Comparing Corporate Reputations: League Tables, Quotients, Benchmarks, or Case Studies?

Dennis Bromley
Department of Psychology, University of Liverpool, UK

ABSTRACT

Finding valid and practicable methods of assessment and comparison should help to clarify the concept of corporate reputation and contribute to professional applications in this area. Four approaches to assessing and comparing corporate reputations are outlined: league tables, reputation quotients, benchmarks, and case studies. The first three approaches are interrelated, and face conceptual and practical difficulties. The fourth approach outlines two contrasting methods: the case comparison method and the Quasi-Judicial (QJ) case method.

INTRODUCTION

The literature on corporate reputation, and reputation in general, has been growing in recent years (Balmer & Soenen, 1999; Bennett and Gabriel, 2001; Bromley, 1993, 2000a; Fombrun, 1995; Fombrun & van Riel, 1997; Green, 1992; Kay 1993; Mercer, 1996; van Riel et al., 1998; and Whetten, 1997). The present report was stimulated by an article entitled 'The Reputation Quotient' (Fombrun, Gardberg & Sever, 2000), and by articles in recent issues of *Fortune* magazine describing surveys of America's most admired corporations.

Although widely debated and researched, the concept of corporate reputation seems to lack an agreed theoretical basis, and this limits practical applications. What follows is an examination of some of the issues and methods involved in assessing and comparing corporate reputations.

LEAGUE TABLES

The best-known league tables of reputation for industrial and commercial companies in the developed world are those published annually in the USA business magazine *Fortune*. These league tables are based on large samples of respondents of various kinds, mostly executives, directors, and securities analysts, who rate a selection of companies on various attributes relevant to corporate success. These samples can be divided into sub-samples representing different business sectors.

Fombrun et al. (2000) draw attention to existing surveys that provide league tables of reputational attributes. The methodological limitations of these surveys include the following:

— biased sampling frames
— target firms selected by size of revenue
— restriction to publicly traded companies
— collusion because of the sector membership of respondents
— over-representation of senior managers, directors, and financial analysts in samples
— respondents may lack direct experience relevant to some attributes (items in rating scales)
— mainly pen and paper mail surveys.

A survey may not sample the stakeholder groups best informed about the survey items. The particular attributes on which

Corporate Reputation Review, Vol. 5, No. 1, 2002, pp. 35–50

respondents are asked to rate a number of selected companies vary from one survey to another, but tend to converge on the concept of corporate success (Kay, 1993).

Reports describing the surveys do not always make clear exactly what instructions respondents were given or what form the questionnaire or rating scale took. For example, some *Fortune* reports refer to an 11-point rating scale, ranging from 'poor' to 'excellent', but Fombrun et al. also refer to responses such as 'not sure' and 'declined to respond'.

In some *Fortune* reports, it is not entirely clear how the 'rankings' are derived from the 'ratings'. The ratings of a set of items (attributes) appear to be added together and averaged. These averages (scores) are then summed across a sample of respondents rating the same set of companies. The scores are ordered from highest to lowest, providing a rank order of merit for a number of companies, ranging, say, from one to ten, thus omitting companies with scores ranked 11 or greater. In one *Fortune* report, American Express ranks first in its sector with a score of 7.44, while Capital One Financial ranks third with a score of 7.05.

The idea of averaging a set of ranks is problematic, because a rank order is an ordinal scale, not a ratio or interval scale. A legitimate method of deriving a single rank from a set of ranks is to sum the ranks in each set and then rank these sums. Using ranked data in elaborate multivariate statistical analyses is questionable. See Kerlinger (1986) for a concise review of the concept of 'levels of measurement', but see also Michell (1990, 2000) for a severely critical review of psychometrics, including the notion of 'levels of measurement'.

THE CONCEPT OF REPUTATION

Fombrun sees corporate reputation as a collective assessment of a firm's past behavior and outcomes that depicts the firm's ability to render valued results to multiple stakeholders. Corporate reputation thus reflects a firm's relative standing, internally with employees and externally with other stakeholders, in its competitive and institutional environment.

The *Fortune* magazine's surveys are based on large, relatively heterogeneous samples of respondents. This draws our attention to the difference between a 'collection' of people and a 'collective'. A 'collective' is a relatively homogeneous group of people with a degree of common interest in a reputational entity, such as a company or a product or a person. A degree of common interest and social interaction provide a basis for shared impressions, beliefs, and attitudes, such as trust and esteem, and for group action, such as a boycott, a strike, or a rush to buy shares.

Reputations, which are socially shared impressions, are based on 'collectives', not on heterogeneous 'collections' of people (Bromley, 1993, 2000a). Surveys of collections of people can generate data (scores and ranks) that mimic reputational data, but such results should be construed as overall indices of success, as in an election, not as collective representations of the candidates.

Commercial and industrial companies, like political candidates and other reputational entities, have as many reputations as there are distinct social groups (collectives) that take an interest in them. A survey of the *Fortune* kind, dealing with large, well-known companies, and using extremely large, widespread, heterogeneous samples of respondents, could be said to portray a sort of 'meta-reputation' — a fusion of a large collection of personal judgments about a standard set of corporate attributes. Members of large, heterogeneous samples of respondents are likely to differ in their values and beliefs from other members, and may not be involved in collective action with them.

Smaller companies, by contrast, are not widely known. Their corporate reputations depend on the relatively homogeneous social networks of communication and influence of their various groups of stakeholders. These shared impressions are formed naturally through business activities in a context of social, psychological, and economic circumstances.

