
Epistemological limitations in quantitative marketing research: implications for empirical generalisations

Dr Gary Robertshaw

Abstract

Increasing attention has been dedicated recently to the subject of empirical generalisations in marketing science and the role of replication studies.

This paper considers the implications of epistemological limitations in a quantitative marketing research context, arguing that these are different and more numerous than those of a general nature, requiring special treatment and solutions. A number of specific problems are discussed including the Hawthorne effect, researcher-participant interaction, context dependency, conditioning, atypical research participants and differing measurement methods.

Quantitative marketing research, often regarded as scientific and objective by academics and practitioners, may be less reliable than generally claimed within the literature. Whilst the use of replication studies and the concept of falsifiability can facilitate genuine progress in marketing knowledge, the contention of this paper is that these need to be applied more fastidiously and with greater recognition of their limitations.

General epistemological background

There are fundamental limitations to the pursuit of all empirical knowledge. The Duhem–Quine thesis, for example, is predicated on the impossibility of testing a scientific hypothesis in isolation because any test of the hypothesis always relies on one or more background assumptions. The hypothesis in question is thus itself incapable of making predictions. Instead, the consequences of the hypothesis rest on background assumptions from which to derive predictions.

A variation on the Duhem–Quine thesis is epistemological holism - the claim that a single scientific theory cannot be tested in isolation. That is, a test of one theory always depends on other theories and hypotheses. That theories can only be tested as they relate to other theories implies that test results that seem to refute a favoured scientific theory have not refuted that theory at all. Rather, it can be argued that the test results conflict with predictions because some other theory is false or unrecognised. For example, the existence of dark matter has been hypothesised to explain the strange motions of some galaxies. In fact, dark matter may not exist and the unexpected motion of galaxies may be due to incorrect prior hypotheses.

Although a collection of theories (i.e. a theory and its background assumptions) can be tested as a whole against the empirical world and falsified if they fail the test, the Duhem–Quine thesis makes it impossible to isolate a single, individual hypothesis from this collection. The context in which a scientific test is carried out and its relationship to the underlying theoretical assumptions must therefore always be considered – a form of ‘contextual relativism’. When researchers have reason to accept that the background assumptions are true, in the sense that they are supported by observation, then they have rational but inconclusive reason for believing that a theory may be wrong if it fails an empirical test.

It is important to recognise, however, that observation alone cannot falsify theories (Chalmers, 1982; Anderson, 1983; Ehrenberg, 1994). For instance, to ‘disprove’ the notion that the Earth was moving, it was argued that birds were not thrown off into the sky whenever they let go of a tree branch. This ‘observational proof’ is no longer accepted as empirical evidence that the Earth is not moving because we have since adopted a different background hypothesis and understanding of how gravity works. Since researchers cannot unambiguously determine which

theory is refuted by unexpected observations means that judgements must be used about which theories to accept and which to reject.

Epistemological limitations are not simply an artifact of consumer research but are inherent in nature and pervade all fields of study, including physics, mathematics and logic. For example, Heisenberg's uncertainty principle, a cornerstone of quantum mechanics, shows that definite and precise outcomes simply do not exist – only statistical probability. In the same manner, Godel's incompleteness theorem is based on the paradoxes that surround self-reference, for example, 'this statement is a lie'. If the statement is true, then it is false; and if it is false then it is true. Godel went on to adapt these difficulties of self-reference to the subject of mathematics. Godel considered the relationship between the description of mathematics and the mathematics itself. By identifying the subject with the object – incorporating the description of the mathematics within the mathematics – Godel uncovered a paradoxical loop that led directly to the inevitability of undecidable propositions (Deutsch, 1998). This led to the inescapable conclusion that it is fundamentally impossible to discover all truths in mathematics.

Objective knowledge is thus inherently limited because it is based on fallible theories, hypotheses and observations. That is, there is no independent, objective and definite 'truth' that can be discovered. We can be successful in developing an ever deeper understanding of phenomena but nature places an inherent limitation on ever acquiring a wholly complete understanding in the sense that our knowledge can never be certain (see also Anderson, 1983; Hunt, 1990, 1993; Peter, 1992; Deutsch, 1998).

Epistemological limitations in a quantitative marketing research context.

