

Mystery customer research: cognitive processes affecting accuracy

Lisa J. Morrison British Psychological Society,
Andrew M. Colman and Carolyn C. Preston University of Leicester



Lisa J. Morrison



Andrew M. Colman



Carolyn C. Preston

Mystery customer research is a technique of quality assessment in the retail sector, where it is called mystery shopping, and also in the service sector. It is growing rapidly in popularity, but research in cognitive psychology suggests a number of potential threats to the reliability and validity of data collected through its use. In particular, various factors associated with the encoding, storage, and retrieval of information by mystery customer assessors are likely to influence the accuracy of the results, and individual differences between assessors should also be taken into account in designing mystery customer surveys. A number of specific recommendations designed to minimise errors arising from memory failures and distortions are outlined and discussed.

Introduction

Mystery customer research, although it is not without its critics (e.g. Brown 1990), is an industry that is currently worth an estimated £10m per annum in the United Kingdom (Miles 1993) and is becoming increasingly popular in the United States (Eisman 1993; Schlossberg 1991), Australia (Dougherty 1987), and elsewhere. One American market research company, Customer Perspectives, spends 95% of its time on mystery customer research (Wolfensberger 1990). In the United Kingdom, a Mystery Shopping

Practitioners Group has been established under the aegis of The Market Research Society.

In The Market Research Society's *Organisations Book* 1994, the Professional Standards Committee of The Market Research Society identified 28 agencies conducting mystery customer research, and by 1995 the figure had risen to 187 (Dawson & Hillier 1995). The use of mystery customer research as an evaluative tool for assessing the quality of both goods and service provision is growing rapidly in popularity, both as a method of self-assessment for client companies and as a technique for comparison between competitor companies. Of 88 commercial companies that responded to Dawson & Hillier's survey, more than two-thirds had commissioned mystery customer research in their own companies, competitors' companies, or both. As a market research technique, it is likely to continue increasing in popularity, because it is widely applicable in virtually any branch of the service and retail sectors and is continually finding new areas of application – for example, it was recently used to assess the feasibility of Mexican pharmacies assisting in AIDS and STD prevention and control through community education and condom promotion (Pick *et al* 1996).

Mystery customer research involves visits by specially trained assessors called mystery shoppers (in the retail industry) or more generally mystery customers to shops, restaurants, banks, or other businesses in which quality of provision is to be appraised. The assessors, posing as ordinary customers, check the attainment of a number of service standards that have been drawn up in consultation with the client company. Examples of service standards that might be assessed include the following: 'Did the cashier deal with me courteously?', 'Did the sales assistant check that I had the right size garment?', 'Was a pool table provided in the pub?', 'Was I served within two minutes?', 'Did the bank teller smile?', 'Were the ashtrays emptied regularly?'

From the results of a mystery customer survey, the standards attained by a particular company can be compared with standards attained by its rivals, and then decisions can be made as to what new standards are realistic, achievable, and potentially most important in the competitive market as it exists at the time. The information thus provided can be of considerable commercial benefit to the client company (Hurst 1993; Jones 1993; Leeds 1992).

Despite the general popularity of mystery customer research, some companies avoid using it because of worries about potential problems that might arise in the absence of stringent guidelines to ensure accuracy of evaluations of individual companies and industries. Although codes of conduct have been introduced by the MRS and ESOMAR, some aspects of the technique evidently remain 'open to interpretation and manipulation' (Dawson & Hillier 1995, p. 417). Thus, in the absence of strict guidelines, a mystery customer survey can to some extent be tailored to fulfil a client company's specific requirements. The list of standards to be assessed is of central importance, and assessors may have to receive generic training in sales and service techniques in order to be able to assess them. For example, they may need to be able to differentiate between acknowledgements and greetings, open and closed questions, or features and

benefits. A very high degree of accuracy in reporting must be maintained to enable the targeting of subsequent improvements by management. It is important that the field force of mystery customer assessors should conduct their assessments in a consistent way.

Potential problems

Published data on the accuracy (reliability and validity) of mystery customer research appear to be non-existent, although properly designed and executed surveys using trained and impartial assessors checking the attainment of clearly defined and objective standards are likely to be more reliable and valid than conventional market research surveys.