REPUTATION QUOTIENTS

Fombrun proposes a multi-stakeholder measure of corporate reputation called a 'Reputation Quotient SM' (RQ). The third revision of the RQ consists of 20 items in eight sub-scales. Applied to a large heterogeneous sample, multivariate analysis yields nine factors. Of seven factors selected, five coincide with the sub-scales. Fombrun concludes that reputation is a construct that combines two factors: Emotional Appeal (one sub-scale) and Rational Appeal (five sub-scales). For example, Emotional Appeal includes trust and admiration; Rational Appeal includes references to product quality, leadership, and risk.

A company's overall rating is obtained by averaging its ratings on all the attributes. These average ratings (scores) are then used to rank companies in order of merit from highest to lowest score. As mentioned above, however, one cannot legitimately average ratings because ratings are ordinal scales, not interval or ratio scales. The problem of the distinction between measurement theory and statistical theory is dealt with below in the section headed 'Psychometrics'.

A number of criticisms and questions can be raised about the league table method of assessing reputation. Although the procedures described above make sense in terms of survey methodology, they raise two questions:

1 What sort of reputation, if any, do they represent?

2 How can the results be converted into a reputational quotient or benchmark?

The answer to the first question is that a company's position (rank) in a league table of companies is analogous to a woman's ranking in a beauty competition, with points awarded by a panel for specified attributes — hence the title, 'America's Most Admired Companies'. The type of league table based on the *Fortune* procedure provides, as *Fortune* claims, guidelines for investors and job applicants.

A company's rank in its business sector is likely to be more firmly based and more informative with regard to reputation than is its rank in business generally. This is because raters drawn from a particular business sector are better informed and more involved. A company's rank in its sector is roughly analogous to an election for a limited number of places, based on proportional representation, using non-transferable votes.

Fombrun does not calculate any arithmetical quotients. A 'quotient', by definition, is a ratio between two numbers; for instance, the Intelligence Quotient (IQ) is the ratio of Mental Age (MA) to Chronological Age (CA), multiplied by 100. If, however, companies were given an overall rating on a set of attributes, then those ratings could be used to calculate a ratio between, say, the rating of a selected company and the rating of the average company. For example, an overall rating of 4.8 by a selected company, and a rating of 6.0 by the average company, would give a ratio of (4.8/6.0) x 100, a reputation quotient (RQ) of 80, suggesting that the selected company is well below the average. Several conditions would have to be met to translate this into a percentile value.

This way of deriving a reputation quotient (RQ) is doubtful, to say the least. As already mentioned, the original data are ratings, not interval or ratio scales. The

assertion that 'numbers do not remember where they came from' might be used to justify the subsequent statistical treatment of rankings and ratings. But, of course, people remember where such numbers came from, and doubts have been raised about their validity. Moreover, meaningful RQ comparisons between companies would, like IQ (intelligence quotient), have to be derived from normative data, based on a large representative sample and a normal distribution of scores.

LIMITATIONS OF THE LEAGUE TABLE METHOD

There are at least four limitations to the league table method of measuring reputation:

— reputational attributes may not be operationally defined
— attributes tend to be stated in abstract, general terms imposed by the researchers, leaving scope for raters' personal interpretations and other influences associated with the timing and circumstances of the survey
— attributes are not equally important from either an objective or a subjective point of view
— samples may be too heterogeneous to constitute a genuine 'collective' representation.

The definition of corporate reputation as 'a collective assessment of a company's ability to provide valued outcomes to a representative group of stakeholders' (Fombrun et al., 2000, p. 243) is not altogether satisfactory because it appears to confine assessments to positive (valued) outcomes. Reputation is essentially an ethical evaluation, and must therefore permit the attribution of negative (undesirable) characteristics. Fombrun reports that reversed (negatively phrased) items in a rating scale tended to confuse respondents, and so were deleted from the final revision of the scale.

BENCHMARKS

A benchmark is defined as a surveyor's mark indicating a point of reference for levelling. The concept of a reputational benchmark is metaphorical. It implies setting up a standard (ideal or average) reputation against which other reputations are compared. A benchmarking system for corporate reputation is not the same as a benchmarking system for corporate performance. A corporate reputation comprises the impressions a company makes on a group of people (a collective).

A benchmarking system, using a version of the *Fortune* method, would involve a representative sample of comparable companies in a particular business sector. Well-informed respondents would rate those companies on a set of operationally defined attributes (based on their subjective impressions). These ratings are assumed to be valid indicators of company attributes. In addition, the scores would have to be normalized, that is, made to fit a normal distribution. This could be achieved either through instructions to, and monitoring of, respondents' behavior, and/or through statistical manipulation of the raw data.

League tables based entirely on objective measures of company characteristics and performance are not indices of corporate reputation. They are, however, sources of information that contribute to the subjective impressions that people form and share with others in the collective processes that generate reputations.

The Free-description Method

An alternative approach to constructing a league table of reputation based on a standard rating scale would be to use a free-description method. This enables respondents to list various attributions, based on their personal interests and experiences

(Bromley, 1993, 2000a). Respondents would be drawn from selected interest groups (stakeholders), depending on the purpose of the research. The more frequently mentioned attributions could then be used to construct a league table of merit using a methodology similar to that associated with the *Fortune* surveys.

The free-description method lends itself to benchmarking reputation by reference to the average rating, as with a 'reputation quotient', dealt with above. Regardless of how they are selected, attributions could also be assigned weights according to their frequency of occurrence or estimated importance.

The free-description method tends to generate a wide range of attributions, most of which are shared by relatively small proportions of respondents, or are idiosyncratic (not shared with any other respondent). Differences in the frequency of occurrence of attributions in a free-description exercise could indicate their relative importance, or at least their familiarity or salience at the time.

