basic idea is centred on departure from chance. Imagine a multiple choice test in which four alternative answers are presented. A student who has not studied for the exam and is answering the questions randomly would be expected to get a score of 25%. The real world is, of course, not so tidy and the actual score might be 35% or even 15%. What if a student has a score of 50%? Can we conclude that he or she has studied for the exam? Unfortunately, we cannot because there is always a random element involved - the aspect that some people call 'luck'. Statistics cannot tell us whether or not the observed result is due to chance. They can only tell us the likelihood (or probability) that it is due to chance.

As many readers will be aware, there is a convention in psychology: if one's results could only have occurred by chance on 1 occasion out of 20, they are deemed 'significant' and one can have confidence in them. If the results could only have occurred by chance on 1 occasion out of 100, they are deemed 'highly significant' and one can have even more confidence in them. 1 in 1,000 is even better and 1 in 10,000 is better still. Many readers will also be aware that there is no logic to this convention. 1 in 12 million may sound very impressive but the odds in many national lotteries are greater than that and yet no one would suggest that the results were due to anything other than chance. This is equally true of 1 in 500 million, 1 in a billion or any other figure we care to choose.

The following work outlines the quest of a famous psychologist for statistical significance and the steps that were taken to achieve it. It also poses some questions about the role

that this kind of evidence plays in the beliefs and practices of psychologists.[69]

The theory and practice of 'n Ach'

Like psychoanalysis, this approach to psychology is closely identified with one person: David C. McClelland, a Harvard psychologist with a strong interest in personality and motivation. This should not, however, blind us to the fact that both were social movements. As Danziger (1997) puts it, McClelland "established a minor industry devoted to 'achievement motivation'" (p. 122). He had many assistants and collaborators and his work inspired others to conduct similar research, both in the United States and overseas.

McClelland was clearly influenced by Henry A. Murray who had developed a complex theory of personality based on 'needs' (Murray, 1938). People were said to differ from each other in terms of the strength of their needs. For example, gregarious people would have a strong need for affiliation, whereas 'loners' would not. Although McClelland was interested in other needs, most of his work was centred on what he called the need for achievement. This was usually abbreviated to 'n Achievement' or 'n Ach'. People were said to differ in the strength of this need. For example, when experimental subjects were faced with a choice between difficult and easy tasks, some would consistently choose the most difficult tasks that they could be expected to complete, while others would choose the tasks involving the least effort. These differences were attributed to differences in 'n Ach'.

Although he tended to use the term, 'achievement' in a very loose way, McClelland believed in later life that the choice of terminology had been unfortunate (McClelland, 1976; Cohen, 1977). Achievement can come in many forms. Painting a great work of art or developing a scientific theory is an achievement; one might also speak of 'military achievements'. McClelland meant the term in a much more narrow sense, economic success in business, which is in itself an interesting socially-inspired definition of the term. Moreover, one might question how success in business is related to the experiments where people choose the difficult tasks. McClelland seems to have assumed that the underlying motive was the same in both situations.

McClelland was not a hereditarian. He regarded different levels of 'n Ach' as the product of differential socialization. For example, some of his work was devoted to studying the interaction between mothers and sons. His rationale for this was that mothers were most likely to be the primary care-givers and sons were most likely to be entrepreneurs. He did not believe that this state of affairs was normal or natural and supported the equality of women on the grounds that women with a moderate level of 'n Ach' were more likely to pass this need on to their children. Apparently, care-givers with high 'n Ach' levels might be too concerned with their own success.

McClelland used this theory to explain economic inequality in the United States. Minorities like African-Americans and Hispanic-Americans were not poor because of discrimination or because they lacked education or capital but because they had never been taught the so-called 'Protestant work ethic' to the same degree as their unhyphenated American counterparts. The theory was also used to explain the inequality of wealth among nations. 'Developed' countries were said to be successful because their peoples were high in 'n Ach', while 'underdeveloped' or 'developing' countries were relatively unsuccessful because their peoples were low in 'n Ach'. The obvious solution to the problem was to devise ways of encouraging high levels of 'n Ach' among ethnic minorities in the United States and among people in the developing world.

McClelland published these views in his most famous work, *The Achieving Society* (1961). One of the reasons for its success was the empirical evidence that he produced. In what was probably his most famous study, McClelland and his colleagues analysed children's readers from 22 different countries in both 1925 and 1950 for their achievement imagery. The idea here was that they would act as a Rorschach test for the country as a whole and each country could then be assigned a score for 'n Ach'. This was another point on which McClelland had been influenced by Henry Murray. In line with the psychoanalytic beliefs of the time, Murray believed that motivation (as expressed through 'needs') was largely unconscious and so one could not simply ask people what their needs were. These had to be inferred indirectly by trained assessors. Murray consequently invented the major alternative to the Rorschach test, the Thematic Apperception Test or TAT. This involved presenting ambiguous images to a person and asking them what was happening in the picture. Since the images were ambiguous, it was thought that people's responses, as in the Rorschach test, would reflect their personal concerns. [70]

It would have been impractical for McClelland to test large numbers of people in 22 different countries, especially given the amount of time that projective testing involves. It was also impossible to test people on dates in the historical past. Children's readers had to suffice. These, according to McClelland, embodied the values that each society wanted to convey to its children. Around 20 stories for each country were used. They were translated into English and rated for 'n Ach' by two American judges who did not know which countries they were from. Having obtained these scores for 'n Ach' in 1925 and 1950, the researchers correlated them with the economic performance of the 22 countries over the period 1925 to 1950. The 'n Ach' scores for 1925 and the economic performance from 1925 to 1950 had a statistically significant correlation of 0.53. The 'n Ach' scores for 1950 and the economic performance from 1925 to 1950 had a statistically insignificant correlation of 0.23. The correlation between the two sets of 'n Ach' scores was also a highly insignificant 0.03. The positive correlation of the economic performance of the 22 countries with the 'n Ach' scores for 1925, but not with the 'n Ach' scores for 1950, led McClelland to the conclusion that high levels of 'n Ach' were the cause of economic growth.