There is a great deal of research in cognitive psychology (see, e.g. French & Colman 1995; Sternberg 1996) that has a bearing on mystery customer research and on factors that are likely to enhance or undermine its validity and reliability. Of prime importance is research into factors related to memory processes that may affect the accuracy of data recorded by mystery customer assessors. Given the importance that is usually attached to the results of mystery customer surveys, it is vital to design and implement them in a manner that is likely to minimise potential memory problems. What follows is a summary of relevant findings from cognitive psychology and some suggestions regarding their implications for mystery customer research.

Reliance on memory

The standard mystery customer procedure involves the assessor visiting the target premises, noting whether the standards that are to be assessed are being satisfactorily attained, and then retiring to a private place to fill in an assessment form. This procedure makes considerable demands on the assessors' memories, and there are two obvious problems that might arise from memory failures on their part. First, an assessor may forget to check on the attainment of one or more of the standards on the list before retiring to fill in the assessment form – there is sometimes quite a long list of items to be remembered. Secondly, having noted all the right details on the list of items (whether the standards were attained), an assessor has to remember them and eventually record them correctly on the assessment form. Theory and empirical research in cognitive psychology suggests that memory problems can arise at three different stages of this memory process:

- First, the assessors' perception and encoding in memory of the relevant details associated with the standards being assessed may be incomplete or inaccurate.
- Secondly, the information may be accurately perceived and encoded but forgetting or degradation of the memory trace may occur during the period of storage before it is recorded on the assessment form.
- Thirdly, the information may be accurately perceived, encoded, and stored, but problems may none the less occur at the retrieval stage when the assessor has to recall information in order to record it on the assessment form.

At each of these stages – encoding, memory, and retrieval – either random errors or systematic distortions may occur in such a surreptitious manner that the assessor concerned may be unaware of them.

ENCODING

Factors that influence memory accuracy at this stage of processing include the following.

Physical factors, such as lighting which, if poor, may reduce the likelihood of accurate perception of details that require careful observation (Yarmey 1986). This may be important in relation to such standards as cleanliness of the premises and efficiency of the maintenance staff in relation to the target premises. Moreover, the time of day at which the assessor visits the premises may affect encoding accuracy due to fatigue, which lessens perceptual sensitivity especially when a large number of observations are required (Guerrien, Leconte-Lambert & Leconte 1993; Parasuraman, Warm & Dember 1987). But beyond pointing out that late night visits (and for some people early afternoon visits) are likely to be affected by fatigue, it is impossible to be more specific because different people have different biological rhythms, and some effects of diurnal variations or circadian rhythms operate in complex ways. For example, there is evidence indicating that effects on memory performance depend crucially on whether the individual is a 'lark' (a morning type) or an 'owl' (an evening type) (Anderson, Petros & Beckwith 1991), and other studies have suggested that there are important factors in addition to fatigue and circadian rhythms that produce time-of-day effects on memory (Leirer, Tanke, & Morrow 1994).

Attentional focus. It is well documented that we do not recall events uniformly. Perception is fallible and selective, thus different people selectively attend to different aspects of an event, person or place. This may result in reconstructive memory distortion in which gaps in memory are 'filled in' with inferences based on assumptions and expectations rather than factual observations. Research has shown that people who reconstruct memories in this way are often unaware that they are doing so; in other words, this problem can occur without any conscious awareness on the part of the assessor.

In relation to this attentional focus, recent studies have shown differential effects on memory for central (relatively important) versus peripheral (relatively unimportant) details. Central details are more likely to be accurately encoded and retained in memory (Burke, Heuer & Reisberg 1992; Christaansen & Loftus 1987; Heuer & Reisberg 1990), but it is essential to acknowledge that 'what is central and what is peripheral in any given situation are entirely in the eye of the beholder' (Spencer & Flin 1993, p 302). For some assessors, whether or not the toilet facilities in a pub are free of vandalism and graffiti may be central and whether or not the grass area outside the pub is well maintained may be peripheral, but for others the relative importance of these two standards may be reversed. Thus, it is difficult or impossible to predict in advance which standards are likely to be affected by attentional focus with a specific assessor, but the

assessors should at least be made aware of this problem.