In the case of quantitative marketing research, epistemological limitations are more pervasive, numerous and complex due to the unavoidable interaction between research participant and researcher, and the influence of contextual externalities. This situation applies to both academic and practitioner spheres of market research. The contention of this paper is that these added limitations are often neglected in developing empirical marketing generalisations, and that they require special attention and solutions.

This perspective has tended to be overlooked in the literature. For example, whilst Hunt (1990; 1993) discusses the general problem of inter-subjective certifiability, the specific epistemological limitations associated with quantitative marketing research are not considered. Similarly, Wright and Kearns (1998) suggest replication as a solution to the problems of fallibility in observation, but do not consider those problems that replication fails to address.

This paper explores the implications of six epistemological problems for empirical marketing generalisations.

Hawthorne Effect

In quantitative marketing research, the fact that an experiment is being conducted can have an important effect on the research participant - the so-called Hawthorne effect. Whilst the topic of research is often not disclosed to participants in studies, participants are aware that they are engaged in a research study. It is this awareness of participation, irrespective of prior knowledge of the research topic, that can lead to some degree of behaviour modification and a subsequent diminution in the integrity of the results obtained.

Even ostensibly unobtrusive methods can affect behaviour because, under the Data Protection Act 1998, research participants must be informed that their actions are being recorded. The Hawthorne effect thus also applies to remote observations and recordings that are designed to

minimise measurement influence on the consumer, such as eye movement scanning and in-store purchase behaviour.

Under the Hawthorne effect it is impossible to determine what the outcome of the research would have been in the absence of participant awareness of the research study being undertaken. This is an important point since it introduces an inherent uncertainty into the reliability of the results obtained that cannot be controlled and expressed in terms of confidence levels. This situation can be regarded as a further manifestation of the Duhem-Quine thesis in that empirical tests are inextricably linked to prior background assumptions; in this case that participant awareness of the research process has no influence on the results obtained. Whilst replication studies that show consistency in research outcomes add confidence to the credibility of results, there remains a not insignificant possibility that the Hawthorne effect is endemic across all the replications. Consequently, measurements of behaviour may not be reflective of 'real life' behaviour.

Inseparability of Researcher and Research Participant

Many quantitative methods are rooted in the belief that the researcher can be removed from the research participant being studied. That is, reality exists independently of the researcher and can be discovered objectively and value free (Neuman, 1997; Deutsch, 1998). Under this paradigm, data and their analysis are value-free and do not change because they are measured. Quantitative measures are thus regarded as independent facts relating to a single, apprehensible reality.

However, this viewpoint is erroneous because all obtrusive research measurements can affect responses and behaviours (Lehmann et al, 1998; Christy and Wood, 1999; Szmigin and Foxall, 2000). For example, Morwitz et al (1993) and Morwitz and Gavan (2004) have shown that interviewer gender can significantly alter the responses provided in personal interviews. Interviewees also tend to provide more agreeable responses in the presence of an interviewer (Lehman et al, 1998), such that results obtained from face-to-face methods can differ with indirect methods such as postal surveys, even when measuring the same underlying construct (Colet, 1999).

The inseparability of researcher and research participant is a fundamental aspect of nature and not simply a peculiarity of quantitative marketing research. In quantum mechanics, for example, the outcome of experiments has been shown to be dependent on whether or not an observer is present. In fact, the whole outcome of an experiment can be determined by the act of observation, as illustrated in the famous 'Schrodinger's cat' thought experiment (Deutsch, 1998). It is thus important to recognise that the effect of measurement on the outcome of quantitative research is not restricted solely to those situations in which there is direct interaction between researcher and research participant. To varying degrees, all measurement techniques fundamentally affect the results obtained. Measured behaviour can never provide a true representation of actual behaviour.

Context Dependency

As Robson (1993) has stated, a purely positivist view is inappropriate when researching consumer behaviour since treating participants as independent objects ignores their ability to reflect on problem situations, and to act on these in an interdependent way. Consumer behaviour cannot be confined to laboratory experimental conditions and does not operate in mechanical ways (Barwise, 1995). Unlike a laboratory where conditions for the effective triggering of causal mechanisms may be created, no such opportunity for quantitative marketing research methods exists (Bass, 1995; Pawson and Tilley, 1997).