McClelland decided to see if his theories could be applied to societies in the more distant historical past. He and his colleagues examined societies as diverse as ancient Greece, Spain prior to the conquest of the Americas, Britain prior to the Industrial Revolution; even the Inca civilization of South America. In all cases, he found that their rise as great powers was preceded by high levels of 'n Ach' and their subsequent decline was preceded by falling levels of 'n Ach'. His conclusions were therefore historically and cross-culturally valid, or so he claimed.

McClelland had fulfilled the usual criteria for statistical significance in all these studies. It is not easy to gather data on societies that no longer exist but the general pattern of these results, taken together with the contemporary results, was seen as overwhelming evidence in favour of the theory. Many of McClelland's contemporaries, not only from psychology but also from other social sciences, were highly impressed with these results. One contemporary reviewer wrote:

This is an amazing book. It is so broad in scope, so ambitious in objective, so ingenious in technique and so full of correlations and tests of significance that few will escape a feeling of awe as they set out to read it. (Erasmus, 1962; p. 622)

As I have noted elsewhere (Brock, 2007, June), the background to this work is the Cold War and the attempts by the western powers to use development aid as a tool to keep the less developed countries on their side. As the Cold War began to intensify, so did the discussion surrounding 'development' and how it might be achieved. Both were at their height during the Kennedy Administration from 1961 to 1963. A consequence of this situation is that economics took on an importance that it had not enjoyed before. For example, Kennedy's ambassador to 'the world's largest democracy', India, was the famous economist, John Kenneth Galbraith. There was almost unlimited funding for economists who specialised in development and how it might be achieved.

McClelland's work, appearing as it did in 1961, was clearly an attempt to jump on this bandwagon. However, the situation should not be seen in personal terms. The theory of 'n Ach' opened up the prospect of psychologists in general having something important to say about one of the burning issues of the day and, not coincidentally, a share of the vast resources that were being channelled in this direction. The Society for the Psychological Study of Social Issues (SPSSI) devoted a special issue of its journal to economic development in 1963. McClelland contributed an article to the special issue on China (McClelland, 1963). Attempts were also made to popularise these ideas. In 1962, the American Psychological Association sponsored a 30-minute educational film, "The Need to Achieve", which helped to make McClelland's work better known outside academic circles.

It thus came to the attention of politicians and funding agencies. The US Agency for International Development (AID) funded a training course at the Harvard Business School in the summer of 1963. The participants were from India, Egypt, Thailand, Colombia, Mexico, Algeria, Tunisia, Chile, Italy and Spain. The last two were still regarded as 'undeveloped' at the time, which no doubt reinforced McClelland's views on the importance of the 'Protestant work ethic'. He was also invited to give expert testimony at a US congressional hearing in 1968 (Herman, 1996). Financial support for McClelland's training schemes was also provided by organisations like the Ford Foundation and the Carnegie Foundation [71]

One of the schemes involved developing high levels of motivation among entrepreneurs in India. A variety of techniques was used. They included getting the businessmen to set goals and to keep a daily record of how close they were to reaching those goals. The businessmen were also encouraged to stay in touch with each other so that they could act as a support network. It was a kind of 'psychotherapy for the poor'. Predictably, McClelland argued that the scheme was a huge success (McClelland & Winter, 1969). 'Success' in this

case was defined in terms of the success of the individual entrepreneurs and not their countries, which remained as economically disadvantaged as they had always been.

By the 1970's, McClelland was one of the most famous psychologists in the world. He is included - together with B. F. Skinner, Hans Eysenck, Noam Chomsky, Leon Festinger, R. D. Laing, Neal Miller etc. - in a volume of interviews with famous psychologists from 1977 (Cohen, 1977). In 1987, he received the American Psychological Association's Award for Distinguished Scientific Contributions. One only needs to search the internet to see how seriously his theories are still taken, particularly in business circles. It is therefore important to emphasise that the present work is not dealing with some marginalised dissident but with a Harvard professor who was one of the most famous psychologists of his time and who was supported and honoured by the institutions of his field.

Conceptual and ideological issues

The theory and practice of 'n Ach' has been criticised on many grounds. Only a brief flavour of the criticism can be given here. One source of criticism is the paucity of evidence on which the theory is based. People have a need to eat, defecate, breathe and sleep. We have a wealth of physiological evidence that will corroborate those needs. No such evidence exists for purely psychological needs like 'self actualisation' or 'n Ach'. The only evidence that exists is the behavioural phenomena that they purport to explain. It is like appealing to the appearance of presents on Christmas day as evidence for the existence of Santa Claus. This lack of supporting evidence may explain why psychologists have postulated a bewildering variety of instincts, drives and needs that bear little or no relationship to each other in name or number (Peters, 1958; see also Danziger, 1997). It may also explain why the psychology of motivation has fallen out of fashion.

Yet another problem involves the type of explanation that is used. If we pose the question, 'Why does Fred give part of his income to the church?' We might appeal to some personal goal (e.g. he wants to go to heaven), we might appeal to some cause that could be external or internal (e.g. he was brought up in a religious home), we might appeal to a personality characteristic (e.g. he is a generous person) or even his membership of a social group (e.g. he is a Mormon). As the philosopher, R. S. Peters pointed out in his classic work, *The Concept of Motivation* (1958), questions about the 'why?' of human action are extremely complex. It is not as if McClelland, as well as other 'needs' theorists like Abraham Maslow, assessed the relative contribution of all the possible internal and external explanations and then decided which one or ones were the most important. They were committed to a certain type of explanation from the start; namely, a psychological explanation that was framed in terms of internal impulses (Danziger, 1997).

