AN EMPIRICAL STUDY OF THE WEIGHT OF SURVEY EVIDENCE IN DECEPTIVE ADVERTISING LITIGATION

ROBERT LANGTON

Lecturer in Commercial Law, University of Adelaide

LINDSAY TROTMAN

Associate Professor in Business Law, Massey University

Issues often arise in intellectual property and Fair Trading litigation in which the fundamental matters in dispute concern the perceptions or behaviour of a large number of people – typically consumers. One way of gathering evidence on this type of issue is by commissioning a market survey. However, courts in Australia and New Zealand have shown a marked reluctance to place much reliance on market surveys in arriving at their findings of fact. In some cases, market survey evidence has been held inadmissible on the basis that it offends the rule against hearsay evidence. In virtually all the remaining cases, it has been found to be of little probative value due to flaws in the design or execution of the survey.

This paper argues that it is possible to design market surveys in such a way that the evidence will be both admissible and likely to be accorded substantial weight. The paper is divided into three sections. The first section briefly discusses the current status of the hearsay objection with respect to market surveys. The second section argues that the weight accorded to survey evidence is likely to depend on the way in which three fundamental issues of survey design are addressed. These issues relate to the survey population, the wording of the survey questions and the context in which the questions are asked. These three issues are examined for the case of a survey designed to prove the claims implied by a television advertisement in the context of litigation under section 9 of the *Fair Trading Act* 1986. The paper describes a type of survey that would, we believe, be regarded as providing highly probative evidence on this issue.

The third section of the paper describes an attempt to assess empirically the weight that would be given to evidence produced by this type of survey. This attempt involves an experiment in which 41 barristers were asked to predict the outcome of nine hypothetical court cases some of which include evidence of such surveys. The results of the experiment suggest that carefully designed surveys can produce highly probative evidence in appropriate cases.

Before turning our attention to the problems of designing probative surveys, we briefly examine the current status of the hearsay objection in New Zealand and Australia.

I. ADMISSIBILITY OF SURVEY EVIDENCE

The traditional objection to the admissibility of survey evidence is that it amounts to hearsay evidence since it is based on the out-of-court statements of people who are not intended to be called as witnesses. Judges in New Zealand and Australia have shown some divergence in their attitude to this objection.

The first reported case dealing with survey evidence in New Zealand is the decision of Mahon J in *Customglass Boats Ltd v Salthouse Bros.*[1] The case concerned an action for "passing off" relating to the name "Cavalier" as applied to yachts and, as is usual in such cases, the plaintiff sought to prove that the name was associated in the minds of the yachting fraternity with yachts of its own manufacture. In an effort to do this, the plaintiff adduced evidence of a professionally conducted market survey. Although no objection was taken to the admission of the survey evidence at the trial, Mr Justice Mahon undertook an extensive review of the approaches to this issue in Canada, the United States and the United Kingdom.

His Honour noted that whether or not evidence amounts to hearsay depends upon the inference the court is being asked to draw from the evidence. In many instances, including the case before him, he felt that survey evidence would not amount to hearsay since it was not offered to prove the truth of the statements contained therein. Even if the evidence were categorised as hearsay, Mahon J would have been prepared to admit it under a well established exception to the hearsay rule as evidence tending to prove the public state of mind.

His Honour noted that in the United Kingdom, survey evidence had been accepted in trademark and patent litigation without any overt discussion as to the basis of its admissibility. He commented that:

> Although the basis of admissibility does not appear to be overtly founded upon anything except established practice and procedure under the trademark and patent legislation, I can for myself see no objection to the classification of such evidence as either proving a public state of mind on a specific question, which is an acknowledged exception to the hearsay rule, or as proving an external fact, namely, that a designated opinion is held by the public or a class of the public, this not being a matter of hearsay at all.

This analysis of survey evidence as either not being hearsay or else falling within an established exception to the hearsay rule has been followed in subsequent cases in New Zealand.[2]

In the first reported decision in Australia, Bray C J described the survey evidence that had been offered at the trial as "not only hearsay, but double, perhaps even treble, hearsay".[3] With one limited exception, subsequent Australian cases have not seen any exception to the hearsay rule as permitting the admission of survey evidence.[4] The exception arose in *Shoshana Pty Ltd v 10th Cantanae Pty Ltd*,[5] where Burchett J was prepared to regard a survey commissioned for purposes unconnected with the litigation as falling within the state of mind exception to the hearsay rule. However, the fact that the survey was not commissioned specifically for the litigation was a crucial element in Burchett J's approach. He stated:

1 [1976] 1 NZLR 36.

2 Mahon J's approach has been approved by the High Court of New Zealand: *Klissers Farmhouse Bakeries Ltd v Harvest Bakeries Ltd* [1985] 2 NZLR 129, *Auckland Regional Authority v Mutual Rental Cars* [1988] 2 NZBLC 103,041 and by the New Zealand Court of Appeal: *Bevan Investments Ltd v Blackball and Struthers (No. 2)* [1978] 2 NZLR 97. It has also been relied on in a "passing off" case in the United Kingdom: *Lego Systems v Lego M Lemelshtrich Ltd* (1983) FSR 155. In *Noel Leeming Television Ltd v Noel's Appliance Centre Ltd* (1985) 5 IPR 249, Holland J admitted survey evidence on the apparently separate ground that it was offered as the basis for expert opinion evidence.

3 *Hoban's Glynde Pty Ltd v Firle Hotel Pty Ltd* (1973) 4 SASR 504 which was an appeal from the decision of the South Australian Licensing Court.

4 *Mobil Oil Corp v The Registrar of Trademarks* (1983) 51 ALR 735; *McDonald's System of Australia Pty Ltd v McWilliams Wines Pty Ltd* (1979) ATPR 40-136, (1979) AR 236.

5 (1988) 18 FCR 285.

If he had conducted the survey specifically for the purposes of the applicant's case, his evidence would not have been admissible.

Thus prior to *Arnotts v Trade Practices Commission*,[6] the hearsay objection remained a significant obstacle to the introduction of market survey evidence in Australia. The *Arnotts* case arose out of a merger between two biscuit manufacturers in Australia. It was alleged by the Trade Practices Commission that the merger would lead to Arnotts acquiring a dominant position in the biscuit market in contravention of s 50(1)(b) of the *Trade Practice Act* 1974 (Cth). Arnotts contended that the relevant market for the purposes of the competition analysis required under that section was not the biscuit market but the much larger "snack food" market in which Arnotts had only a small market share. In order to demonstrate that a wide variety of foods such as cakes, fruit, and muesli bars were substitutable for biscuits, Arnotts commissioned a survey by a professional market research firm. The survey involved interviews with 1200 members of the public carried out by 120 interviewers as part of a regular survey of consumer attitudes on behalf of a number of commercial clients.

At the trial, Beaumont J initially refused to admit the survey evidence in the absence of testimony from any of the interviewers. Arnotts then called three of the interviewers who each produced their own questionnaires and testified as to the manner in which the interviews had been conducted. Mr Justice Beaumont then exercised a discretion to admit the rest of the questionnaires on the basis of rule O 33 r 3 of the *Federal Court Rules* which allows the Court to dispense with the rules of evidence where compliance would involve unnecessary or unreasonable expense or delay. Having admitted the evidence, Beaumont J found it of little weight mainly because, in his view, the questions were ambiguous and the interviewers were inadequately trained.

On appeal, the Full Federal Court endorsed Beaumont J's view that the evidence was worthy of little weight. However the Full Court took the opportunity to emphasise that the question of admissibility of survey evidence should not involve an analysis of whether the evidence amounts to hearsay. This view was based partly on the fact that the Court has a discretion to admit hearsay evidence and partly on the view that surveys sometimes provide the most reliable form of evidence of the matters sought to be proved. In the words of the Full Court:

However, it is not very profitable – at least in this Court – to spend time in determining whether a particular survey is hearsay evidence. Even if it is, the Court will have discretion under O.33 r.3 to permit the evidence to be adduced. To call the persons who responded to the survey will almost always result in appreciable expense and delay. Given the existence of the discretion, it seems more sensible to concentrate attention upon the necessity for, and the reliability of, the survey evidence; rather than to worry about its compliance with rules regarding hearsay evidence which were developed before this type of problem arose.[7]

The effect of the *Arnotts* decision is that the issue of admissibility of survey evidence is now approached the same way in both Australia and New Zealand. This approach is to ignore technical objections based on the hearsay rule and to admit the evidence provided the survey can be shown to have been properly conducted. The approach was summarised by the Full Court as follows:

6 (1990) ATPR 41-061.
7 (1990) ATPR 41-061 at 51,807.