Research findings must therefore be considered in the context in which the data gathering occurs. That is, the response to questions depends partly on the circumstances of the research participant at the time and other alternatives considered. As Bickart (1993) and Simmons et al (1993) have shown, even seemingly innocuous differences in surroundings can affect responses. Whilst in situ market research can reduce the influence of surroundings on response, it cannot eliminate these influences. Further, exogenous events occurring at the time of the data gathering may also influence the results obtained (Morwitz and Gavan, 2004). For example, Tweney (1998) reported increased public awareness of data protection issues following a number of accidental disclosures of personal details over the Internet.

As a further example, suppose a company's sales increase by 20% following a 20% increase in advertising in a particular location, but remain unchanged in the remaining locations. The company may deduce that the advertising caused sales to increase by 20%. However, this deduction could be wrong for several reasons. It could be due to randomness, and other explanations such as a major competitor coincidentally withdrawing from the market in that location, or an economic upsurge in the same location. Though the use of control and tests cells can establish concurrent variation and precedence in the results obtained, these can never eliminate the possibility of alternative explanations for the research findings. This is especially the case in quantitative marketing research where external, confounding influences are more numerous.

Whilst commercial market research can be both exploratory and descriptive in nature as well as solely seeking to derive causal inferences, it is important to note that consistency in repeat observations does not necessarily demonstrate causality. This is because the findings may be due to some coincidental and unconsidered externality.

Context dependency also places a limit on generalising empirical findings. This was illustrated by Deutsch (1998) in a hypothetical story about a farmer who starts bringing his chickens more food than usual. How this observation is extrapolated to predict the farmer's future behaviour depends entirely on the context in which the observation is considered. According to the 'benevolent-farmer' theory, it is evidence that the farmer's benevolence towards chickens has increased. But according to the 'fattening-up' theory, the behaviour is ominous – it is evidence that slaughter is imminent. The fact that the same observational evidence can be extrapolated to give two diametrically opposed predictions is not some incidental limitation of the farmyard environment; it is true of all observational evidence. That is, the extent to which the results obtained can be generalised and used to make predictions must be considered in the context in which they are gathered.

Conditioning

Consumers interact with their environment, and are shaped and nurtured by that environment, leading to conditioning of responses. Conditioning manifests itself in a number of different forms and can lead to different research outcomes. In some cases, research participants may knowingly falsify responses where they believe a socially desirable response is required, such as questions relating to eating habits, alcohol consumption, environmental and charitable concerns (Lehmann et al, 1998; Robertshaw and Marr, 2005).

Research participants may also operate under the preconceived notion that the answers they provide should be consistent. Thus, rather than providing true responses in the sense that they reflect actual behaviour, some research participants will respond in a manner they perceive to be consistent throughout the measurement process. Other research participants tend to either overstate their true beliefs and opinions, or to understate these as an expression of humility such

that more emphasis is placed on the extremities of the measured values (Christy and Wood, 1999; Szmigin and Foxall, 2000).

Furthermore, repeated participation in research studies can give rise to an 'expert' consumer leading to more careful consideration of responses. Therefore it cannot be assumed that repeated measurements have no effect on respondents and that respondent behaviour will remain constant over time. This is important because extant sampling approaches typically do not consider previous exposure of participants to testing and the conditioning effects of repeated participation.

Whilst sampling frames can be expected to contain similar proportions of expert and non-expert consumers and similar levels of response exaggeration such that the net effect is cancelled out in replication studies, there remain a number of areas in which the biasing effects of conditioning cannot be ignored. Firstly, it is more likely that sampling frames will be over-represented with respect to expert consumers since individuals who choose to participate in research studies are more likely to be exposed to repeat studies compared to non-participants. Secondly, responses relating to market sectors where elements of social desirability are more pronounced, for example, questions concerning charitable donations, drinking habits and political affiliations, are more likely to be falsified due to social conditioning.

Atypical Research Participants

Those individuals electing to participate in marketing research studies are by definition atypical of the general population in at least one respect in that they have elected to become involved whilst others have abstained (Greenleaf, 1992; Lehmann et al, 1998; Christy and Wood, 1999; Szmigin and Foxall, 2000).