The number of attributions contributing to a benchmarking exercise would depend on how frequently they were cited by the sample. For example, attributions made by more than 25 per cent of members of a sample, for a given set of companies, could be included in a list of benchmark criteria.

Consider the following hypothetical example:

— 36 members of a stakeholder group provide free descriptions of 11 companies representing a sector of the telecommunications industry. The method elicits nine attributions, each mentioned by ten or more members of the group.
— the members (respondents) then rate each of the 11 companies on each of the nine attributions, on a scale ranging from, say, 0 (Disagree/Poor) through 5 (Average) to 11 (Agree/Excellent). Respondents are asked to distribute their ratings so that their average over the nine attributions is 5.
— a statistical correction is applied if necessary to achieve this result. The nine attributions are weighted in proportion to their frequency of mention. For example, 'Good prospects' might achieve a weight of 1.7 (17 members cite this attribute in their free description); 'New management' might achieve a weight of 1.2 (12 members cite it).
— it would be important to check whether weighted scores had any advantage over unweighted scores. The resulting scores are summed over the nine attributions for each of the 11 companies. These summed scores are used to rank the 11 companies from 1 (highest score) to 11 (lowest score).
— the distribution of scores is normalized to a mean of 5 and a standard deviation of 1, or some other convenient transformation. This procedure does not avoid the problems associated with levels of measurement, but at least it offers an explicit and clear procedure for comparing corporate reputations as 'collective' representations rather than 'collections' of standard ratings.
— the particular merit of this method of calculating RQs would be to indicate how far, relatively speaking, each company deviates from the average for its sector, and from other companies.

RQ = (company score/average score) x 100
This is analogous to the way IQ is interpreted, at least by professional psychologists (see above).

Readers will no doubt wonder if the research effort required to carry out this sort of investigation would yield a sufficient improvement on existing methods of comparing reputations.

PSYCHOMETRICS

The word 'psychometrics' is used to refer to the way numbers are assigned to subjective estimates, such as judgments of a brand's attractiveness or a company's resolve. These are different from objective measures of a company's attributes or performance, such as p/e ratio or number of employees.

The traditional rating procedure is based on two assumptions:

— that the various attributes share a common factor, namely a company's ability to provide valued outcomes, and
— that the attributions (impressions of attributes) are quantifiable, and ratings constitute valid and reliable measures or indicators of these quantities.

Fombrun demonstrates a common factor in the RQ scale. Michell (1990, 2000) and Kline (1998), however, have serious reservations about the way psychometrics is used in the social and behavioral sciences. This section offers brief comments on their reservations.

In order to construct a useful psychometric scale for assessing corporate reputation, the ideal procedure is to identify a representative set of attributes, or sets of attributes within categories, that are relatively independent of each other and yet highly correlated with the overall score or a criterion. It is likely that the usual attributes for corporate reputation, such as those in the *Fortune* surveys, are interrelated, especially for firms within a particular business sector.

A 'true' score is a theoretical concept in psychometrics. With reference to corporate reputation, a true score consists of the aggregate score a company would achieve on all the quantitatively scaled attributes in a universe of attributes relevant to the issue of corporate reputation. Some of these attributes may not refer to corporate success. A test of corporate reputation comprises a random (representative) set of descriptive categories from this universe. For the purposes of measurement, an estimate of a company's 'true' score is the sum of the 'sub-scores' on a set of test items in each of the descriptive categories.

The number of categories required, and the number of items in each category, are issues to be settled by empirical inquiry and practical convenience, within a legitimate psychometric framework. The larger the number of categories, and the larger the number of items within a category, the better, as far as reliability is concerned, assuming the test is within the respondents' performance capacity.

It is important to remember that subjective attributions represent real or supposed attributes. Real attributes have an independent existence. Thus, attributes may or may not be truly represented by a psychometric assessment.

The interest group members who are asked to make attributions about a company should comprise either the entire group or a representative sample of it. The procedure for eliciting the attributions, such as a rating scale, checklist, or free description, should be standardized, as should the instructions to respondents and the method of recording and analysing the data. Such attributions should be univariate, reliable, and valid (measuring what they are supposed to measure) in the sense of yielding a 'true' mean score plus or minus the standard error of the mean.

A company's reputation score(s) make sense only in relation to the scores of other companies with which they are sensibly compared — for example, in terms of size, sector, location, and so on. Psychometric methods of assessment should be internally consistent, reliable over time, discriminating, valid, and replicable. Scores may be standardized, that is, transformed to fit a particular sort of distribution, such as a normal distribution.

Unfortunately, the numbers assigned to corporate attributes in a psychometric exercise seem not to be 'quantitative' in the strict sense of that word (Kline, 1998; Michell, 1990, 2000), even if the numbers are assigned according to rules, as in a typical *Fortune* survey, and have some actuarial value. For example, there are no real 'measurement units' for subjective attributions. The meaning of the numbers depends on understanding how the corporate attributions have been defined and assessed.

The meaning assigned to the results of any subsequent manipulation of the psychometric data, such as by analysis of variance or factor analysis, depends on two unconnected sets of assumptions:

— the measurement assumptions underlying the original data
— the assumptions underlying the subsequent statistical analysis.

There may be no guarantee that these assumptions have been met.

Some corporate attributes appeal to one sort of stakeholder, other attributes appeal to another sort, although the success of a company depends on producing valued outcomes for all stakeholders. The recent growth of interest in coordinating stakeholder relationships reflects this awareness. A solution to the problem would be to devise separate measures of corporate reputation for each stakeholder group. Companies could then be compared across a profile of stakeholder assessments. It should be possible to weight the importance of each stakeholder group and combine their assessments to arrive at a single overall assessment (Keen & Greenall, 1987).