But, in a civil case in which a market survey may cast light on relevant issues, it is desirable in principle to admit into evidence a report of a professionally conducted survey, upon proof that it has been satisfactorily conducted using relevant and unambiguous questions; and without requiring evidence from each of the participants.[8]

However, the Full Court was also careful to point out that the decision to admit survey evidence left open the issue of its probative value in the particular case. It stated:

We have already indicated our opinion that Australian law should follow the American lead in acknowledging that market survey evidence may play a useful role in cases such as the present. In doing so, we do not mean to suggest that survey evidence will always, or even usually, be decisive. It will merely be one element in the overall picture, its importance varying from one case to another.[9]

While emphasising that the importance of survey evidence would vary from one case to another, the Full Court appeared to recognise the possibility that in some cases, survey evidence could be decisive. It is this possibility that inspired the research described below.

II. Deceptive Advertising Litigation

This research aimed to investigate the possibility that survey evidence might prove decisive in litigation concerning the allegation that an advertisement constitutes "misleading or deceptive conduct" in contravention of section 9 of the *Fair Trading Act* 1986 or equivalent legislation.[10] A variety of issues can arise in such litigation. These include:

- whether the advertisement conveys the alleged representation
- whether the representation amounts to puffery
- whether the representation is true
- whether the representation would cause loss to any party
- whether interim relief should be granted.

The scope of this paper is restricted to a consideration of the role that survey evidence might play with respect to the first of these issues – whether the advertisement conveys the representation alleged by the plaintiff.[11] Further, it is limited to cases where the representation is not explicitly stated in the advertisement but is alleged to be implied by it.

The question of whether an advertisement conveys an implied representation is one that seems to arise most often with respect to television commercials. Professor Preston has noted that most of the cases in which survey evidence has been introduced in the United States have concerned television commercials.[12] His explanation for this is as follows:

While nothing in the record discusses that point, an easy explanation is that television must produce most of the types of conveyed messages that require extrinsic evidence. Explicit meanings and obvious implied meanings require little or no such evidence, while non-obvious

8 (1990) ATPR 41-061 at 51,808.
9 (1990) ATPR 41-061 at 51,810.
10 Section 9 states: "No person shall, in trade, engage in conduct that is misleading or deceptive or is likely to mislead or deceive". For present purposes, the section is identical to section 52(1) of the *Australian Trade Practices Act* 1974 (Cth). Equivalent provisions are also to be found in the Fair Trading legislation passed by each State legislature in Australia.
11 The terms "plaintiff" and "defendant" are used throughout the paper to refer to the parties to the litigation although most cases concern applications for injunctions.
12 Preston I L "Extrinsic Evidence in Federal Trade Commission Deceptiveness Cases" [1987] *Columbia Business Law Review* 633.

implied meanings, which do, seem to occur far more often in television than in print or radio advertisements.[13]

Given the American experience, this research focused on the probative value of surveys designed to prove the implied claims in a television advertisement.

III. The Design of a Probative Survey

In broad terms, the probative value of any evidence depends on the extent to which it is seen as directly addressing one of the factual matters in dispute and the extent to which it is seen as reliable. Both of these matters need to be considered in any attempt to design a probative survey for the type of issue identified above. The design of any survey requires addressing three fundamental questions:

i. Who should be surveyed?

ii. In what context should the questions be asked?

iii. What type of questions should be asked?

Each of these questions is discussed in turn below.

i. Who should be surveyed?

One of the most commonly cited descriptions of the steps the Court will follow in deciding section 9 cases suggests that the first step is for the Court to define the section of the public which is likely to be affected by the conduct complained of.[14] Despite this statement, it is not clear who the Courts would see as the "class of persons likely to be affected" by a television advertisement.

One approach is to define this class as the general public on the basis that virtually everyone has some chance of being exposed to a television commercial. A second approach is to argue that only those people who actually see an advertisement can be affected by it. A third approach is to argue that only people who are potential purchasers of the advertised product should be included, since they are the only ones whose purchase behaviour could be affected by the commercial.[15] A fourth approach might be to say that the most relevant group are those considered by the advertiser to be within its target market.[16] In cases of "niche marketing", this might be a much narrower group of people than either those who actually see the advertisement or those who are potential purchasers of the product class.

In the absence of clear guidelines from the Courts, a practical solution to this problem is to define the population of interest as the general public. This approach reduces the risk of the Court rejecting the entire survey as irrelevant to its inquiry. It also allows the results of the survey to be broken down into any sub-groups the Court deems appropriate. In order to allow for this possibility, it would be advisable to use a reasonably large sample of about 1000 subjects.

13 Ibid at 654. The term "extrinsic evidence" is used to refer to all types of evidence other than that provided by the advertisement itself, which is referred to as "intrinsic evidence".

14 *Taco Company of Australia Pty Ltd* v *Taco Bell Pty Ltd* (1982) ATPR 40-303 at 43,749.

15 Healy and Terry comment that "Representations made in television advertisements, for example, are generally made to all the world. This proposition stands, of course, unless the goods or services offered are not likely to be bought by one sector of the community" Healy D and Terry A, *Misleading and Deceptive Conduct*, CCH Australia, Sydney, 1991 para 426.

16 This group would, in some cases, be specified in the company's marketing plan or advertising brief.

ii. In what context should the questions be asked?

Tests of television commercials normally adopt one of two basic formats. The main differences between these are, first, whether the audience views the commercial under artificial or natural conditions and, second, whether the audience answers questions immediately after seeing the commercial or some time later.

In tests using natural viewing conditions, subjects watch the commercial at home in the normal way without any prior contact with the researchers. Within twenty four hours of the broadcast of the commercial, they are contacted by telephone and asked if they remember seeing the commercial. If they say they do, they are asked a series of questions about the contents of the commercial.

Tests using artificial viewing conditions are conducted in a theatre. The audience for each night is recruited to attend a preview of a televison programme in return for answering some questions about it. Recruitment procedures are designed to produce a representative sample from the city in which the test is conducted. Subjects are shown a television programme, and then five commercials. Immediately after each commercial, subjects fill out their responses to a page of questions about that commercial.

In the United States, judicial support can be found for each approach. However, tests using artificial viewing conditions are generally seen as more directly addressing the issue of the implied messages in a television commercial. This attitude is reflected in the following passage from the judgment in *American Home Products v Johnson & Johnson*:[17]

> Initially we were disposed to think that the G&R method of testing would be better, because the consumer saw the commercial in the normal context – viewing TV at home. However, the verbatims caused us to seriously doubt the reliability of this method, at least when the need was to find out how the consumer interpreted the advertisement's language, i.e. what the consumer took the message to be from the particular language used.
> The G&R technique ... revealed ... how an ad for a familiar product seemed to leave its impact by triggering the viewer's past association with the product as well as by hooking the viewer's interest in the particular new advertising copy being aired. But we are not confident that the G&R technique really tested what message the consumer took from the language used as opposed to the broader issue of what mental processes viewers went through ...
> We think that the format of the ASI [theatre] presentation was likely to cause the audience to attend to the programs and ads more closely and thus more accurately reflect what the average consumer who heard and saw the ad took the message(s) to be, although it probably less accurately reflects the real impact ... on the viewer.[18]

Tests using artificial viewing conditions and immediate questioning are seen as directly addressing the precise issue before the Court. In addition, they are probably seen as providing more trustworthy evidence. This is because tests conducted in a theatre obviate the need for interviewers. The evidence produced at the trial comprises the questionnaires filled in by the survey subjects. By contrast, tests in which subjects view the commercial at home demand the use of a group of telephone interviewers. Thus, the evidence produced at the trial is the interviewers' records of a number of telephone conversations.

The Courts have shown much concern over the potential for bias offered by inadequately trained and supervised interviewers.[19] Bias could be

17 *American Home Products v Johnson & Johnson* 436 F Supp 785 (SDNY 1977). This case was decided under s 43(a) of the *Lanham Act* 1946 which allows a person to bring a civil action against a competitor's advertising on the basis that it is misleading.

18 *American Home Products v Johnson & Johnson*, above at 794.