There may also be subtle differences in personality and attitude not accounted for in the sampling procedure, which can lead to bias in the results obtained. For example, one study has shown that differences can exist in human values between research participants and non-participants, which are often missed in typical sampling frames (Robertshaw and Marr, 2005). This can have ramifications because differences in values can lead to differing responses in ostensibly unbiased sampling frames, which have been based on demographic criteria such as age, income and occupation.

Other researchers (Schwartz and Bilsky, 1987; Kamakura and Novak, 1999) have shown that different values segments lead to differences in behaviour and responses. For example, one segment was motivated by security, typified by risk aversion and less likelihood of engaging in independent decision-making, creativity and action. At the other extreme, another segment was found to be more likely to engage in social activities such as participation in surveys. This situation can lead to an undetected, over-representation of particular values types within sampling frames that lead to differing responses, which are not truly reflective of the target population.

Sampling frames tend to be based on general, demographically defined characteristics such as age, occupation, geography, income and gender with disregard for more abstract characteristics such as values (Kahle and Kennedy, 1989; Robertshaw and Marr, 2005). The omission of higher abstraction but consequential sampling criteria such as values, because of lack of prior consideration or impracticality for inclusion, can thus lead to bias in research outcomes. Many claims for empirical generalisation in the marketing literature fail to acknowledge this situation and the potential for bias arising from atypical research participants.

Measurement systems

In quantitative marketing research, differing systems of measurement can yield differing results when applied to the same research problem. This is an important point because there is no universal methodology that dictates the use of a specific measurement technique in a given situation. The choice is subjective and relies to some degree on the researcher's intuition including the number of scale points to employ, question phraseology, type of questions posed and type of measurement scale.

For example, some researchers claim that the absence of a mid-point on even number scales fails to provide respondents with an expression of neutrality. Others have argued that mid points on odd number scales are undesirable since they allow for simplification of the response process through non-commitment on the part of respondents (Szmigin and Foxall, 2000). The selection of odd or even number scales represents just one area where researcher intuition can influence the results obtained.

The use of a greater number of scale points in an attempt to increase construct discrimination creates an increased likelihood of response error due to respondent fatigue and the scale points becoming unwieldy thereby losing relevance (Johnson et al, 1990; Lehmann et al, 1998). More recent research (Viswanathan et al, 2004) has shown that consumers deal with complex information along a dimension by recoding it into chunks or categories, thereby conserving their capacity to process information. This is due to a limit on human ability to discriminate beyond seven levels. Maximum discriminations may therefore be merely an artefact of the measurement scale and not reflective of true consumer discriminations. There thus exists a compromise in terms of the information generated by a scale for purposes of the researcher and limitations in the research participant's ability to discriminate beyond a certain number of response categories. Essentially, there exists no universally agreed optimum number of scale points to employ in the measurement process, and differing numbers of scale points can yield differing results.

This situation also directly affects the outcome of quantitative reliability measures, such as Cronbach's alpha, which is represented by the formula:

$$\alpha = \frac{k\Gamma}{1+(k-1)\Gamma}$$

Where k represents the number of scale point items, and Γ represents the average inter-item correlation among measures of the construct.

The higher the score, the more reliable the quantitative findings are considered to be. It can be seen from the formula that an increase in the number of scale items increases Cronbach's alpha. However, as Viswanathan has shown, research participants cannot make meaningful discriminations beyond a certain number of points. The measure of reliability is therefore partly a product of the researcher's subjective choice of scale points and partly the research participant's ability to discriminate between scale points. By extension, differing numbers of scale points will lead to differing reliability measures.

Further, the choice of particular descriptions on measurement scales is subjective and differing consumer interpretations of the meaning of these descriptions can elicit differing responses. This situation can be exacerbated due to items in the first position on a question scale typically receiving a disproportionate number of mentions due to heightened visibility (Lehmann et al, 1998).

Consider, for example, a measurement scale designed to measure product usage, using the following scales:

Scale 1

Very often	Often	Occasionally	Hardly ever	Not at all

Scale 2

Very often									Never
1	2	3	4	5	6	7	8	9	10

Scale 3

Regularly	Sometimes	Not much	Never

Whilst scales 1-3 are designed to measure the same underlying construct, the particular method used is subjectively determined by the researcher. This subjectivity unavoidably permeates the empirical findings, which can lead to different findings when applied to the same underlying construct.