The above suggestions identify but do not remedy the faults or limitations in the psychometric assessment of corporate reputation. The question is whether there are viable alternatives to psychometric assessment.

CASE STUDY METHODS

Case study methods have a long history in business studies, for both research and teaching. Early references include McNair and Hersum (1954) and Towl (1954). McNair and Hersum describe the historical origins of the Harvard case method. This method uses incomplete cases to explore the diversity of, and the uncertainty in, making business decisions. It is a form of learning using simulations, thus avoiding risk.

Coincidentally with interest in corporate reputation, the literature on case study methods has been growing in recent years, after a long period of relative neglect. Interest in the method, of course, extends well beyond business studies (Bromley, 1986, 2000b; Easton, 1992; Fishman, 1999; Gomm et al., 2000; Hamel, 1993; Ragin, 1987; Ragin & Becker, 1992; and Yin, 1984, 1993).

The traditional case method in business studies appears not to be based on an explicit scientific method. Some may regard it as an art. This is not to say that there are no worthwhile case reports, but rather that there appears to be no systematic method of carrying out, reporting, or cataloguing cases. In an article currently under review, Bromley (2000b) offers a philosophy of science and a scientific procedure for the case method in business studies. The following sections draw on this report.

In view of the rate of change in commerce and industry, and the fluidity of corporate reputation, a short, standardized method of investigating, assessing, and comparing corporate reputations is needed. Possible uses of the case method would be to examine in detail the following:

— companies whose attribute ratings deviate markedly from the average, as explained above
— companies where there seems to be a

mismatch between their reputation and objective measures of their performance
— companies dealing with a risk to their corporate reputation or brands.

The Case Comparison Method

Ragin (1987) draws a clear distinction between the statistical approach and the case study approach in social research. The same distinction could also be drawn in the behavioral sciences, as in the study of personality and social psychology, where the experimental and quantitative traditions are even stronger.

For Ragin, the choice is between (1) the method most appropriate to the analysis of quantifiable variables across a relatively large number of cases, and (2) the method most appropriate to the analysis of nominal categories across a relatively small number of cases. There is, in fact, a third method, the Quasi-Judicial (QJ) case method (Bromley, 1986, 2000b). It is dealt with in a later section.

A typical multivariate statistical analysis of, say, a league table of corporate reputations, tries to assess the average effects of several quantitative variables across a large, diverse sample of companies. Using statistical controls to isolate the effects of variables produces average effects across a sample of cases, but may not answer questions about how characteristics combine and interact to produce a particular outcome.

By contrast, the case comparison method, recommended by Ragin (1987), tries to assess which combination of qualitative factors determines how cases (companies) can be classified into types, and which combination of factors determines a particular outcome, such as the success or failure of a business enterprise, a substantial change in corporate reputation, or a relatively high or low level of public esteem.

An essential difference between the two methods is that in the case comparison method the combination of attributes (characteristics) of cases is dealt with as a whole, whereas in a multivariate analysis they are dealt with in a fractional way. Consider the difference between an average score and a score profile. Two or more companies may achieve the same average rating, and the same rank in a league table, even if they differ, possibly substantially, on some or all of the attributes. They would therefore have rather different reputations. For example, consider the scores of two firms on five attributes:

Attribute:	A	B	C	D	E	Mean
Firm 1	4	7	8	2	4	5
Firm 2	8	2	5	7	3	5

The mean scores do not reveal differences in the score profiles, which may be important in classifying cases or in identifying causal processes. The mean can be derived from a profile, but not vice versa.

The choice between using multivariate analysis and the case comparison method naturally depends on a researcher's interests and the way a problem is formulated, which in turn depends on theoretical assumptions, the available evidence, and explanatory concepts.

Ragin summarises features of Boolean algebra and shows how the case comparison method is made more rigorous by applying Boolean procedures. Readers are referred to Ragin (1987) for further details, and to the detailed example below.

Ragin also considers the problem of generalising from a small number of cases. Bromley (1986) notes that it is easy to generalize, even from a single case. The problem is how to test the validity and scope of such generalizations. Case laws (generalizations) are made in judicial inquiries (legal science) without resort to statistical analysis or Boolean algebra! Scientific knowledge is not restricted to findings from quantitative or experimental inquiries.

Case-oriented scientific investigations take account of three features of cases:

— their internal complexity
— the specific context in which they occur
— and, to some extent, their origins or historical background.

In many instances, the emphasis is on the proximal factors affecting an outcome, whereas in others the emphasis is on the longer-term sequence of events. The same effect or outcome may be produced in different ways over time because of historical shifts in cultural and economic circumstances, as with production methods, consumer values, government legislation, or demographic changes.

For Ragin, cases are treated as wholes, not broken up into fractional parts, which are assigned to different variables. Furthermore, causation is the result of a conjunction or combination of particular conditions. Cases are compared according to whether or not they share a particular set of attributes or circumstances.

It is not possible within the scope of a short article to do justice to Ragin's philosophy of science for the case comparison method. For example, he makes important points about the way investigators perceive similarities and differences between cases, about the limitations of case methods and multivariate methods, and about how various approaches can be combined. Ragin shows how to combine the problem of constructing a useful typology with the problem of analysing causal complexity. It is fairly obvious that the causes of success or failure in business, or changes in corporate reputation, can be numerous and interconnected. The example below illustrates how the method can be used to classify organizations and to compare corporate reputations.