Such issues are not merely the result of abstract ruminations by some clever philosopher. They matter in important ways. Consider the issue: 'why are some countries rich and others poor?' We can go through similar list of possibilities. Some writers might appeal to the legacy of colonialism, the stagnant or even falling prices of typical 'third world commodities like cocoa, coffee, sugar, cotton and bananas, unequal terms of trade where rich countries have access to the markets of poor countries but not vice-versa, the huge debt burden that many poor countries carry and so on (Ellwood, 2006). These are all social, political or economic explanations. We might also offer psychological explanations, such as the people in rich countries being high 'IQ' and the people in poor countries being low in 'IQ' (Lynn & Vanhagen, 2002), or the people in rich countries being high in 'n Ach' and the people in poor countries being low in 'n Ach'. In other words, the people in poorer countries are either stupid or lazy, depending on one's theory of choice. Again, there is no attempt to look at all the possible explanations for this situation and to assess the relative contributions of each one. Psychologists who start out with psychological theories and then look for possible applications for them will inevitably end up with psychological explanations. [72]

This phenomenon, which is often characterised as the 'psychologization' of social problems (e.g. Herman, 1996), is seen as ideologically pernicious by many writers. McClelland's theories are not a million miles away from the old colonialist view that the 'natives' are poor because they lack the appropriate work ethic. This seems difficult to reconcile with the existence of sweatshops, child labour and, more generally, working hours and conditions that were made illegal in most western countries in the nineteenth century. It is hardly surprising that many people regard McClelland's views as arrogant and ethnocentric.

What may seem ideologically pernicious to some has an inherent attraction for others. In placing the responsibility for poverty and for overcoming poverty on the shoulders of the poor, the theory does not question the division of wealth; neither domestically nor on a global scale. It also does not look at the ways in which the rich might profit from the existence of the poor and thus have a vested interest in keeping them poor. It also holds out the prospect that the poor can be rich at some indeterminate point in the future, thus making them more docile and accepting of their fate. Some help must be provided by the rich but at little cost to themselves. Providing a course of 'psychotherapy' is always cheaper and easier than addressing an inequitable distribution of wealth.

These considerations will cut no ice with many psychologists, who are generally unimpressed by the arguments of philosophers, much less by those of ideologues who might object to 'scientific evidence' because it runs counter to their political views. What they are impressed by is statistics and, in particular, tests of statistical significance. It was McClelland's statistics that were the most impressive aspects of his work. It is therefore important to look in detail at how his significant results were obtained.

How McClelland obtained his significant results

1. Measuring 'n Ach'

What exactly were the measures of 'n Ach' that McClelland and his colleagues used? Anyone who goes into the details will discover that it was a very messy business indeed. In assessing levels of 'n Ach' in different countries, some of us may be tempted to identify each country with a 'culture' but that would be a mistake. One of the countries that McClelland looked at, India, has 23 languages that are officially recognised for government business, while South Africa has 11. Both countries have many more languages that are not officially recognised and most of these languages are associated with a distinct culture. All that McClelland could do in these cases was pick a sample of stories from a sample of the different linguistic and cultural groups.

Other problems include the fact that children's readers in some countries had not changed between 1925 and 1950. Perhaps the most astonishing admission that McClelland made was that all the children's readers for Japan in 1950 had been written by Americans while the country was under American military rule in the aftermath of World War II (Cohen, 1977). This did not prevent him from taking them as representative of 'Japanese culture'.

There also seems to have been a great deal of subjectivity involved in the judgements of achievement imagery. Only two judges were used - both American - and even they could only come up with similar figures "after some practice" (McClelland, 1961; p. 74). Whether or not the figures they produced are anything other than random or arbitrary is open to debate.

McClelland was well aware that his measures of 'n Ach' needed to be independently validated. This is true of all psychological tests. IQ scores are usually correlated with performance on exams and personality scores are usually correlated with the results of other methods of personality assessment. In an attempt to validate his results, McClelland got students in 6 of the 22 countries to write essays and had the essays independently scored for achievement imagery. The results bore little or no relationship to the scores that had been given to the children's readers for the same countries in 1950, namely the most recent figures that he had. McClelland tried to explain these results away, arguing among other things that they were from a limited sample of the population: male, young, upper middle-class and from particular localities. One wonders how he might have reacted had the results correlated with his children's reader scores. He concluded: "in the end, the proof of the pudding is in the eating: do they [n Ach scores] enable us to predict which countries will develop more rapidly economically?" (1961, p. 79). With this statement, McClelland effectively abandoned any attempt to independently validate his measures of 'n Ach' and defined them as 'anything that correlates with economic growth'. This was, however, only the first step in the quest for the holy grail of statistical significance. [73]

2. Measuring 'economic development'

If McClelland's 'n Ach' scores were problematic, the same is true of his measures of 'economic development'. Economists have a standard measure for assessing the performance of a country's economy. It is Gross Domestic Product or GDP and represents the value of the goods and services that a country produces. This value is usually expressed in US dollars and can be divided by the population of the country to produce 'per capita GDP'. While it may not be perfect as a measuring tool, it is considered by economists to be the

best measure available.

McClelland decided, however, that he was not going to use per capita GDP as a measure of economic development, arguing that expressing a figure in US dollars did not take into account the cost of living in each country. Generally speaking, the cost of living in poor countries is lower than the cost of living in rich countries and so it is misleading to compare their wealth in this way. The criticism is perfectly valid.