There is evidence within the literature to suggest that different measurement scales and measurement methodologies do in fact lead to different research outcomes. For example, Colet (1999) discusses the dilemma of different answers to the same business question and inquires how and why different measurement approaches yield different results. Luce and Tukey (2002) discovered that different methods of interviewing respondents could lead to different results, whilst Pineau and Slotwinder found that different research methodologies can yield different results leading to different business decisions.

A number of attempts have been made to circumvent the preceding issues and reduce researcher influence, the most notable being the Juster scale, which categorises responses according to probabilities rather than requiring absolute responses. The Juster scale is based on the premise that verbal intentions are really disguised probability statements and that self-predictions of future behaviour are usually conditional; they depend on what happens over the period concerned. Its chief application is in predicting future purchase intentions, though adaptations of the scale have been used in differing fields of study, for example voting intentions. However, whilst the probabilities associated with the Juster scale improve predictive accuracy over absolute choices, it remains susceptible to the same limitations in respect of scale points and descriptions and its uses are restricted to those situations where intentions are being measured (Gendau et al, 1991).

Scale 2 (Juster scale adaptation)

Almost no chance (1 in 100 chance)	Very slight possibility (1 in 10 chance)	Slight possibility (2 in 10 chance)	Some possibility (3 in 10 chance)	Fair possibility (4 in 10 chance)	Fairly good possibility (5 in 10 chance)	Good possibility (6 in 10 chance)	Probable (7 in 10 chance)	Very probable (8 in 10 chance)	Almost sure (9 in 10 chance)	Practically certain (99 in 100 chance)
0	1	2	3	4	5	6	7	8	9	10

Falsifiability and replication as solutions to epistemological problems

According to Popper (1935), the first criterion for any empirical theory is that it be falsifiable; it must be possible to conceive of an observation which would contradict the theory. If we cannot do this, Popper would say that the theory is unscientific, and provides no basis for epistemic progress (Wright and Kearns, 1998). For example, Newton's theory of gravity was falsifiable, and superseded by Einstein's general theory of relativity, with the latter successfully predicting outcomes that were at odds with the former. In contrast, neither astrology nor solipsism is falsifiable because there is no direct evidence that could refute their predictions or viewpoint. In the same manner, advances in empirical marketing theories can only be achieved where such theories can at least be falsified in principle.

When a conjectured theory disagrees with observation then it is abandoned. If one of our originally held theories is abandoned in favour of one of the newly proposed ones, then we tentatively deem our problem-solving to have made progress. This progress can only be tentative because subsequent problem solving will involve altering or replacing even these new, apparently satisfactory theories (Deutsch, 1998). Empirical knowledge thus grows by conjecture and refutation, a process known as evolutionary epistemology.

In this respect, replication studies are critical in putting theories to the test of falsification, using a variety of measurement methods under different conditions to ensure that the original results are not spurious and that findings do not contradict theory. Developing generalisations relies on replication as a means of extending the theory to new situations, and of investigating the influence of different variables. Replication studies therefore give added credence to a theory by mitigating the limitations of single observations through the completion of studies with further samples. This helps to lessen the problem of sampling error by undertaking measurements at different times and places to moderate the effects of confounding influences, and using additional observers to reduce the influence of subjectivity, bias and error.

A problem persists, however, with lack of acknowledgement of the fallibility of empirical knowledge within the marketing literature (Peter, 1992; Hunt, 1993), coupled with a plethora of one-off empirical studies that claim to be conclusive in their findings (Wright and Kearns, 1998). Hubbard and Armstrong (1994), for example, point out that marketing has a preponderance of un-replicated studies. The findings of such studies cannot be viewed with confidence until they have been reproduced by different researchers using different data collection methods under different conditions. Falsifiability and replication not only provide effective methods for ensuring that research results are robust, they also play a vital role in the development of empirical generalisations by testing whether the initial findings still occur under different conditions (Lindsay and Ehrenberg, 1993; Hubbard and Armstrong, 1994; Wright and Kearns, 1998).