The aim of the case comparison method is to examine how different attributes or conditions combine in one type of case as compared with another. The frequency with which the different types of case occur is not relevant to the immediate issue of *how* they differ, although frequency of occurrence may be relevant to other issues. This stands in marked contrast to a multivariate approach to the same problem. A more important consideration is the choice of a diverse range of case characteristics (attributes or conditions) thought to be relevant in the context of the theory that underpins a case-oriented investigation. Factors may combine in different ways to produce a particular outcome, such as the success or failure of businesses, their relative standing in a league table, or their response to damage to their corporate reputation.

Having identified the conditions or attributes thought to be responsible for a specified outcome, the next step is to list all possible combinations of these conditions, and examine how each combination is related to the outcome. In some instances, the relationship will be clear; in other instances, the relationship will be unknown. Where it is not possible to determine a prior specified outcome for a particular case, it may still be possible, on the basis of the results for other cases, to predict the likely outcome. An interesting feature of the case comparison method is that it can reveal matters of interest about cases which do not share the key attributes that affect an outcome in a typical type of case.

Emphasising the proximal causes of significant case outcomes draws attention to the combination of conditions that produces such outcomes. This is not to say that the long-standing historical circumstances of a case are unimportant, but rather that causes do not operate at a distance: it is the proximal causes that trigger an outcome. For example, long-standing poor management by itself may not have been responsible for a company's bankruptcy or poor public image until it combined with worsening economic conditions and increased competition.

The analysis of such combined causes is central to Ragin's qualitative case comparison method. The point is that different combinations of conditions may produce the same effect. In the above example, a further condition, such as a partnership or takeover, might have avoided bankruptcy or a poor public image.

The proper analysis of cases, whether in business studies or in other disciplines, provides the scientific basis for professional applications. Case studies are necessary in those areas where experimental controls cannot be applied, or where the assumptions underlying legitimate measurement and statistical controls cannot be met.

Example

This example uses extracts from data published in the *Financial Times* for the weekend 7/8 April, 2001. The data provide numerical details supporting a league table of 97 universities in the UK based on 20 weighted attributes. Most of these attributes are objective and quantitative; a few are based on subjective marks. In order to simplify the example, 14 universities and four attributes were selected. The 14 selected universities are distributed fairly evenly across the range for each of the four selected attributes.

Two outcomes were considered: employment prospects of graduates, and overall league table rank. Success (S) for each outcome was defined as being in the top seven, and failure (F) in the bottom seven, of the 14 selected universities.

The four attributes were as follows:

A = Library resources/FTES
B = Research assessment grade
C = Teaching assessment grade
D = Staff/student ratio

There are 16 possible combinations of these four attributes, representing 16 different 'types' of case. They range from the absence of any of the attributes to the presence of all four. Note the use of binary data: presence (1) or absence (0) in Table 1.

Table 1 shows 16 different types of case defined by their combination of attributes, together with their observed frequency of occurrence, and success (S) or failure (F) for two outcomes. A question mark (?) indicates where an outcome is unknown. When examining the attributes associated with success, lower case letters denote absence (low score), and capital letters denote presence (high score). These are shown in the right-hand columns of Table 1, headed 'Combinations'. The first argument is that universities with relatively high scores on attributes A, C and D should enhance graduates' employment prospects. The second argument is that relatively high scores on attribute B have a disproportionate effect on a university's overall reputation (league table ranking).

Table 1 illustrates the 16 ways in which the four attributes are combined. Eight of these conceivable combinations are not represented among the 14 universities selected for this example. Also shown are the frequencies with which the remaining eight types of combination occur, the success or failure for Outcome I (employment prospects), and Outcome II (overall league table rank).

From Table 1, readers can easily work out which combinations of conditions lead to relative success (S) for Outcomes I and II. However, this is a simple example. With larger, more complicated sets of data, resort to Boolean algebra is helpful. A Boolean 'product' is a particular set of attributes. The data on the relatively successful universities are represented as 'sums of products', using Ragin's notation. Outcome I (employment prospects) is as follows: S = abcd + aBCd + Abcd + ABcD + ABCD. Boolean minimization reduces this complexity by eliminating one or more attributes that distinguishes

Table 1: Attributes and Outcomes

Type	*A*	*B*	*C*	*D*	*Frequency*	*Outcome*		*Combinations*	
						I	II	I	II
1	0	0	0	0	3	S	F	abcd	
2	0	0	0	1	0	?	?		
3	0	0	1	0	1	F	F		
4	0	0	1	1	1	F	F		
5	0	1	0	0	0	?	?		
6	0	1	0	1	0	?	?		
7	0	1	1	0	1	S	S	aBCd	aBCd
8	0	1	1	1	1	F	S		aBCD
9	1	0	0	0	2	S	F	Abcd	
10	1	0	0	1	0	?	?		
11	1	0	1	0	0	?	?		
12	1	0	1	1	0	?	?		
13	1	1	0	0	0	?	?		
14	1	1	0	1	2	S	S	ABcD	ABcD
15	1	1	1	0	0	?	?		
16	1	1	1	1	3	S	S	ABCD	ABCD

Notes
Sixteen possible combinations of four attributes of UK universities, together with their frequency of occurrence, and their association with (S) success or (F) failure on each of two outcomes: (I) employment prospects of graduates and (II) overall *Financial Times* league table position. Success is defined as being in the top seven of the 14 selected universities for each outome. Obviously, outcomes are not known for combinations that do not occur in the data, and these unknown outcomes are labelled with a question mark. The last two columns show which attributes are associated with each outcome, using Ragin's notation. See text for further details.

between two products: ABcD + ABCD = ABD; abcd + Abcd = bcd; aBCd cannot be further reduced. This result shows that any one of three sets of conditions (ABD + bcd + aBCd) can account for the relatively high employment prospects enjoyed by graduates of seven of the 14 selected universities. The Boolean sums of products for Outcome II (overall league position) are as follows: S = aBCd + aBCD + ABcD + ABCD. Boolean minimization reduces this complexity: ABCD + aBCD = BCD; ABcD + ABCD = ABD; aBCd + aBCD = aBC. This result shows that any one of three sets of conditions (BCD + ABD + aBC) can account for the relatively high overall league table rank of seven of the 14 selected universities. Thus, the first argument is not justified by the results. The combination bcd, meaning relatively low scores for research, teaching and s/s ratio, can be associated with good employment prospects for graduates. The second argument receives some support because each of the three sets of conditions includes B.