19 See the comments of French J in *State Government Insurance Corporation v Government*

introduced in the way interviewers read the questions and in the way they record the answers. This concern was one of the three reasons the Court held the survey evidence to be worthy of little weight in the *Arnotts* case.

In Australia, the artificial nature of the survey situation has sometimes been given as a ground for dismissing survey evidence. In *Interlego AG v Croner Trading Pty Ltd*:[20]

> I think the hypothetical situation which was created was so artificial as to make it quite a dangerous guide to what the reactions of actual shoppers purchasing Lego or Tyco products from the shelf of a supermarket or a toy shop might have been.[21]

Similarly, in *State Government Insurance Corporation v Government Insurance Office of NSW*,[22] French J stated:

> The utility of the results is also limited by the fact that the survey sought to elicit responses and identify associations in a context removed from that of consumer decision-making connected with the acquisition of insurance and financial services. The associations tested by the first question arose in an artificial situation from which few, if any, useful inferences can be drawn about actual consumer reactions and decision-making.[23]

However, these statements were made in the context of cases concerning similar names or products and thus are of doubtful relevance to surveys designed to prove the claims implied by a television advertisement. On this issue, it is submitted that tests using artificial viewing conditions would probably be found more probative than those using natural viewing conditions.

iii. What type of questions should be asked?

Tests using artificial viewing conditions can include both open-ended and forced-choice questions. Open-ended questions are those where subjects are asked to respond in their own words such as "What do you think the advertisement is saying?". Forced-choice questions ask subjects to select their response from a range of predetermined options.

Two common problems with open-ended questions have been identified in Federal Trade Commission (FTC) decisions in the United States. The first is that categorising the responses is an inherently subjective and somewhat arbitrary process.[24] The second problem is that open-ended questions tend to elicit only the central theme of the commercial which may not be the focus of the litigation.

In the *Wonder Bread* case,[25] viewers were asked "What is the most important thing that the commercial told you about Wonder Bread?" Despite the fact that only 50 out of 789 responses suggested that Wonder Bread induces remarkable growth, the Commission held that the survey

Insurance Office of NSW (1991) ATPR 41-110 at 52,697.

20 (1991) ATPR 41-124, Sheppard J felt unable to place any weight on the survey evidence adduced in that case largely because, in his view, the circumstances in which the questions were asked were too artificial. Subjects were approached leaving a supermarket. They were shown a sample of the defendant's product and then asked who they thought the manufacturer was. Sheppard J refused to place any weight on this evidence.

21 Ibid at 52,835.

22 (1991) ATPR 41-110.

23 Ibid at 52,698.

24 For a dramatic illustration of the arbitrariness of the process see the criticisms of French J in *State Government Insurance Corporation v Government Insurance Office of NSW* (1991) ATPR 41-110 at 52,692. The FTC has usually demanded that all the responses be available for its perusal and has sometimes undertaken its own categorising: *Bristol-Myers* 102 FTC 21 (1983).

25 *ITT Continental Baking Co* 83 FTC 865 (1973).

was not inconsistent with their conclusion that this claim was implied. In the Commission's view:

> The questions were not designed and would not be likely to elicit consumers' perception of the latent or implied messages contained in the advertising such as those challenged in the complaint. Rather, the questions asked were designed to and usually only elicited the interviewee's recall of the explicit message projected by the advertisement.[26]

Similarly, in *American Home Products,*[27] the Commission stated:

> It is accepted in marketing research that an open-ended question is not representative of everything stored in subjects' minds ... Open-ended questions lead most subjects to play back only one theme or point ... Respondents will play back the dominant theme or primary impression and having done that, will stop.[28]

Some of the problems associated with open-ended questions can be overcome by using forced-choice questions. Such questions can be designed to correspond to the exact allegation under investigation and the problem of categorising individual answers is avoided. However, forced-choice questions have been criticised as encouraging guessing[29] and as being ambiguous.[30] In addition, the way in which the response options are formulated has often been seen to bias the results.

In *Rhodes Pharmacal,*[31] subjects were asked whether certain advertisements meant that Imdrin would provide relief from the pain of arthritis and rheumatism or whether they meant that it would provide a cure for these ailments. Only 9% of subjects indicated the latter option which corresponded to the FTC's allegation, but the Commission held that:

> A survey implying that the advertisements either represented a cure or represented pain relief was improperly designed to determine if they did not represent both.[32]

A similar criticism was made of the range of response options in the Firestone case. Consumers were asked which of four statements came closest to what they thought an advertisement said. Fifteen percent indicated the option corresponding to the allegation. However, the Commission felt that the structure of the question prevented subjects indicating that they had perceived more than one of the statements.[33]

On the issue of whether forced-choice questions are seen to provide more probative evidence than open-ended questions, Professor Preston comments:

> ... A definitive conclusion is difficult, however, because the evident superiority of well-constructed forced-choice questions is offset by their scarcity, owing both to the need to customize them to the specific situation (with consequent extra cost), and to the problems of constructing

26 Ibid at 977.
27 *American Home Prods* 98 FTC 136.
28 Ibid at 415.
29 See the detailed discussion of this issue in *State Government Insurance Corporation v Government Insurance Office of NSW* (1991) ATPR 41-110 at 52,692.
30 The ambiguity of the questions was one of the main reasons the survey evidence was given little weight in the *Arnotts* case: *Arnotts Ltd* v *Trade Practices Commission* (1990) ATPR 41-061 at 51,809.
31 *Rhodes Pharmacal Co* 49 FTC 263.
32 Ibid at 283. A similar criticism was made of the range of response options in the *Firestone* case. Consumers were asked which of four statements came closest to what they thought an advertisement said. Fifteen percent indicated the option corresponding to the allegation. However, the Commission felt that the structure of the question prevented subjects indicating that they had perceived more than one of the statements. *Firestone, Tire & Rubber Co* 81 FTC 398 at 454.
33 *Firestone Tire & Rubber Co* 81 FTC 398 at 454.

them adequately. Thus, although good forced-choice questions may be better, they often are not available.[34]

One type of forced-choice question that avoids the need for customisation is a question of the form: "Is the advertisement claiming that X ?" where X corresponds to each claim alleged to be implied by the advertisement. A simple range of response options, such as "Yes", "No" or "Don't Know", could be provided for such a question. A possible objection to this type of question – that it is a leading question – is discussed in the final section of the paper.

IV. Audience Reaction Tests

Based on the above discussion, it is tentatively concluded that the most probative type of survey with respect to the issue of whether a television commercial conveys an implied claim would be one that has the following features:

- The subjects for the survey are approximately 1000 people who are selected so to be representative of the general population in the region where the commercial is broadcast.
- Subjects are shown the advertisement in a theatre and asked to respond to written questions immediately afterwards.
- The questions asked are of the form "Is the advertisement claiming that X ?" where X corresponds to each representation alleged to be implied by the advertisement.
- Subjects are asked to select one of three possible options in response to each question: "Yes", "No" or "Don't know".

In the remainder of this paper, the term "audience reaction test" is used to refer to a survey having these features. Analysis of such a survey would involve simply calculating the percentage of subjects who selected each of the three possible response options for a particular question. The term "ART score" is used in the remainder of this paper to refer to these percentages. This acronym stands for **A**udience **R**eaction **T**est score.

Using this terminology, it is tentatively concluded that ART scores – the results of audience reaction tests – would provide the most probative form of survey evidence with respect to the issue of the implied claims conveyed by a television commercial. This tentative conclusion raises two questions:

- How much weight would be accorded to evidence of ART scores in litigation under section 9 of the *Fair Trading Act* 1986?
- Is our tentative conclusion correct or would some other type of survey provide more probative evidence?

Both of these questions are empirical questions. The first is concerned with the impact of ART scores compared to all other (non-survey) types of evidence which could be adduced on the same issue. The second question is concerned with the impact of ART scores compared to other types of survey evidence which could be adduced. The remainder of the paper describes an empirical method of addressing these two questions.

34 Preston I L, "Extrinsic Evidence in Federal Trade Commission Deceptiveness Cases" [1987] *Columbia Business Law Review* 633 at 667.

V. An Empirical Method of Estimating the Probative Value of ART Scores

The method devised with respect to the first question involved an experiment in which expert barristers were asked to predict the outcome of hypothetical court cases.[35] Forty-five barristers with experience in trade practices litigation were invited to participate.[36]

Participants watched a video tape containing nine hypothetical court cases each involving the allegation that a television advertisement constituted "misleading or deceptive conduct". After each case, they were asked to estimate the likelihood of the advertisement being found to contravene section 9 of the *Fair Trading Act* on the basis of certain assumptions.