Attitudes and social pressures. Preconceptions and prejudices are known to influence recall (Boon & Davies 1993). Classic conformity experiments such as those of Asch (1956) demonstrated how some people's perceptions can be altered by social pressures. In mystery customer surveys, assessors may be subject to subtle social pressures and may wish to give favourable reports of customer service because of a natural empathy with the people working in the target establishments, and if they feel warmly towards these people, objectivity may be difficult to maintain.

Attitudes and social pressures are especially likely to interfere with accurate reporting when subjective or ambiguous judgements are involved. Some of the standards that are assessed in mystery customer surveys require difficult judgements, even when only Yes/No answers are required on the assessment form. For example, one could argue that tidiness is a spectrum, from extremely tidy at one extreme to extremely untidy at the other. To answer Yes or No to the question, 'Was the shop tidy?', or 'Was the fitting room tidy?' the assessor has to make a judgement as to where on this spectrum the division between Yes and No lies. This judgement will vary from individual to individual, depending on attitudes, preferences, and previous experience, so two different assessors confronted with identical situations, even if they are alert, conscientious, and well trained, may give different answers to this question and to others like it.

Despite the fact that assessors are trained, an individual's preferences and personal opinions regarding what is and what is not acceptable will inherently bias judgements regarding cleanliness, friendliness and courtesy of staff, appropriateness of music, and so forth.

Special knowledge. Expertise leads to better recall of events and circumstances (Logie, Baddeley & Woodhead 1989; Logie, Wright & Decker 1992; Yuille 1984). Thus, even if mystery customer assessors are all trained to roughly the same level of competence, some may have considerably more expertise than others, and this is bound to affect levels of consistency in assessment reports. Mystery customers' own personal experiences of the service under scrutiny are also liable to affect their reports. For example, previous hospital experience has been shown to influence patients' ratings of satisfaction and service quality in hospitals (Joby 1992).

Encoding and recall times. A survey reported by Dawson & Hillier (1995) indicated that, of the 88 respondent companies that had commissioned mystery customer research, just under 80% thought that assessment visits should be no longer than half an hour, and of these 40% felt that 10 minutes was too long. The respondents were right to worry about the lengths of assessment visits, because recent research suggests that the duration of time during which a fixed amount of information is encoded has a marked effect on the accuracy with which it is later recalled (Reeder & Logue 1995). Thus in an assessment visit requiring a significant amount of memory encoding, a reduction in the length of

the visit should be expected to reduce the amount of information accurately retained by the assessors.

STORAGE

Cognitive psychologists have known for over a century (since the work of Ebbinghaus 1885) that, with the passage of time, details in memory may be lost through decay of memory traces, interference from competing memories, and other processes. Moreover, as the delay between exposure and reporting increases, information in memory becomes altered and reinterpreted to fit in with prior knowledge (Bartlett 1932). Consequently, memory becomes more reconstructive and less reproductive.

During storage, memory is highly susceptible to the uptake of extraneous information, although this becomes a major problem only if the original memory trace is relatively weak (Loftus, Levidau & Duensing 1992; Tousignant, Hall & Loftus 1986). Research into the reliability of eyewitness testimony has shown that recollections of peripheral details are more likely than central features to be altered by information acquired after the event – information conveyed through the questions asked by the police, information read in newspapers, and so on (Hall, Loftus & Tousignant 1987; Marquis, Marshall & Oskamp 1972; Marshall 1966). For example, in relation to a question such as, ‘Was the car park well-maintained?’, mystery customer assessors are advised to think in terms of potholes and ‘significant hollows’, but if they cannot actually recall the state of the car park but do recall driving over a bump, they may report that the car park was in a bad condition even if the bump in question was actually a ‘sleeping policeman’ designed to prevent speeding.

More importantly, research findings have shown explicitly that an observer’s recollection of events and people is a function not only of what was actually perceived but also of the observer’s expectations (Doob & Kirschenbaum 1973; Shepherd & Ellis 1973; Wall 1965). Anything that interferes with accurate perception or storage of relevant details is liable to lead to biased reports affected by the assessors’ prior expectations rather than the objective facts (Baker 1961). A large number of research studies (reviewed by Penrod, Loftus & Winkler 1982) have shown that personal prejudices, expectancies, and social attitudes can all influence and distort information stored in memory. Moreover, these expectancy biases occur most often in situations – such as those confronting mystery customer assessors – in which a large amount of information has to be remembered (Guastello, Traut & Korienek 1989; Macrae, Hewstone & Griffiths 1993).