Naturally, innumerable conditions are taken for granted when considering business issues. In a research investigation, however, the main interest is in a particular set of critical conditions or attributes thought to be relevant to the theory that underpins the investigation. A different pattern of outcomes might have shown a

different combination of attributes. The logic of case comparisons depends on the theory (argument) that underpins an investigator's choice of cases, attributes and outcomes. This makes some theoretical assumptions explicit. A case comparison approach to companies would be useful in showing how different sorts of firms respond to similar threats to their corporate reputation. Restrictions on space do not permit further consideration of the case comparison method or the role of theory construction in relation to case study methods generally.

The Quasi-Judicial (QJ) Case Method

There are two important features of the QJ case method:

— it adapts a 'legal science' approach to human social issues that are not amenable to a 'natural science' approach
— it emphasizes the structure and function of real-world arguments, using a procedure called 'substantive logic' developed by Toulmin (Toulmin et al., 1979). Substantive logic, as explained below, enables investigators to construct or dissect complex arguments, using simple diagrams and labels.

HISTORICAL NOTE

Half a century ago, Towl (1954) drew an analogy between a business case report and a legal court report of a systematic body of precedents. Towl thus anticipated the arguments underpinning the Quasi-Judicial (QJ) case method. The *Harvard Business Reports* were abandoned in 1932 because it appeared that a broad body of precedents was of limited use, and because a problem-oriented approach to specific issues seemed more promising. This may have been true at the time; the question now is whether a more rigorous, scientific, case study method and a more demanding business climate can revive the 'legal science' approach. The *Harvard Business Reports* (from 1925 onwards), and the *Harvard Bureau of Business Research Bulletins,* provide material for researchers interested in the reasons why the legal science model for the case method in business studies was abandoned.

More recently, Easton (1992) offers a critique of the traditional use of the case method in business studies, and an up-to-date guide to practical applications, but appears to be unconcerned with the problem of classifying cases or making inductive inferences about cases of a particular type. Green (1992) uses cases to illustrate how failure to identify, or deal effectively with, risks to corporate reputation or brand image can result in serious damage.

In contrast to other case methods, the QJ case method uses a systematic inductive procedure to construct arguments about a specific case, or a set of similar cases. The QJ method tries to establish the causal connections referred to in the substantive arguments used to describe and analyse a case. These are the arguments used to explain or justify a particular business outcome. In this respect, the QJ method is analogous to the case method in forensic inquiry or accident investigation, or in historical research (Winks, 1968).

Applying the QJ method to a set of recorded (completed) business cases may enable one to collect cases that seem to show a common logical/causal pattern, and to categorize them as cases of a certain type. This is analogous to the way case law is established in jurisprudence. Well-documented, carefully analysed cases in various categories, such as brand management or corporate communication, should be useful to business professionals. A body of case law in business studies should help professionals to appreciate some of the key issues in the management of corporate reputation in their sector of industry. Fishman (1999)

describes what a case law approach would imply for professional clinical psychology.

The QJ method is well adapted to deal with cases that are non-routine — unique or unusual cases that do not fit a familiar pattern. The method was originally developed to deal with 'individual cases' of personal adjustment — the study of individual persons in problematical situations — as in clinical psychology and counselling (Bromley, 1986, 1990), where unique psychological factors and circumstances determine the outcome.

Perhaps the nearest business case studies come to the legal model is when they deal with breaches to a company's code of conduct. It is usual, in cases of this sort, for a company to set up a formal panel of investigators and to follow a formal procedure, paying very close attention to the rules of evidence and argument. Other obvious examples would be internal investigations into serious management failures. These sorts of sophisticated case studies contrast with casual, anecdotal reports of incidents, which are not scientific case studies but might provide a starting point for a rigorous inquiry.

The Parallel with Legal Science

The QJ case method is based on the philosophy of jurisprudence (legal science). More precisely, it is modeled on the 'inquisitorial' method in legal science, not on the 'adversarial' system practised in the UK. It aims to be scientific in the general sense of being empirical and rational, and is not unduly bound by legalistic procedures and precedents, or by cultural values, hence the 'Quasi' prefix. All scientific case studies, however, should meet certain requirements relating to empirical data and rational argument. The elimination of unsubstantiated or irrelevant evidence, and the elimination of formal logical fallacies, is of course necessary in order to establish the validity of a substantive, real-world, argument.

A business case based on quasi-judicial principles should offer cogent arguments that reach sensible conclusions about the problems faced by a company in a particular set of circumstances, taking account where necessary of the firm's corporate history. A simple mnemonic symbolizes the relationships between a company, the situation with which it is trying to cope, and the outcome of that interaction:

Ci x Sj –> Bij

— Ci represents a particular company
— Sj represents a particular set of circumstances
— x represents the interaction between Ci and Sj
— Bij represents the behavior (performance) of the company, and the outcome of the interaction between Ci and Sj
— → represents the causal connection between the interaction and the associated outcome Bij, such as the company's success or failure in dealing with a threat to its reputation.