Two standard assumptions were provided for every case. The first specified the claim alleged to be conveyed by the advertisement. The second was an admission by the defendant that, if this claim was made, it would be false. The assumptions were designed so that the only matter in dispute was whether the advertisement conveyed the implied claim alleged by the plaintiff.

For the first three cases, only the standard assumptions were included. The first case (British Airways) was used to familiarise participants with the way in which the cases would be presented. This involved first showing the commercial, then stating the two standard assumptions, then showing the commercial again and finally inviting participants to estimate, on the assumptions provided, the probability that the advertisement would be found to contravene section 9 of the *Fair Trading Act*. In estimating this probability, participants were asked to select a number between zero and ten according to the following scale.

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

Following the first three cases, the way in which audience reaction tests are conducted was explained to participants.

Next, three cases were presented in which, in addition to the two standard assumptions, a third assumption was included to the effect that the judge had admitted evidence of audience reaction tests. The first of these cases (Reach toothbrushes) was used to familiarise participants with the way in which evidence of audience reaction tests would be presented. After the advertisement had been shown and the two standard assumptions had been presented, a third assumption was provided which specified the wording of the survey question asked and the percentage of survey subjects

35 The barristers selected were all those who had appeared as counsel in at least one case concerning the application of section 9. Given the short time in which the section had been in force in New Zealand, only 45 barristers qualified for inclusion in the study.

36 Of these 41 agreed to participate in the study.

who selected each of the three response options. Then the commercial was shown again, and participants were asked for their prediction of the probable outcome in the usual way.

The ART scores presented in these three cases were constructed by the researchers for the purposes of the experiment. The ART scores were one of the following possible scores:[37]

	Yes	*No*	*Don't know*
"Low score"	5%	65%	30%
"Medium score"	35%	35%	30%
"High score"	65%	5%	30%

For the last three cases, participants were asked to estimate the likely outcome twice, under two different scenarios. In the first scenario only the standard assumptions were included. Once the initial estimates were obtained, a third assumption giving evidence of ART scores was introduced and a second estimate requested.

The 41 participants were randomly assigned to one of three groups (Tapes A, B or C). The video tapes watched by members of each group were identical except in one important respect. The order of the cases was different for each group. The order of the cases on each tape and the assumptions provided for each case are shown below.

Assumptions	*Tape A*	*Tape B*	*Tape C*
1. Standard	British Air	British Air	British Air
2. Standard	Simpson	Mitsubishi	Simpson
3. Standard	Honda	Pine'o Cleen	Honda
4. Medium ART	Reach	Reach	Reach
5. Low ART	Mitsubishi	Simpson	Pine 'o Cleen
6. High ART	Pine'o Cleen	Honda	Mitsubishi
7. A Standard B High ART	Panadol	Palmolive	Toyota
8. A Standard B Medium ART	Palmolive	Toyota	Panadol
9. A Standard B Low ART	Toyota	Panadol	Palmolive

The reason for varying the order of the cases was to enable a comparison to be made between those predictions which were based on evidence of ART scores and those made without such evidence. For example, in the Pine'o Cleen case, the predictions of barristers in Group B were based on only the two standard assumptions, those in Group A, on evidence of a high ART score and those in Group C on evidence of a low ART score.

After predicting the outcome of the nine hypothetical cases, participants were asked to assess whether a number of changes to the design of the survey would affect the probative value of the evidence. This part of the

37 The labels "low", "medium" and "high" were not included on the tapes.

study aimed to address the second question which was whether some other type of survey might provide more probative evidence than ART scores.

VI. Results of the Study

The results for four of the first six hypothetical cases are shown on the following pages in the form of bar charts.[38]

In these charts, the horizontal axis corresponds to the probability scale used by participants in making their estimates of the likelihood of the plaintiff succeeding. According to this scale, an estimate of zero represents the view that there is "no chance or almost no chance" of the advertisement being found to constitute misleading or deceptive conduct while an estimate of ten corresponds to the view that it is "certain or practically certain" that the Court would find the advertisement deceptive.

The vertical axis on each chart shows the number of barristers who selected each point on the scale.

For each case, the estimates of each group are shown separately. This is because the assumptions on which the estimates were based depended on the tape viewed. Where the predictions were based on only the two standard assumptions, this is denoted by the phrase "No ART score". Where ART score evidence was included, the value of the ART score is indicated.

As can be seen from the results of these four cases, [39] there was a large variation between barristers' estimates of the probability of the Court finding each advertisement deceptive. This was so even for estimates based on identical assumptions. It was realized that this variation would mean that the influence of ART score evidence would be found to be statistically significant only where the evidence had a dramatic effect on the estimates.[40]

In an attempt to distinguish the variation due to evidence of ART scores from the variation due to differences between barristers, the last three cases[41] required each barrister to predict the outcome both before and after being informed of the ART score evidence.

All participants made their initial estimates based only on the standard assumptions. They were then provided with the ART score evidence and asked to make a second estimate. As with the first four cases, the value of ART score depended on the tape viewed.

The results for these last three cases are presented in a slightly different format. The horizontal axis of the bar charts shows the *difference* between the first and second estimates of each participant. If a participant did not change his or her estimate after ART score evidence was introduced, this is shown as zero. If the second estimate was two points lower on the probability scale than the first, this is shown as minus two.

38 As mentioned above, the first and fourth cases were not true test cases, but were included for demonstration purposes only.

39 See below, pp 159-162.

40 This, in fact, proved to be the case. Using a single factor analysis of variance test on all the estimates for the four cases, the ART score evidence was found to have a statistically significant impact on the estimates ($p < 0.01$).

41 See below, pp 163-165.

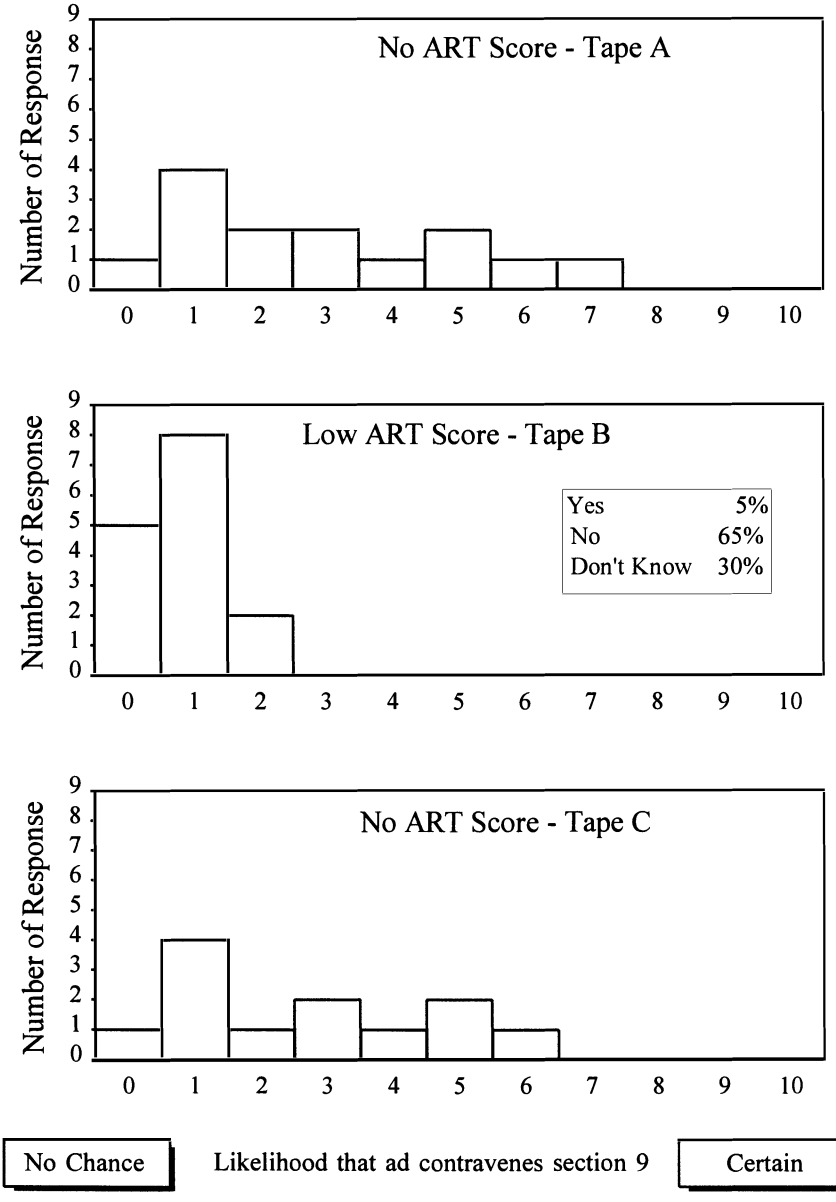
SIMPSON CASE

In this example, Tapes A and C contained only the two standard assumptions.