It is one thing to criticise the standard measure of economic development; it is quite another to come up with something better. McClelland thought he had something better and that was electricity production. There are, of course, parts of the world that are still without electricity and they would have been more extensive at the time that McClelland was writing. It is also intuitively plausible that the production of electricity in a country is loosely related to its industrial development. In spite of this, using electricity production as a measure of economic development brings with it a new set of problems and this might explain why economists have preferred to stick with GDP.

The amount of electricity that a country produces is related to the cost of producing electricity in that country. For example, countries like Canada and Norway can produce hydro-electric power very cheaply. Some of it is exported to other countries and some of it is used in industries like aluminium production which require very high temperatures. The cost of producing electricity will also influence the price of electricity and determine whether it is used for heating or to power trains in a specific country, as opposed to alternatives like oil, coal and gas.

All this may help to explain why, according to McClelland's figures, Canada (1855 kilowatt hours per capita) was nearly twice as 'developed' as the USA (962 kilowatt hours per capita) in 1929, while Norway (2850 kilowatt hours per capita) was more than ten times as 'developed' as Great Britain (269 kilowatt hours per capita) in the same year. In spite of this, both Canada (+2265) and Norway (+2583) continued to 'develop' at much higher rates than the United States (+1598) and Great Britain (+846) from 1929 to 1950. This is, of course, nonsense. All that the figures show is that Canada and Norway produce a lot of electric power. McClelland was aware of these and other problems with his figures but he chose to ignore them.

The unusual and controversial decision to use electricity production, rather than GDP, as a measure of economic development may start to make sense when we realise that McClelland's statistics would have been nowhere near as impressive if he had used the more conventional measure. He was keen to point out that 78% of the countries that were above average in 'n Ach' in 1925 had above average economic growth from 1925 to 1950, whereas only 25% of the countries that had below average 'n Ach' had above average economic growth over the same period. It was this that formed the basis of his claim that 'n Ach' had been the cause of economic growth. However, when GDP is used as a measure of economic performance, only 44% of the countries that had above average 'n Ach' in 1925 had above average economic growth from 1925 to 1950, while 57% of the countries that had below average 'n Ach' in 1925 had above average economic growth over the same period. These figures suggest, if anything, that low 'n Ach' is a more accurate predictor of economic growth than high 'n Ach'.

In fact, the overall correlation of 0.53 between 'n Ach' in 1925 and 'economic development' from 1925 to 1950 would have been a statistically insignificant 0.25 had McClelland used GDP as a measure of economic development instead of electricity production. [74]

McClelland's choice of electricity production as a measure of economic development is only one of a series of fortunate 'coincidences' that have led some writers to openly question whether his methodological decisions were made before or after the data had been analysed (Lewis, 1998; Schatz, 1971).

3. Measuring 'economic growth'

An even more striking example of this phenomenon is the way in which McClelland defined 'economic growth'. This is a straightforward matter as far as most economists are concerned: economic development is measured in terms of GDP and economic growth is measured by percentage increase in the GDP. McClelland rejected not only GDP as a measure of economic development; he also rejected percentage increase as a measure of economic growth. While it has to be said that there was some merit to his criticism of GDP as an accurate measure of economic development, even if the alternative that he used was much worse, there is no merit whatsoever to his criticism of percentage increase as a measure of economic growth.

First of all, he argued that countries with a low rate of economic development tend to have higher growth rates than countries that are already developed: "the lower the initial level, the more likely it is that the percentage increase will be larger" (1961, p.88). Historically, this is untrue. On the whole, developed countries have had much higher rates of economic growth than what are euphemistically called, 'developing countries'. Secondly, McClelland maintained that growth figures for countries with small units of currency would be lower than those with large units of currency: "One cannot help wondering how much of the fact that Japan has twice as high a rate of growth as France is due to the fact that Japanese income is expressed in yen, and French income in much more numerous francs" (1961; p. 88). This is like arguing that percentage increases for US economic growth can be lowered by expressing the figures in cents rather than dollars. The argument is nonsensical and looks suspiciously like an act of desperation.

Instead of using percentage growth, McClelland used a complicated regression analysis that predicted rates of economic growth based on the country's economic development at the starting point. If the actual growth was higher than the predicted growth, it would be classed as a high rate of growth, whereas if the actual growth was lower than the predicted growth, it would be classed as a low rate of growth. The obvious question here is: 'Who decided what the predicted growth should be in each case?' McClelland of course. An interesting feature of this regression analysis is that it predicts that developed countries will have a higher rate of growth than developing ones. For example, if the Bulgarian economy grew by 40% over the period from 1925 to 1950, it would be classed as a high rate of growth, whereas if the US economy grew by 40% over the same period, it would be classed as a low rate of growth.

By now the reader will not be surprised to learn that there is yet another fortunate coincidence at work: had McClelland used percentage increase as a measure of economic growth, as opposed to this complicated regression analysis, 10 of the 22 countries in his sample would have had above average economic growth from 1925 to 1950. Of these 10 countries, 5 had above average levels of 'n Ach' and the other 5 had below average levels of 'n Ach' in 1925. It may be recalled that McClelland argued: "in the end, the proof of the pudding is in the eating; do they [n Ach scores] enable us to predict which countries will develop more rapidly economically?" (1961, p. 79). Had he had used the traditional measure of economic growth, he would have been forced to conclude that his 'n Ach' scores had no predictive value at all. He could have got the same results by flipping a coin.

4. Choosing the 'sample'

There is at least one other fortunate coincidence in McClelland's results. Another methodological decision was to exclude tropical countries from the study on the grounds that they had uniformly slow rates of growth. It is not difficult to imagine that these uniformly slow rates of growth would have prevented him from obtaining statistically significant results. The strongest evidence for the theory is, therefore, based on a sample of less than 9% of the countries in the world and it is, by McClelland's own admission, not a representative sample. [75]

5. Choosing the time period

We might also ask: 'Why 1925 to 1950?' The figures have a certain neatness to them, being exact divisions of 100 and it might be easy to conclude that they were chosen at random. Even if they were chosen at random, it does not preclude the possibility that a different time period may also have led to a different set of results. A study by Lewis (1998) suggests that this is almost certainly the case.