This pattern of facts and relationships is most easily seen in a summary account of a case that gives the gist of the argument. Naturally, this summary depends on the truth and validity of the details of the argument. A QJ case study collects observational, documentary and material evidence, preferably corroborated, about each of the elements represented by Ci x Sj → Bij.

A case study maps real-world data onto this abstract, general framework, and expresses it in terms of a detailed substantive argument, which is, in effect, a theory. The evidential facts, the observational data, do not speak for themselves. They have to be spoken for by a theory, an argument using assumptions, concepts, rules, and generalizations. This imposes a pattern of meaning onto the available evidence, thus explaining the nature of the interaction and

generating hypotheses about performance and outcome.

A narrative gives a largely descriptive, sequential account of a case (an episode or set of related episodes), whereas an explanation involves the attribution of causes, relationships, and implications. Different investigators may have different theories, assumptions, and data, and so reach different conclusions.

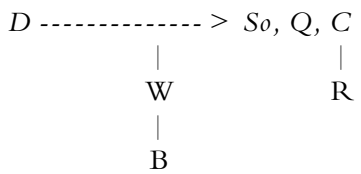
In routine cases, such as firms with cash flow problems or poor public relations, a familiar pattern of facts regarding Ci x Sj → Bij may emerge. This allows an investigator to impose a familiar, standard interpretation based on previous cases. For example, an investigator may explain a company's poor performance by reference to poor monitoring by a bank combined with inexperienced management. In frequently occurring cases, investigators tend to rely on familiar, easily recalled, preestablished patterns of argument. The danger is that cognitive biases may hinder the search for newer and better arguments (Kahneman et al., 1982).

Substantive Arguments

A substantive argument about a case refers to evidence, assumptions, implications, reservations and rules of inference. These need to be made explicit, otherwise the argument is incomplete and possibly misleading. The interrelated elements in an argument constitute an explicit theory about the case in question, and by implication constitute a theory about other cases of that sort. This can happen even while other cases of the same sort have yet to be identified.

Generalization is a basic cognitive process independent of sophisticated sampling theory. A generalization or theory becomes a search instrument; this is best illustrated, perhaps, not so much in business case studies as in accident and forensic cases.

A substantive logical argument is an explicit, open-ended method of reasoning about matters of fact in the real world. It contrasts with a closed, formal logical argument, the validity of which is tested by reference to rules of inference, independently of empirical data. The overall structure of a case study can be divided into an interrelated set of problem areas (Bromley, 1986; Easton, 1992). Each problem area is described in terms of an argument that imposes a pattern of meaning onto the available data (relevant evidence). Following Toulmin et al. (1979), the basic structure of a substantive logical argument dealing with a problem area is symbolized as follows:



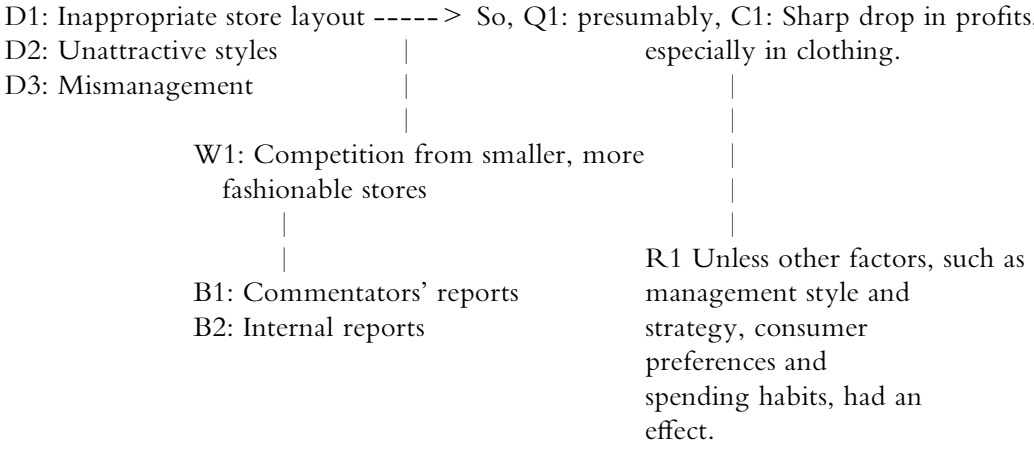
— C (claim): the conclusion
— D (data): relevant empirical evidence
— Q (modal qualifier): subjective likelihood or probability of the claim being true
— W (inference warrant): assumption, rule, or theory that justifies the claim (C) on the basis of the data (D)
— B (backing): background, contextual information justifying W. Requiring B (backing) to be made explicit as a justification for W (inference warrant or theory) forces disclosure of the grounds for believing W to be true
— R (rebuttal): conditions under which the argument would fail.

Colloquially speaking, the above elements in a substantive argument pose the following questions, which are used to construct or dissect an argument:

— C: What am I (you) trying to prove?
— D: What evidence do I (you) have to go on?

EXAMPLE

The following example is based on media reports about a well-known chain of retail stores:



— Q: How likely is it that my (your) conclusion is correct?
— W: What entitles me (you) to draw that conclusion from that evidence?
— B: What is the justification for my (your) line of reasoning?
— R: What assumptions am I (you) making?

A more detailed case report would contain separate arguments dealing with each of a number of problem areas. Some of these arguments would be interrelated in the sense that the conclusion (C) of one argument may provide the data (D) or other component (W, B or R) of another argument. Substantive logic encourages us to think in terms of a 'web' rather than a 'chain' of argument. See Bromley (1986) for concrete, detailed, diagrammatic examples, especially Case J (pp. 212–214) and Figures 9.4(i) to 9.4(iv).