The first assumption alleged that the advertisement conveyed the representation that *the Simpson Acquarius is the only washing machine of its size where the bowl is top suspended.* The second assumption admitted this representation would be false.

Tape B contained an additional assumption providing evidence of a Low ART score.

The predictions obtained from barristers who watched each tape are shown below.

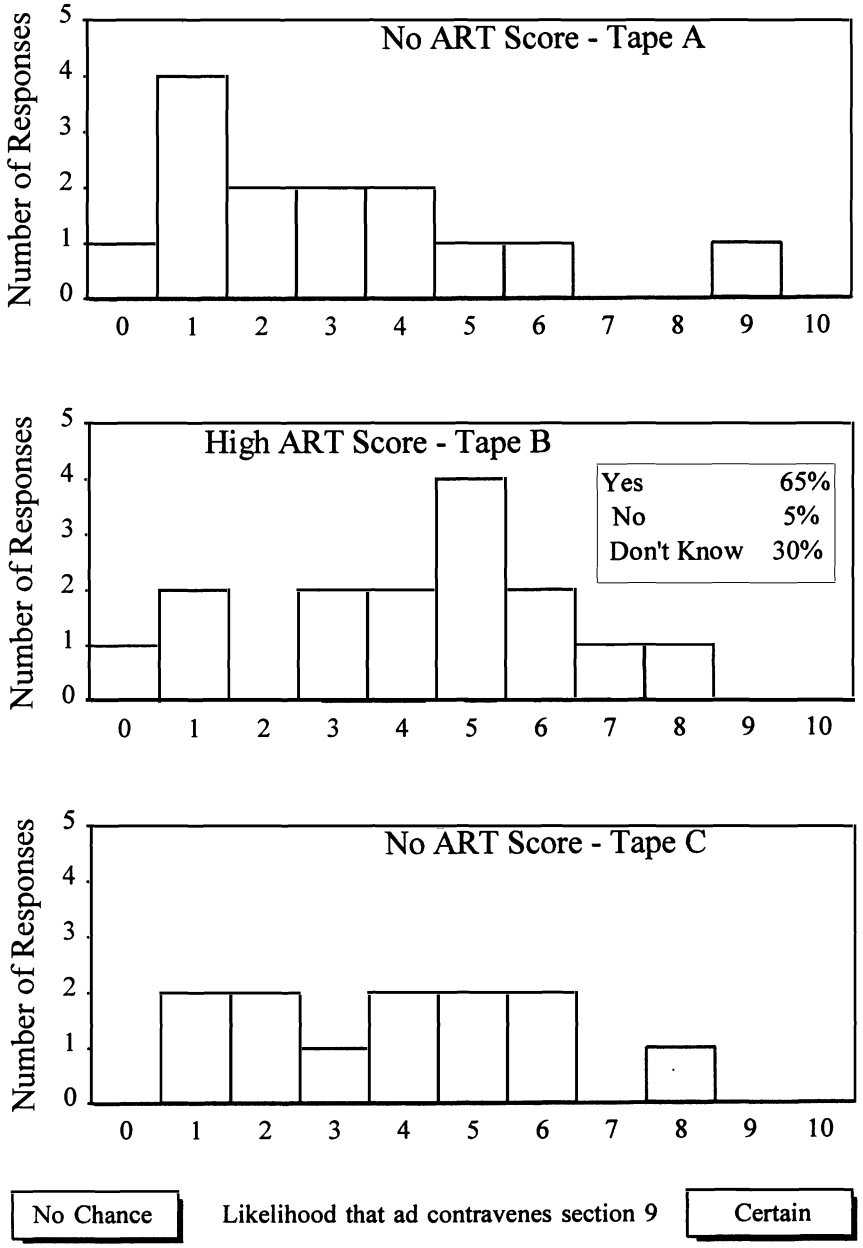


HONDA CASE

In this example, Tapes A and C contained only the two standard assumptions.

The first assumption alleged that the advertisement conveyed the representation that *the Honda Concerto is the only hatchback in which the noise level does not exceed 65 decibels when driven at 100 kms/hr*. The second assumption admitted this representation would be false.

Tape B contained evidence of a High ART score.

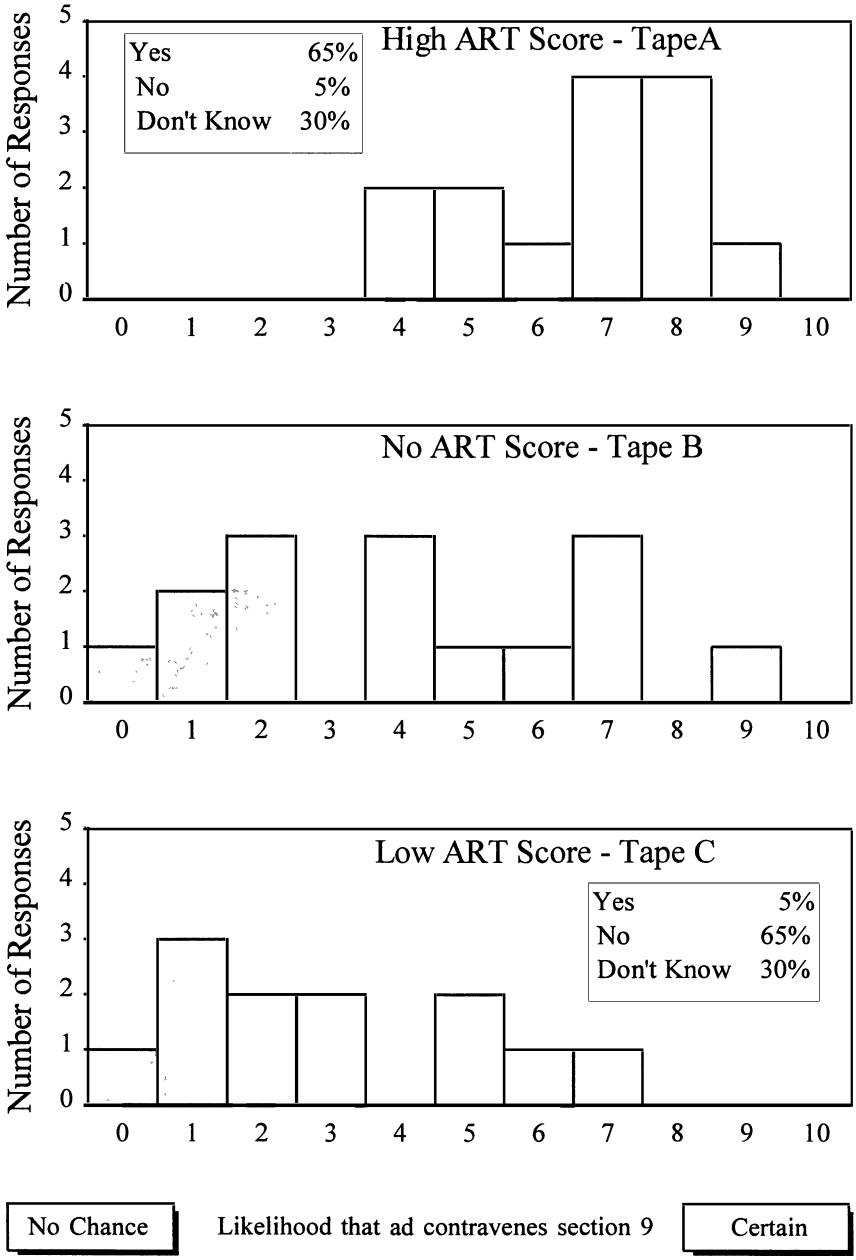


PINE'O CLEEN CASE

In this example, Tape B contained only the two standard assumptions.