It will be recalled that McClelland found a statistically significant correlation between 'n Ach' scores in 1925 and economic performance from 1925 to 1950. At the same time, the correlation for 'n Ach' scores in 1950 and the economic performance from 1925 to 1950 was insignificant. There is no reason why McClelland's 'n Ach' scores for 1950 cannot be correlated with the economic performance for a similar time-period after 1950. If McClelland's theory is correct, there should be a statistically significant correlation here as well. This is what Lewis (1998) did. He took McClelland's 'n Ach' scores for 1950 and correlated them with the economic performance of the same 22 countries from 1950 to 1977.

There was, however, one crucial difference between his study and that of McClelland.

It will be recalled that McClelland objected to the use of GDP, expressed as a figure in US dollars, on the grounds that it did not take into account the cost of living in each country. By the time that Lewis conducted his study, economists working for the World Bank had produced GDP figures that had been adjusted to take this factor into account. Given that McClelland's objection to the use of GDP was no longer valid, Lewis assumed, perhaps naively, that McClelland would have no objections to its use. This was the only difference between the two studies. Although he had grave doubts about the validity of the regression analysis that McClelland had used as a measure of economic growth, Lewis retained it so that the results would be comparable to McClelland's own.

To cut a long story short, the correlation between 'n Ach' scores in 1950 and the economic performance of the 22 countries from 1950 to 1977 was a highly insignificant 0.015. The results were not even in the expected direction; that is, the economic performance of the countries that were below average in 'n Ach' in 1950 was slightly better than those that were above average in 'n Ach', though not to a statistically significant degree.

On the assumption that there may have been errors of measurement in the 'n Ach' scores for 1950 and/or that high levels of 'n Ach' may take some time to work, Lewis correlated the 'n Ach' scores for 1925 with economic performance from 1950 to 1977 and did obtain marginally significant results. Unfortunately, they were significant in the wrong direction; that is, below average 'n Ach' scores could predict above average economic performance to a statistically significant degree.

The plausibility of McClelland's results

As noted above, McClelland was keen to point out not only that his 'n Ach' scores for 1925 correlated with 'economic growth' from 1925 to 1950 to a statistically significant degree but his 'n Ach' scores from 1950 did not correlate with 'economic growth' over the same period. The latter was a very insignificant 0.03. The implication here was that 'n Ach' had been the cause of the economic growth.

The obvious point to be made here is the one that psychologists make to their undergraduate students but often forget about when it comes to their own research: *correlation does not imply causation*. If we were to find an even higher correlation between the consumption of hamburgers and economic performance, it does not automatically follow that increasing the consumption of hamburgers will lead to economic growth. A correlation between 'x' and 'y' should not lead anyone to conclude that 'x' is the cause of 'y', any more than 'y' is the cause of 'x' or that both 'x' and 'y' are caused by a third factor, 'z'. It could also be that the association is nothing more than a random coincidence, as it may well have been in this case. [76]

One point that may be less obvious to psychologists is that McClelland's findings contradict what economists know, or at least think they know, about the economic performance of the countries concerned (Schatz, 1971). The countries that had the highest economic growth in the 1920's and the countries that had the highest economic growth in the 1950's are virtually the same. If 'n Ach' scores had been a genuine predictor of economic growth, we would have expected a reasonable correlation between the 'n Ach' scores for 1925 and the 'n Ach' scores' from 1950, as well as a reasonable correlation between the latter and the economic performance of the 22 countries from 1925 to 1950. There was neither.

Even before Lewis (1998) correlated the 'n Ach' scores for 1950 and the economic performance of the 22 countries from 1950 and 1977, it had been noticed that McClelland's 'n Ach' scores for 1950 were intuitively implausible. The highest scores were given to Turkey, Argentina, India and Lebanon, while relatively wealthy countries like the United Kingdom, the Netherlands, Switzerland and Japan were rated as low in 'n Ach'. McClelland tried to explain these results by saying "many of the backward countries realize their backwardness and are now motivated to close the gap between themselves and the more industrially developed countries" (1961, p. 102). This did not happen. One of the reasons why many people in poorer countries ultimately became disillusioned with 'development' theory is that the highest rates of economic growth since World War II have been made by the developed countries, not the so-called 'developing' ones. As a consequence, the gap between rich and poor countries has been gradually widening (Dube, 1988, Escobar, 1995; Leys, 1996).

The evidence from history

The weakness of McClelland's evidence can be seen even more clearly with his so-called 'historical' studies. According to McClelland: "We have done a lot of history. Generally speaking economic growth occurs when n Achievement levels are high and economic decline occurs when n Achievement levels are falling" (Cohen, 1977; p. 35). These words are correct in a legalistic sense but they are also misleading.

Take, for example, McClelland's study of ancient Greece. Here he could not use children's readers to assign 'n Ach' scores to different historical periods. He had to rely on whatever written material he could find, including poetry and plays. It was also not possible to find figures for GDP or electricity production. McClelland's figures for 'economic development' were based on a book from the 1930's showing the countries and regions where Greek vases had been found in archaeological digs. This was taken to be a sign of the extent of Greek trade in different periods and the extent of this trade was to be taken to be a sign of 'economic development'. If all that were not dubious enough, McClelland's interpretation of his results was very suspect.