Falsification and replication thus provide elegant solutions to many of the epistemic limitations associated with quantitative marketing research. The question arises, however, of the extent to which falsifiability and replication address the specific epistemological limitations raised in this paper, and whether these fail under certain situations. The contention of this paper is that falsification and replication must be applied in a more rigorous and considered framework for them to be truly effective as mechanisms for facilitating empirical generalisations, and that there remain conditions under which such generalisations are limited.

Weaknesses in falsifiability and replication as solutions to epistemological problems

Consistency in repeat observations and research findings from replication studies is generally taken to support the reliability of a given theory. However, replication consistency may not be a wholly dependable indicator of reliability. Further, different and sometimes contradictory

conclusions can be derived from the same findings. This situation requires further consideration of how epistemological limitations associated with quantitative marketing research affect empirical generalisations.

- i. According to the Duhem-Quine thesis it is not possible to test a scientific hypothesis independently since any test of the hypothesis is always dependent on background assumptions. There remains the possibility, therefore, that a prior, erroneous assumption or false premise has been perpetuated throughout the replication studies through insentience. Furthermore, consistency in repeat findings may be attributable to some unconsidered or extraneous influence. This is a particular problem in quantitative marketing research where outside influences are more numerous and pervasive, and where the likelihood of unconsidered externalities is therefore greater.
- ii. The Hawthorne effect may be present to a similar degree across replication studies, such that modification of behaviour is endemic and the findings not reflective of actual behaviour under 'real life' situations. This situation is likely to be more pronounced where the replication studies employ measurement methods that have the same level of participant awareness of the research exercise being undertaken. The Hawthorne effect thus introduces an inherent uncertainty into the reliability of results obtained that cannot be expressed in terms of confidence levels, because it cannot be known what the outcome of the research would have been in the absence of participant awareness of the research study being undertaken.
- iii. All research measurements are intrusive to some extent and can affect responses and behaviours. Fundamentally, the researcher or research mechanism cannot be separated from the research participant, and consequently the outcome of empirical tests is always influenced by the actual process of measurement. These effects can range from interviewer gender to more subtle influences such as the phraseology used in self-completed surveys and questionnaires. It is therefore logically impossible for measured behaviour to replicate actual behaviour since the process of measurement plays an intrinsic role in the results outcome.
- iv. Context dependency exists not only in relation to the environment in which measurements are undertaken but also with respect to the background in which the results are extrapolated. This is an important but often overlooked consequence of context dependency; consistency in repeat observations does not necessarily lead to consistency in predictions and explanations. In fact, it is impossible to extrapolate observations unless these have already been placed within an explanatory framework, as Deutch's (1998) hypothetical story of the farmer illustrates. Thus, different conclusions and theoretical frameworks can emerge from the same set of repeat observations.
- v. Conditioning can lead to bias as research participants tend to respond according to what they perceive the research objectives to be, rather than providing responses that are reflective of their true behaviour. This is particularly the case when the researcher is physically present during the measurement process, which can lead to participants being more inclined to provide answers that they believe will meet with the researcher's approval. Conditioning can also manifest itself in different forms, for example, where participants strive to provide what they consider to be socially responsible answers. Furthermore, repeated participation in research studies can also lead to the development of an expert consumer and over-representation of this type of consumer in replication studies, leading to bias.

- vi. Those consumers who agree to participate in research studies may not be representative of the general population, which can lead to persistent bias in repeat studies. This is more likely if the same demographically-based sample frames are used that fail to encapsulate the differing emotive and belief values of target populations, which can affect results outcomes.
- vii. The results obtained may be to some extent an artefact of the particular measurement type, scale and phraseology used, rather than being an independent, objective measure. A number of studies have revealed, for example, that different measurement systems can yield different results when applied to the same research problem. Where the same measurement approach is used in repeat studies, there is therefore the possibility that consistency in findings is in part due to consistency in the measurement system employed.

Clearly, the preceding limitations do not exist in isolation since each limitation can combine with every other limitation to produce a greater, overall deleterious impact on the reliability of research findings. Furthermore, these limitations operate at a more abstract level that is difficult or impossible to quantify in terms of confidence levels.