The first assumption alleged that the advertisement conveyed the representation that *more Pine'o Cleen is sold in a year than any other brand of disinfectant.* The second assumption admitted this representation would be false.

Tape A contained evidence of a High ART score, while Tape C included evidence of a Low ART score.

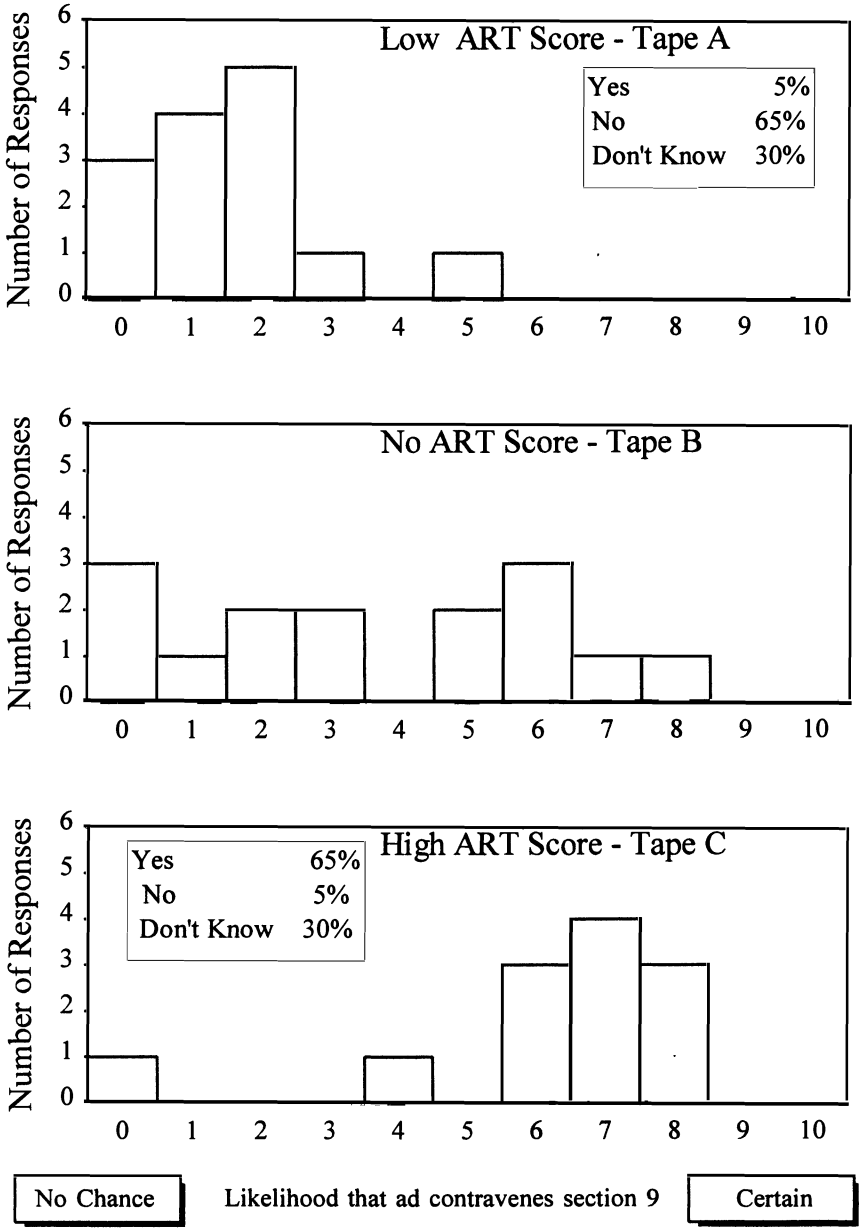


MITSUBISHI CASE

In this example, Tape B contained only the two standard assumptions.

The first assumption alleged that the advertisement conveyed the representation that *Mitsubishi has won some of the major international four wheel drive rallies in the last three years.* The second assumption admitted this representation would be false.

Tape A contained evidence of a Low ART score, while Tape C included evidence of a High ART score.

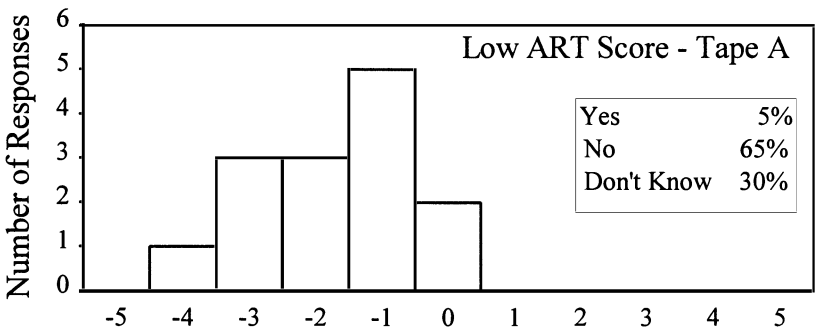


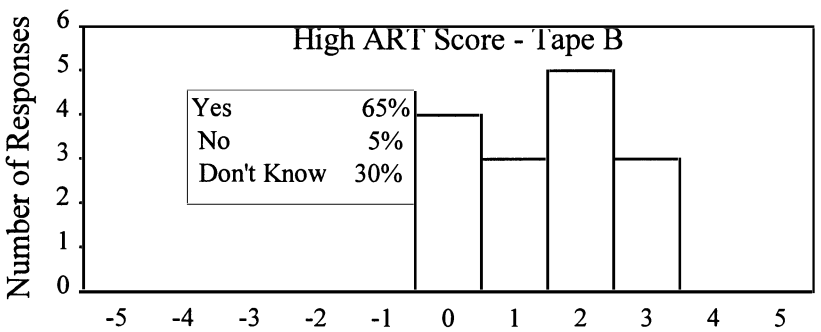
PALMOLIVE CASE

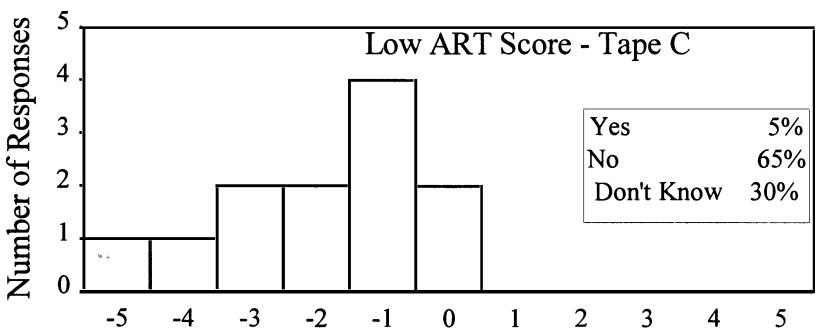
In this example it was alleged that the advertisement conveyed the representation that *soaking one's hands in undiluted Palmolive is beneficial for the skin of most people.*

All participants made their first estimates based only on the standard assumptions.

The ART score evidence that formed the basis of the second estimate is shown in the bar charts below.







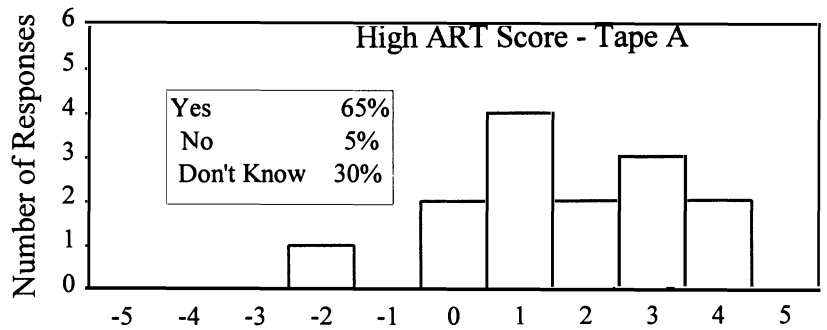
Less Chance — Likelihood that ad contravenes section 9 — More Chance