Essentially, the results show that 'n Ach' declined continuously from 700 to 200 BC. Against this backdrop, economic development rose sharply from 600 to 400 BC and was followed by an equally sharp decline, making it difficult to account for the rise and fall in trade in terms of 'n Ach' alone. For one thing, there is a gap of 250 years between the high point of 'n Ach' and the high point of trade. Higgins (1968) commented sarcastically that McClelland is unlikely to interest development agencies in policies that require 250 years to take effect. The reader may recall that McClelland was making fine distinctions between 'n Ach' in 1925 and 1950 in the study discussed above. Now high levels of 'n Ach' and 'economic development' that are centuries apart will suffice.

Let us look at McClelland's statement again: "Generally speaking economic growth occurs when n Achievement levels are high and decline occurs when n Achievement levels are falling" (Cohen, 1977; p. 35). The statement is correct only because the word, 'high' is left vague (how high does 'high' have to be?) It is also true that 'trade' declined while 'n Ach' levels were falling. However, it is also true that 'trade' expanded significantly while 'n Ach' levels were falling as well. Interestingly, McClelland made no mention of this point.

Perhaps emboldened by his 'success' with ancient Greece, McClelland decided to look at 'n Ach' levels in the pre-literate society of the Incas. Here there were no written records of any kind on which 'n Ach' scores could be based. Using the results of experiments which suggested that people who are high and low in 'n Ach' produce different kinds of "doodles", McClelland decided to analyse the designs on Inca vases and came to the conclusion that some designs were indicative of high 'n Ach' while others were indicative of low 'n Ach'. He could thus give ratings to different historical periods based on the designs on vases that had been found. There were, of course, no figures for GDP or electricity production. There were not even figures to suggest the extent of Inca trade. McClelland therefore decided to use the rate of public building as a measure of 'economic development' instead [77]

The results show a sharp decline in 'n Ach' from 800 to 200 BC and a more gradual decline from 200 BC to 150 AD. Then there is a dramatic rise from 150 to 600 AD followed by a sharp decline. Meanwhile, the rate of public building rose consistently throughout the whole period, irrespective of the level or the direction of 'n Ach'.

Even these results were seen by McClelland as evidence in support of his theory. However, in order to do this, he had to make a change. He decided that public building was not a good reflection of 'economic development' after all. Successive generations, especially in small valleys, can use the buildings that previous generations have left behind and therefore do not have the same need to build. How this explanation might be related to the finding that public building consistently rose throughout the whole period is not clear. McClelland then noted that the gradual decline in 'n Ach' levels from 200 BC to 150 AD and the much sharper decline after 600 AD (but not the sharp decline from 800 to 200 BC) preceded military conquest of the Incas by outside civilizations. Thus 'economic development' came to be defined not in terms of GDP, electricity production, trade or public building but military failure or success. By moving the goalposts in this way, McClelland was able to conclude:

The data confirm for another culture area, as different from the traditions of Western civilization as it is possible to get, that n Achievement is a key factor in the rise

and fall of economic civilization. (1961, p. 175)

We have now moved full circle from defining 'n Ach' as 'anything that correlates with economic development' to defining 'economic development' as 'anything that correlates with n Ach'. With these definitions, statistical correlations were inevitable. If they did not occur, one could simply discard the definitions of 'n Ach' and/or 'economic development' that were being used and replace them with something else.

Some tips for the budding researcher

McClelland's work provides some useful pointers on how to achieve statistical significance when doing psychological research.

1. Take your chances

Chance unquestioningly plays an important role in the results that psychologists produce. The phenomenon where insignificant results are discarded and only significant results published helps to encourage this situation. If you do 20 experiments, there is a good chance of obtaining at least one significant result and, if that one result is published, its readers are likely to be misled into thinking that the result is not due to chance.

There is no need to do so many experiments, however, since chance is only one factor in producing significant results. The dice can also be loaded in favour of the researcher in a number of ways. Methodological decisions have to be made and there are no hard and fast rules to guide these decisions. Researchers typically make their choices and then try to defend them on logical grounds. However, in the absence of any hard and fast rules, they can, and often will, make methodological decisions that lead to support for their views.

2. Define the terms

One type of methodological decision involves defining the terms. How do we define 'n Ach'? Originally, this was done with experimental results or the results of personality tests. Later, it became scores based on children's readers, scores based on more general literature and ultimately designs on vases. It was not, however, defined in terms of scores from student essays since these failed to produce significant correlations with McClelland's other scores.

With such flexibility in defining 'n Ach', a researcher always has a fighting chance of correlating it with measures of 'economic development'. However, when there is enormous flexibility in defining the economic measures as well, significant correlations are almost guaranteed. How are we going to define something like 'economic development'? Usually, it is expressed in terms of GDP but, as we have seen, there is no reason why a researcher should follow past precedent. The definitions of 'economic development' that we have seen so far so far include electricity production, the size of the trading area, the rate of public building and military failure or success. No one can definitely say that these definitions are correct or incorrect. They are simply methodological decisions that have been made. We might have done things differently but we cannot say that they are 'wrong' in any absolute sense. This situation gives the researcher a great deal of flexibility in deciding which definition(s) to use. [78]

The important point here is that some definitions lead to statistically significant results while others do not. GDP vs. electricity production is an example. Whether or not McClelland chose electricity production as a measure of economic development before or after his data had been analysed cannot be proven one way or the other. McClelland's economicist critics have noted, however, that he had been happy to use GDP as a measure of economic prosperity in a previous study. Here he wrote: "The best measure of the economic prosperity of a country is probably something like income per capita" (McClelland, 1958; p. 534). His book, *The Achieving Society* (1961), opens with a comparison of GDP (not electricity production) in different countries around the world. Nothing is said here about the alleged inadequacy of GDP as a measure of economic prosperity. What does seem clear is that McClelland picked and chose his definitions as the occasion suited him.