CONCLUSIONS

The traditional league table approach to assessing and comparing corporate reputations faces a number of difficulties associated with defining and measuring reputation. These difficulties include doubts about the measurement assumptions underlying psychometric methods (using subjective judgments) and consequent doubts about the assumptions underlying multivariate statistical analysis of psychometric data.

Even assuming that carefully constructed psychometric assessments can generate scientifically acceptable numerical (quantifiable) data, the effort and care required to construct such scales are considerable.

Calculating reputation quotients (RQs) or benchmarks for comparing corporate reputations calls for a departure from the traditional league table method, but still depends on questionable assumptions about the legitimacy of psychometric assessment.

Two case methods depend less or not at all on psychometric and statistical assumptions. The case comparative method deals with smaller or larger samples of cases, and with smaller or larger numbers of attributions. It relies on the logic of Boolean algebra. The Quasi-Judicial (QJ) case method is modelled on judicial procedures for dealing with individual cases. The pattern of argument revealed by substantive logic may permit findings to be generalized to a class of similar cases.

References

Balmer, J.M.T. and Soenen, G.B. (1999) 'The acid test of corporate identity management', *Journal of Marketing Management,* 15 (1–3): 69–92.

Bennett, R. and Gabriel, H. (2001) 'Corporate reputation, trait covariation, and the averaging principle. The case of the UK pensions mis-selling scandal', *European Journal of Marketing,* 35(3–4): 387–413.

Bromley, D.B. (1986) 'The case study method in psychology and related disciplines', John Wiley & Sons, Chichester.

Bromley, D.B. (1990) 'Academic contributions to psychological counseling: 1. A philosophy of science for the study of individual cases', *Counselling Psychology Quarterly,* 3 (3): 299–308.

Bromley, D.B. (1993). 'Reputation, image, and impression management', John Wiley & Sons, Chichester.

Bromley, D.B. (2000a) 'Psychological aspects of corporate identity, image, and reputation', *Corporate Reputation Review,* 3(3): 240–252.

Bromley, D.B. (2000b) 'The scientific basis of the case method in business studies', Unpublished article, Department of Psychology, The University of Liverpool, Liverpool, UK.

Bromley, D.B. (2001) 'Relationships between personal and corporate reputation', *European Journal of Marketing,* 35 (3–4): 316–334.

Easton, G. (1992). 'Learning from case studies' (2nd ed.). Prentice-Hall, New York.

Fishman, D.B. (1999) 'The Case for Pragmatic Psychology', New York University Press, New York.

Fombrun, C.J. (1995) 'Reputation: Realizing Value from the Corporate Image', Harvard Business School Press, Boston, MA.

Fombrun, C.J., Gardberg, N.A. and Sever, J.M. (2000) 'The reputation quotient sm: a multi-stakeholder measure of corporate reputation', *The Journal of Brand Management,* 7 (4): 241–255.

Fombrun, C.J. and van Riel, C. (1997) 'The reputational landscape', *Corporate Reputation Review,* 1 (1): 5–13.

Gomm, R., Hammersley, M. and Foster, P. (eds.) (2000) 'Case Study Method', SAGE publications, London.

Green, P.S. (1992) 'Reputation Risk Management', Pitman Publishing, London.

Hamel, J. (1993). 'Case Study Methods', SAGE publications, Newbury Park, CA.

Kahneman, D., Slovic, P. and Tversky, A. (eds). (1982) 'Judgment Under Uncertainty: Heuristics and Biases', Cambridge University Press, Cambridge.Kay, J. (1993) 'Foundations of Corporate Success', Oxford University Press, Oxford.

Keen, C. and Greenall, J. (1987) 'Public Relations Management in Colleges, Polytechnics, and Universities', Heist Publications, Banbury.

Kerlinger, F.N. (1986) 'Foundations of Behavioral Research' (3rd. ed.). Holt, Rinehart, & Winston, New York.

Kline, P. (1998). 'The New Psychometrics: Science, Psychology and Measurement', Routledge, London.

McNair, M.P. (ed.), and Hersum, A.C. (1954) 'The Case Method at the Harvard Business School', McGraw-Hill Book Company, Inc., New York.

Mercer, J. (1996) 'Reputation and International Politics', Cornell University Press, Ithaca.

Michell, J. (1990) 'An Introduction to the Logic of Psychological Measurement', Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Michell, J. (2000) 'Normal science, pathological science, and psychometrics', *Theory and Psychology,* 10 (5): 639–668.

Ragin, C.C. (1987) 'The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies', University of California Press, Berkeley, CA.

Ragin, C.C. and Becker, H.S. (eds.) (1992) 'What is a Case? Exploring the Foundations of Social Inquiry', Cambridge University Press, Cambridge.

Toulmin, S., Rieke, R. and Janik, A. (1979) 'An Introduction to Reasoning', Collier Macmillan, London.

Towl, A.R. (1954) 'The use of cases for research', in M.P. McNair (ed.) with the assistance of A.C. Hersum, 'The Case Method at the Harvard Business School:' 223–229. McGraw-Hill Book Company, Inc., New York.

van Riel, C.B.M., Stroeker, N.E. and Maathuis, O.J.M. (1998) 'Measuring corporate images', *Corporate Reputation Review,* 1 (4): 313–326.

Whetten, D. (1997) 'Theory development and the study of corporate reputation', *Corporate Reputation Review,* 1 (1–2): 26–34.

Winks, R.W. (ed.) (1968) 'The Historian as Detective: Essays on Evidence', Harper & Row, New York.

Yin, R.K. (1984) 'Case Study Research: Design and Methods', Sage Publications, Beverly Hills, CA.

Yin, R.K. (1993) 'Applications of Case Study Research. Sage Publications, Newbury Park, CA.