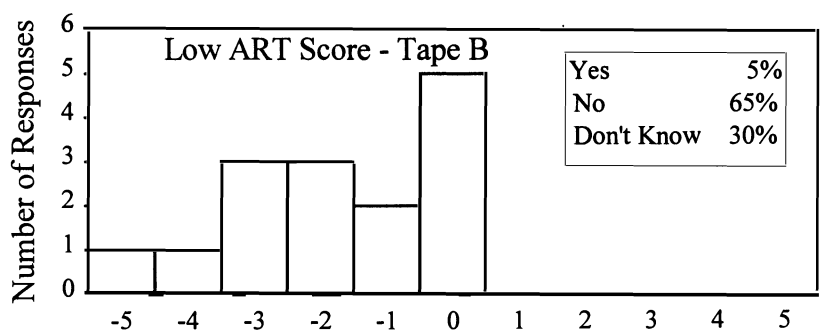
PANADOL CASE

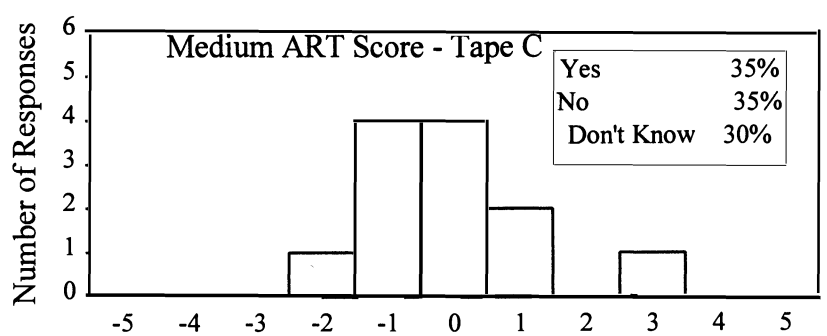
In this example it was alleged that the advertisement conveyed the representation that *Clinical tests show that Panadol is more gentle on the stomach than any other brand of pain reliever.*

All participants made their first estimates based only on the standard assumptions.

The ART score evidence that formed the basis of the second estimate is shown in the bar charts below.







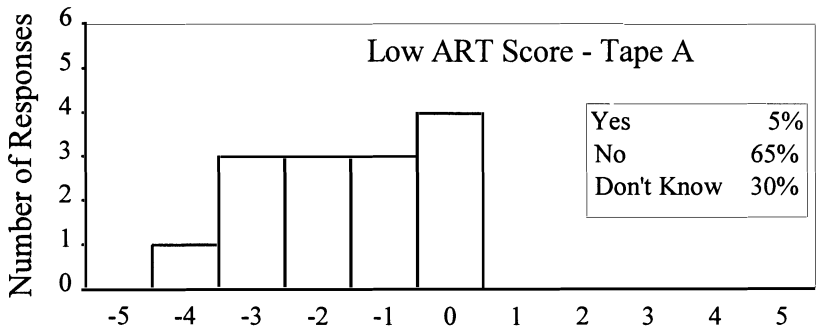
Less Chance — Likelihood that ad contravenes section 9 — More Chance

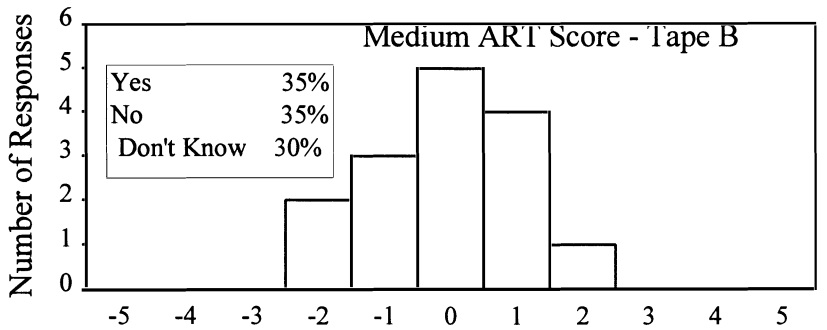
TOYOTA CASE

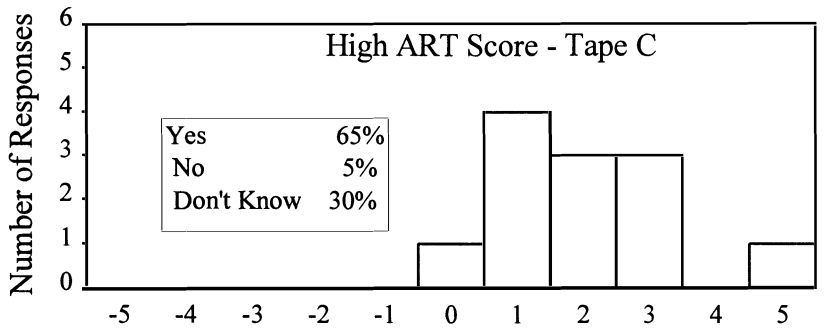
In this example it was alleged that the advertisement conveyed the representation that *Six million dollars was spent on research and development directly related to the Toyota Corona.*

All participants made their first estimates based only on the standard assumptions.

The ART score evidence that formed the basis of the second estimate is shown in the bar charts below.





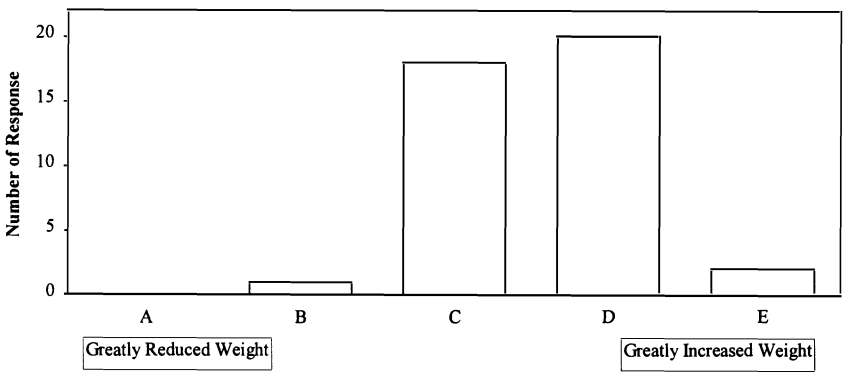


Less Chance — Likelihood that ad contravenes section 9 — More Chance

After predicting the outcome of the nine hypothetical cases, participants were asked whether a number of possible changes to the format of audience reaction tests would affect the weight of the evidence. In describing the effect of these changes, participants used a five-point scale ranging from "greatly reduced weight" at one end to "greatly increased weight" at the other. The horizontal axis in the following bar charts corresponds to this scale. The vertical axis shows the number of participants who selected each point on the scale.

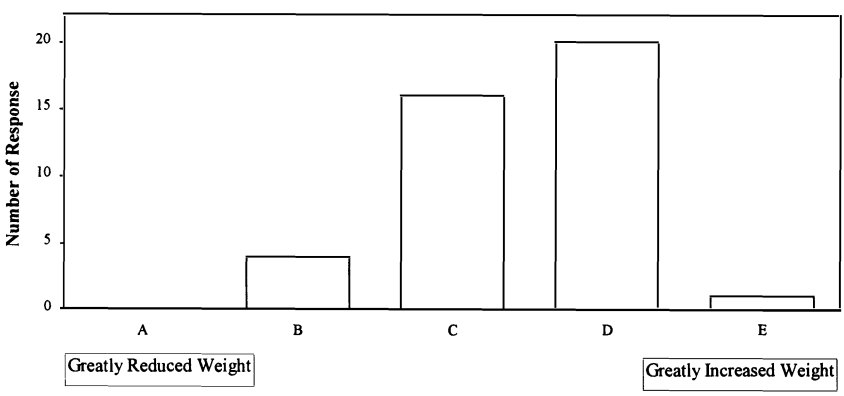
The first change investigated related to the number of subjects included in the tests. The vast majority of participants thought that doubling the sample size (i.e. from 1000 to 2000) would have either no effect or only slightly increase the weight given to the evidence as is shown below.