Conclusion

In seeking to develop objective knowledge about phenomena researchers must recognise that such knowledge is based on fallible measurements. However, this does not require a relativist epistemology as Hunt (1993) has argued. Whilst acknowledging a degree of uncertainty in quantitative findings, researchers can still gain reasonably accurate knowledge through measurement. The fact that this knowledge is fallible puts quantitative marketing research in a position no different from other empirical fields of study. The added problem for quantitative marketing research, and consumer research generally, is that the ability to form empirical generalisations is made more difficult by the presence of a greater number of more complex and interacting limitations.

This paper has discussed several such limitations; the Hawthorne effect, researcher-participant interaction, context dependency, conditioning, atypical research participants and differing measurement methods. Whilst the principles of falsifiability and replication provide elegant methods for addressing many of these limitations, they fail to provide complete solutions. This situation calls for increased acknowledgement of the imprecise nature of research findings and the methods by which these have been obtained in tandem with a more considered use of replication studies.

Firstly, a more diverse range of measurement methods should be used including a variety of researcher intrusion levels, differing data collection techniques and measurement types, scales and phraseology. Relatively speaking this serves to negate the potentially biasing influences of the Hawthorne effect, researcher-participant interaction and the same measurement devices across replication studies.

Secondly, greater recognition is required in the literature that consistency in research outcomes may not necessarily be an indicator of reliability, since this consistency may be partly an artefact of the repeated use of a particular measurement methodology. Only when a range of measurement methods have been applied over a period of time by different researchers can consistency in replication study findings considered to be reliable. This precludes claims for reliability and conclusiveness from single studies.

Thirdly, replication studies should employ multiplicity in sampling methods that go beyond simple demographics and which include a consideration of participants' frequency of involvement in previous studies to reduce biased samples arising from atypical and expert research participants. In this respect, the use of consumer research panels in replication studies should be particularly avoided.

Fourthly, different conclusions can be derived by different researchers from the same, repeat findings. That is, it is impossible to extrapolate observations unless these have already been placed them within an explanatory framework. In certain situations competing theories can be given equal credence by replication studies, instead of lending support for one theory over another. It is important to recognise, therefore, that consistency across replication study findings does not necessarily lead to a resolution between competing theories.

Finally, an accepted theory or group of theories provide the best explanation currently available to us, but may not hold in future when subjected to new types of criticism under different conditions. This is especially true for quantitative marketing research where results have a more limited lifetime and applicability because background contexts and consumer behaviours are in perpetual flux. Replication studies are therefore not only fundamental in testing the reliability of currently accepted theories but also play a significant role in confirming that such theories still hold as the external environment changes. Context dependency thus places an inherent limitation on the lifespan of quantitative marketing research findings.

References

- Anderson, P. F. (1983), Marketing, Scientific Progress, and Scientific Method, *Journal of Marketing*, 47, Fall, 18-31
- Barwise, T. P. (1995), Good Empirical Generalisations, *Marketing Science*, 14, 3, Part 2, G29-G35
- Bass, F. M. (1995), Empirical Generalisations and Marketing Science: A Personal View, *Marketing Science*, 14, 3, Part 2, G6-G19
- Bickart, B. A. (1993), Carryover and Backfire Effects on Marketing Research, *Journal of Marketing Research*, 30, February, pp52-62
- Chalmers, A. F.. (1982), *What is this thing called Science? second edition*. University of Queensland Press: St Lucia, Queensland.
- Christy, R. and Wood, M. (1999), Researching Possibilities in Marketing, *Qualitative Market Research: An International Journal*, Volume 2, Number 3, pp189-197
- Colet, E. (1999), Different Answers to the Same Business Questions: Part ii – Resolving this Dilemma, DS Star, The On-Line Executive Journal for Data-Intensive Decision Support, Volume 3, Number 32, 10th August
- Deutsch, D. (1998), *The Fabric of Reality*, The Penguin Press, London
- Ehrenberg, A.S.C. (1994), Theory or Well Based Results: Which Comes First? In Laurient, G.; Lilien, G; and Pras, B (eds) *Research Traditions in Marketing*, 79-108.