There is no need to portray these decisions in cynical terms. A researcher who is convinced of the truth of his theory, as McClelland clearly was, may decide that a particular set of measurements do not reflect the underlying phenomenon after all. Examples of this mentioned so far include the abandoned attempt to validate his 1950 measures of 'n Ach' by questioning students in different countries or his decision to abandon the rate of public building as a measure of economic prosperity in the Inca study. Note also that there is no question about whether McClelland did the latter before or after the data had been analysed. The decision was made after the data had been analysed and no attempt was made to conceal it. This clearly shows that neither McClelland nor his colleagues thought that they were breaking any rules in looking at the consequences of different measures.

This phenomenon of defining terms to suit one's purposes can often be seen in politics. Thus 'unemployed' in many countries does not refer to the number of people who are out of work. It refers to the number of people who are receiving unemployment payments. In order to receive unemployment payments, one must meet a strict set of criteria and anyone who does not meet these criteria is not officially 'unemployed'. Similarly, those who are deemed unfit to work are 'sick' or 'disabled', people who have reached a certain age are 'retired', while others are forced onto training schemes and classed as 'students' for the duration of their course. We all know the ways in which politicians massage the statistics in order to suit themselves and yet many of us have a faith in 'science' that is charmingly naive.

3. Choose the statistical tests

Yet another way in which the dice can be loaded in one's favour involves the choice of statistical tests. Some might think that there is little room for manoeuvre here but there is. As with the definition of terms, there are past precedents but, as we have seen, researchers may decide for a variety of reasons to depart from them. In some cases, past precedents do not exist.

In a classic study of the sociology of science, Mackenzie (1981) analysed a bitter controversy between statisticians in London in the first decade of the 20th century. The controversy was over whether parametric or non-parametric statistics were most appropriate for analysing a particular set of results. The difference centred on whether the underlying population from which the sample had been taken conformed to a particular distribution; in this case, the 'normal curve'. Two things should be noted here: first of all, it seems odd that anyone should get hot under the collar about the choice of a statistical test; secondly, they were dealing with an assumption, not something that could be proved or disproved. As Mackenzie shows, the argument was not really about statistics. It was political. The battle lines were drawn between eugenicists and their opponents and the key to understanding the controversy is that one type of statistical test led to results that supported eugenicist theories and the other type of test led to results that did not. [79]

A similar situation exists with McClelland's decision to use a regression analysis, rather than percentage growth, in calculating 'economic growth'. As with the example given above, there were no hard and fast rules on how to calculate it, merely past precedents that may or may not be followed. As with the case study examined by Mackenzie, the crucial difference between the two measures is that one led to significant results and the other did not. Again, there is no need to portray these decisions in cynical terms. If a statistical test does not lead to significant results, the problem may lie with the results or it may lie with the test. The decision as to which of the two it might be is often a matter of opinion.

4. Choose the samples

McClelland's decision to use 'data' from 22 different countries might sound impressive. However, as noted earlier, it is not a representative sample of countries. McClelland (1961) stated that he had decided to exclude 'tropical countries' from the study because their uniformly slow rate of growth would lead to insignificant results. Although he does not state it explicitly, he may have done previous studies that had included these countries in order to come to this conclusion. It is not the kind of thing that would be immediately obvious to someone on a purely rational basis.

One cannot criticise McClelland's sample on the grounds that it is not representative of all the countries in the world. McClelland would simply reply that it is not meant to be. His conclusions relate only to the countries that are not entirely within the tropical zones. Here he believed that other factors were at work, though he does not say what they might be.

or how they are related to the climate of the country. He simply says that they are outside the scope of his study.

I have seen no evidence to suggest that the dates of McClelland's sample, 1925 and 1950, were chosen after other alternatives had been tried. This should not, however, lead us to rule out the possibility that they were nothing more than a fortunate choice. The study by Lewis (1998) comparing McClelland's 'n Ach' scores for 22 countries in 1950 with their economic performance from 1950 to 1977 would suggest that this is the case.

McClelland's results are reminiscent of a maze in which there is only one path to the exit and a number of dead ends. One has to choose the correct path in every case in order to reach the exit. Similarly, McClelland had to make a series of methodological decisions in order to obtain statistically significant results. Had any one of these methodological decisions been different, statistical significance would not have been achieved. Also, just as people trying to find their way out of a maze rarely choose the correct path at the first attempt, there is plenty of evidence to suggest that McClelland came to a few dead ends before deciding to retrace his steps and explore the possibilities in a different direction.

5. Interpret the results

While 'significance' is usually a precondition for having one's results taken seriously, there is still a great deal of flexibility in how the results can and should be interpreted. With studies relating to IQ, for example, a hereditarian like H. J. Eysenck and an environmentalist like Leon Kamin can look at the same set of results and come to radically different conclusions about them (Eysenck vs. Kamin, 1981). This inherent flexibility enables researchers to interpret results in order to suit their own purposes.

We have seen how McClelland mentioned that 'economic decline' occurred in Greece when 'n Ach' levels were falling but failed to mention that 'economic growth' occurred when 'r Ach' levels were falling as well. This is similar to the politicians who selectively quote the statistics that support their views and quietly ignore the statistics that do not support their views or even cast doubt on them.

The flexibility in the interpretation of results extends as far as deciding which results are to be taken seriously and which are not. We have seen how McClelland dismissed the results from 'achievement imagery' in student essays as being based on an unrepresentative sample. Again, this was done after the fact and no attempt was made to conceal it. Claiming that samples are unrepresentative of a population is a common way of dismissing results since, as the Mackenzie (1981) study shows, no one knows for certain what the underlying population is. That can only be assumed. A less common way of dismissing results is to make allegations of fraud, as was famously done in the case of Cyril Burt. However, the difficulty of proving the allegation, and the seriousness with which it is taken, leads to 'incompetence' or 'bias' being the allegations of choice. [80]

Conclusions

The point was made at the outset that statistics cannot tell us if we are witnessing real effects or if the results are entirely due to chance. This is true irrespective of whether the probability is 1 in 20, 1 in a 100 or 1 in a million. The winner of a lottery jackpot may have had a 1 in 30 million chance of winning but few people would suggest that there was anything other than chance involved. No amount of significance testing will prove that 'n Ach' exists, let alone that it has any measurable effects, since we can never rule out the possibility that McClelland's significant correlations were due to chance. Indeed, it seems like a plausible explanation in this case.