Double Sample from 1000 to 2000



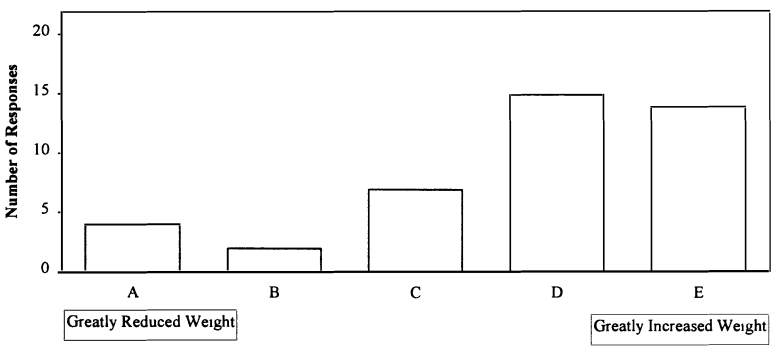
The second variation considered was the number of times the advertisement is shown before subjects respond to the survey questions. The vast majority of participants felt that showing the advertisement twice rather than once would have either no effect or only slightly increase the weight accorded to the evidence. A few believed this would slightly reduce the probative value of the evidence as is shown below.

Showing Ad Twice Instead of Once



The third variation examined was the possibility of using open-ended questions rather than forced-choice questions. Based on Professor Preston's suggestion, it was tentatively concluded that responses to forced-choice questions would be given greater weight than those to open-ended questions. This conclusion was not supported by participants. One third of participants thought that using open-ended questions would greatly increase the weight given to the evidence. Another third thought such evidence would be given slightly greater weight. Only a few supported the conclusion that such evidence would be given much less weight as is shown below.

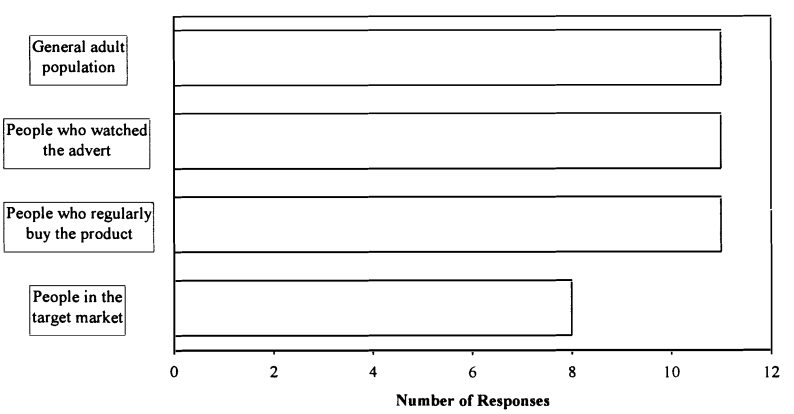
Using Open Ended Question



The final possible variation investigated related to the group from which subjects were selected. Participants were asked from which of four possible groups should subjects for audience reaction tests be selected, in order to provide the most probative evidence. These four groups were defined as follows:

- The general adult population
- The people who watched the programme(s) in which the advertisement was shown
- The people who regularly buy products of the type advertised
- The people who fall within the "target market" as defined by the advertiser

There was very little consensus among participants as to whose reactions the Court would find most relevant as is shown below.



VII. CONCLUSIONS

Several forms of evidence may be adduced on the issue of whether a television advertisement conveys the implied claim alleged by the plaintiff. The main forms of extrinsic evidence include:

- testimony of individual consumers
- expert opinion testimony (unrelated to any survey)
- survey evidence

The weaknesses of the first two forms of evidence are well recognised. The main problem with testimony of individual consumers is that they may be carefully selected and briefed prior to the hearing in such a way that substantial doubt is cast on the extent to which their responses are representative of the general television audience. The problem with expert opinion testimony is that no generally accepted theory of advertising exists from which one can predict with any confidence how consumers will interpret a particular advertisement.

These considerations have led to numerous calls for the Courts in Australia and New Zealand to place more reliance on survey evidence. For example, Warren Pengilley has argued that:

> [T]he results obtained from alternative methods of gathering evidence in areas properly subject to survey are so much less reliable than the evidence obtained by such properly conducted surveys.[42]

In FTC deceptive advertising proceedings in the United States, surveys are generally described as providing more probative evidence than either expert opinion testimony or testimony from individual consumers. Professor Preston states:

> The record shows that the best extrinsic evidence of deceptiveness is methodologically acceptable consumer research that pertains directly to the question of what message is conveyed.[43]

Despite this statement, considerable doubt remains about the actual impact of survey evidence in FTC decisions. This is due to the existence of the intrinsic evidence which is the evidence provided by the advertisement itself. It is difficult to determine from the reported decisions whether the Commission's findings were based on the survey evidence or on its own interpretation of the advertisement. As Professor Preston has noted:

> The FTC, which is empowered to decide without extrinsic evidence if it so chooses, might arbitrarily choose to accept as relevant any indirect evidence that suits its predilection and reject as irrelevant any that does not ... The same potential for arbitrariness could also exist with respect to methodologically flawed research. While there is no question that such flaws may lead to rejection, the record shows that they sometimes do not prevent the research from yielding findings.[44]

The FTC typically describes survey evidence as corroborating its own view of the claims implied by the advertisement. It rarely describes the evidence as decisive on the issue.

Similar attitudes are evident in section 9 cases in Australia and New Zealand. The Courts usually describe the question of whether conduct is

42 Pengilley W, "Survey Evidence in Trade Practice Cases" 5 AMLB 1 at 3.
43 Preston I L, "Extrinsic Evidence in Federal Trade Commission Deceptiveness Cases" [1987] *Columbia Business Law Review* 633 at 691.
44 Ibid, at 693.

misleading or deceptive as a question of fact. However, they repeatedly emphasise that this question is one for the Court to decide. This suggests that the most important factor in the Court's finding as to the claims implied by a television advertisement is the judge's own interpretation of the commercial. Indeed, it might be argued that gathering any extrinsic evidence on this issue is simply a waste of time and money. According to this view, a judge is unlikely to pay any attention to a survey that conflicts with his or her own perception of the claims implied by a television advertisement.

The research described here sought to evaluate this view. The major finding of the research is that barristers experienced in section 9 litigation generally believe that ART scores would influence a judge with respect to the issue of whether a television advertisement conveys the implied claim(s) alleged by the plaintiff.

Analysis of all the estimates for the nine cases shows that the mean prediction of the chances of the advertisement being found deceptive where no evidence of ART scores was provided was 42% (ie 4.2 on the scale). This compares with a mean of 25% for predictions based on a low ART score and a mean of 61% for predictions based on a high ART score. Thus, experienced barristers believe that most judges would be willing to forgo their own interpretation of the advertisement where this interpretation was shown to conflict with evidence of audience reaction tests. On this basis, it is concluded that audience reaction tests provide highly probative evidence of the claims implied by a television commercial.

The possibility that some other type of survey would provide even more probative evidence should not be overlooked. This possibility was explored only briefly in the study. The initial findings suggest that increasing the size of the survey sample or the number of times the advertisment is shown to the audience would have little effect on the probative value of the evidence. Furthermore, it appears that alternative conceptions of the survey population would have little impact on the weight attributed to the evidence.

Future research might address the issue of whether surveys using open-ended questions would be given more weight than those using forced-choice questions. Most participants in this study felt that this was likely to be the case. This finding probably reflects a perception of forced-choice questions as being leading questions. It may also reflect participants' unfamiliarity with the United States decisions and the problems associated with open-ended questions. In our view, Professor Farmer's comments on this issue deserve particular attention:

> It is surely inevitable that an analysis of any question, or series of questions, whether contained in a survey or asked in court, will always be able to reveal certain assumptions contained in the question or questions which might be said to a make a particular question to some extent leading. The mere asking of a question as a sequel to an earlier question on the same subject-matter will necessarily involve predication of the earlier question and hence to some extent pre-condition the response. It is respectfully submitted therefore that, in analysing the objectivity of the form of a questionnnaire used in a survey, a reasonably realistic approach should be taken. Surveys should not be judged more harshly in this respect than other kinds of evidence.[45]

The extent to which this comment influences judicial attitudes to survey evidence in Australia and New Zealand remains to be seen. The early signs

45 Farmer J A, *Australian Trade Practices Reporter (Vol 2)* at 15-095.

in Australia suggest that survey evidence is no more enthusiastically received since the *Arnotts* decision than previously.[46]

However, it is possible that in some cases where survey evidence is described as worthy of little weight, the evidence has, nevertheless, influenced the Court's decision. The exploration of this possibility would seem to require empirical research methods of the kind adopted in this study.

46 Survey evidence was admitted but found to be worthy of little weight in both *State Government Insurance Corporation v Government Insurance Office of NSW* (1991) ATPR 41-110 and *Interlego AG v Croner Trading Pty Ltd* (1991) ATPR 41-124.