- Gendau, P, Esslemont, D, & D Day (1991). A Comparison of Two Versions of the Juster Scale Using Self-Completion Questionnaires. *Journal of the Market Research Society* 33, 257-263.
- Greenleaf, E. A. (1992), Improving Rating Scale Measures by Detecting and Correcting Bias Components in some Response Styles, *Journal of Marketing Research*, 29, May, p176-188
- Hubbard, R. and Armstrong, J. S. (1994), Replications and extensions in marketing: Rarely published but quite contrary, *International Journal of Research in Marketing*, 11, 233-248.
- Hunt, S. D. (1990), Truth in Marketing Theory and Research, *Journal of Marketing*, 54, July, 1-15. *Journal of Empirical Generalisations in Marketing Science*, Volume Three, 1998. Page 20
- Hunt, S. D. (1993), Objectivity in Marketing Theory and Research, *Journal of Marketing*, 57. April, 76-91.
- Johnson, M.D., Lehmann, D.R. and Horne, D.R. (1990), The Effects of Fatigue on Judgments of Inter-product Similarity, *International Journal of Research in Marketing*, Volume 7, Issue 1, August, pp35-43
- Juster, F (1966). Consumer Buying Intentions and Purchase Probability. Occasional Paper 99, National Bureau of Economic Research, Columbia University Press
- Lehmann, D. R., Gupta, S. and Steckel, J. H. (1998), *Marketing Research*, New York, Addison-Wesley
- Lindsay, R. Murray and Ehrenberg A.S.C. (1993), The Design of Replicated Studies, *The American Statistician*, 47, 3, 217-228.
- Luce and Tukey (2002), Online Surveys and Conjoint Analysis, www.sawtoothsoftware.com/productforms/ssolutions/ss16.shtml
- Morwitz, V. G., Johnson, E. and Schmittlein, D. (1993), Does Measuring Intent Change Behaviour? *Journal of Consumer Research*, 20, June, pp46-61
- Morwitz, V. G., and Gavan J. F. (2004), The Mere Measurement Effect: Why Does Measuring Intentions Change Actual Behaviour? *Journal of Consumer Psychology*, 14 (1&2), 64-74
- Pawson, R. and Tilley, N. (1997), *Realistic Evaluation*, Sage Publications, London
- Pineau, V. and Slotwiner, D. (2003), Probability Samples vs Volunteer Respondents in Internet Research: Defining Potential Effects on Data and Decision-making in Marketing Applications, Knowledge Networks Inc, <http://www.knowledgenetworks.com/insights/docs/Volunteer%20white%20paper%2011-19-03.pdf>
- Peter, J. P. (1992), Realism or Relativism for Marketing Theory and Research: A comment on Hunt's "Scientific Realism", *Journal of Marketing*, 56, April, 72-79
- Popper, K. R. (1935), *Logik der forschung*. Julius Springer Verlag.

- Robertshaw, G. S. and Marr, N. E. (2005), Voluntary Disclosures of Personal Information: A Quantitative and Qualitative Comparison of Differences between Contributors and Abstainers, *Journal of Empirical Generalisations in Marketing Science*, Article 9
- Robson, C. (1993), *Real World Research: a Resource for Social Scientists and Practitioners-Researchers*, Basil Blackwell, Oxford
- Simmons, C. J., Bickart, B. A. and Lynch, J. G. (1993), Capturing and Creating Public Opinion in Survey Research, *Journal of Consumer Research*, 20, September, pp316-329
- Szmigin, I. and Foxall, G. (2000), Interpretive Consumer Research: How Far Have We Come? *Qualitative Market Research: An International Journal*, Volume 3, Number 4, pp187-197
- Tweney, D. (1998), The Consumer Battle over Online Information Privacy has just Begun, *InfoWorld*, Volume 20, Number 25, 22nd June, p66
- Viswanathan, M., Sudman, S. and Jordan, M. (2004), Maximum Versus Meaningful Discrimination in Scale Response: Implications for Validity of Measurement of Consumer Perceptions about Products, *Journal of Business Research*, Volume 57, Issue 2, pp108-124
- Wright, M. and Kearns, Z. (1998), Progress in Marketing Knowledge, *Journal of Empirical Generalisations in Marketing Science*, Volume 3, pp1-21