There is an analogous situation in parapsychology which has produced a good deal of evidence for the existence of extra-sensory perception or ESP. A typical experiment here would involve a person selecting one of four cards and then concentrating on that card. A second person in another room has the same set of four cards and then tries to choose the same card. We would expect 25% of the answers to be correct if only chance is operating. As with any other psychology experiment, statistics can be used to calculate the probability of obtaining the observed results by chance. If the results depart significantly from chance, it is seen as evidence for the existence of ESP. In spite of the fact that parapsychologists have reported significant results of a billion to one (Broughton, 1992), few psychologists, or indeed members of the wider public, are willing to accept the existence of ESP.

All this raises the interesting conclusion that statistical evidence is not the crucial factor in determining our beliefs on psychology that it is often made out to be. To return to one of the examples used in the previous section, I am not aware of any hereditarians or environmentalists in the intelligence testing field who have undergone a dramatic conversion in the light of statistical results. They are much more likely to challenge the results than they are to change their views. It must also be clear from the above that McClelland had a missionary zeal about him. He was completely convinced that his theories were correct. Proving them by means of statistical significance tests was merely a formality. If this did not happen, the fault lay with the statistics, not with his theories. This is also the situation in parapsychology when it goes into the detail (Collins & Pinch, 1982).

Tests of statistical significance might be better seen as a social practice that serves to legitimate the authority of psychologists, just as the complicated procedures in a church, a parliament or a courtroom serve to legitimate their authority. Few members of the public understand statistics and even those with statistical expertise do not have the time or the energy to analyse in detail every set of statistics that they encounter. It is therefore much harder to challenge anyone who claims to have irrefutable evidence in support of their views. The fact that these tests can be so easily manipulated may even add to their appeal.

References

- Brock, A. C. (2007, June). The quest for empire: Psychology, economics and the third world. Paper presented at the joint conference of Cheiron (International Society for History of Behavioral and Social Sciences) and ESHHS (European Society for History of the Human Sciences) in Dublin, Ireland .
- Broughton, R. S. (1992). *Parapsychology: The controversial science*. New York: Ballantine.
- Cohen, D. (1977). *Psychologists on psychology*. London: Routledge & Kegan Paul.
- Collins, H. M. & Pinch, T. (1982). *Frames of meaning: The social construction of extraordinary science*. London: Routledge & Kegan Paul.
- Danziger, K. (1997). *Naming the mind: How psychology found its language*. London: Sage. [81]
- Dube, S. C. (1988). *Modernization and development: The search for alternative paradigms*. Tokyo: United Nations University .
- Ellwood, W. (2006). *The no-nonsense guide to globalization* (2nd ed.). Oxford: New Internationalist Publications.
- Erasmus, C. J. (1962). Review of "The Achieving Society". *American Anthropologist*, 64 , 622-625.
- Escobar, A. (1995). *Encountering development: The making and unmaking of the third world*. Princeton, New Jersey: Princeton University Press.
- Eysenck, H. J. vs. Kamin, L. (1981). *Intelligence: The battle for the mind*. London: Macmillan.

Herman, E. (1996). *The romance of American psychology: Political culture in the age of experts*. Berkeley, California: University of California Press.

Higgins, B. (1968). *Economic development: Principles, problems and policies*. London: Constable and Company.

Lewis, J. (1998). Revaluating the effect of N-Ach on economic growth. In M. A. Saigon & J. T. Passé-Smith (Eds.), *Development and underdevelopment: The political economy of global inequality*, 2 nd edition (pp. 187-194). Boulder, Colorado: Lynne Renner.

Leys, C. (1996). *The rise and fall of development theory*. Bloomington, Indiana: University of Indiana Press.

Lynn, R. & Vanhafen, T. (2002). *IQ and the wealth of nations*. Westport, Connecticut: Praeger.

Mackenzie, D. A. (1981). *Statistics in Britain , 1865-1930: The social construction of scientific knowledge*. Edinburgh: Edinburgh University Press.

McClelland, D. C. (1958). The use of measures of human motivation in the study of Society (pp. 518-552). In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, New Jersey: Van Nostrand.

McClelland, D. C. (1961). *The achieving society*. New York: Van Nostrand.

McClelland, D. C. (1963). Motivational patterns in Southeast Asia with special reference to the Chinese case. *Journal of Social Issues* , 19 , 6-19.

McClelland, D. C. (1976). *The achieving society* (2 nd ed.). New York: Irvington .

McClelland, D. C. & Winter, D. G. (1969). *Motivating economic achievement*. New York: The Free Press.

Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.

Peters, R. S. (1958). *The concept of motivation*. London: Routledge & Kegan Paul.

Schatz, S. P. (1971). Achievement and economic growth: A critical appraisal. In P. Kilby (Ed.), *Entrepreneurship and economic growth* (pp. 183-190). New York: Free Press. [82]

Correspondence: Adrian C. Brock / School of Psychology / University College Dublin / Belfield Dublin 4 / Ireland, e-mail: adrian.c.brock@ucd.ie

IF YOU WOULD LIKE TO SEND IN YOUR COMMENTS ON THIS ARTICLE, PLEASE USE THE [SUBMISSION FORM](#).

IN ORDER TO VIEW THE REFLECTIONS ON THIS ARTICLE PLEASE CLICK [HERE](#).