RETRIEVAL

It is clear from a great deal of research evidence that we all remember more than we can recall at any one time, so that there is a distinction between available and potentially accessible information (Tulving 1983). The implication of this is that the format of the assessment form on which assessors record their recollections can affect the accuracy of what is recorded on it. More detailed and precise

questions are not necessarily the answer: there is evidence to suggest that excessively detailed questioning may even decrease accuracy of recall by encouraging reconstructive memory distortion and by introducing suggestive questioning, that is, 'leading questions' (Lipton 1987). What is required is a format that is neither too coarse nor too detailed, and research is needed to determine the optimal balance for mystery customer assessment forms.

Research has shown that recall is often enhanced by contextual reinstatement, that is, by returning to the context in which the memory was encoded (Bekerian & Bowers 1983). Although it is difficult to see how mystery customer assessors can exploit this directly, they may be able to use it indirectly. If they have difficulty recalling certain details while filling in an assessment form, they might find it helpful to shut their eyes and imagine themselves back in the place where their observations were made. Nevertheless, recent research also suggests that how well information transfers from one environment to another depends on how similar they 'feel' to the individual rather than how similar they 'look' (Eich 1995). Even when target events are encoded and retrieved in the same physical setting, memory performance suffers if the individual's emotional state changes between the encoding and retrieval phases, and this phenomenon is called *mood-dependent memory*. Consequently, assessors should ideally be in a relatively neutral mood state both when they make their assessments and when they record the results. Strong emotions are liable to impede the memory process, and of course they are also likely to distort the results by influencing any interpersonal interactions that take place during the assessment visit.

In addition to the factors already mentioned, individual differences between assessors are bound to influence their reports. Factors such as gender and age have been shown to affect the reliability and accuracy of details recalled from memory. The superiority of females in accuracy and completeness of eyewitness testimony was established many decades ago (Whipple 1909) and has been confirmed by numerous subsequent research studies. Also, it has been shown that the reliability of memory increases rapidly during childhood but decreases slowly from middle age (Ceci, Ross & Toglia 1987; Yarmey & Kent 1980).

Other psychological processes

Apart from problems related to memory failures, there are other psychological processes that are worth considering. A mystery customer assessment is, by its nature, a report of an individual rather than of a representative sample of the customer population. An encounter between staff and customer is a two-way interaction and is influenced by the behaviour and appearance of both participants. One important implication of this is that individual differences between assessors are bound to be reflected in the assessment.

The effects of some of these individual differences have been well documented. For example, it has been found in a department store setting that men tend to get service priority over women, and that style of dress and gender interact to influence service priority (Stead & Zinkhan 1986; Zinkhan & Stoiadin 1984). Similar gender differences have been found in other service settings, for

example Hall (1993) found that waiters and waitresses preferred serving men and saw women customers as less friendly and harder to serve, but customers saw waitresses as more friendly than waiters. Galin & Benoliel (1990) found that the effect of the gender and dress of staff on their performance rating depended on the gender and dress of the raters themselves: staff of the same gender and style of dress as the rater received the highest ratings. Also, casually dressed raters tended to give higher ratings overall than smartly dressed raters. On a broad level, this suggests that the features of interactions between staff and customers, and also the perceptions of these interactions, are affected by both staff and customer differences. This may, for example, translate into lower service ratings from women assessors.

These points suggest that different mystery customer assessors may have different experiences in the same target establishments and also that similar encounters may be interpreted very differently by different assessors. The implications of this should always be borne in mind when interpreting aspects of mystery customer reports that relate to personal interactions. In these cases, perceptions are dependent to a large degree on the characteristics of the assessors and their responses to the transaction, and they may not necessarily be representative of the entire customer base. Only further research can reveal whether the effects of differences between assessors are outweighed by the differences between service or goods providers. This is covered by the suggestion made earlier that research is urgently needed into the reliability and validity of mystery customer assessments.

Some of the problems arising from memory failures and individual differences between assessors, and particularly problems of establishing the reliability of data that depend on unverified recall of information, have arisen in different areas of market research. In the 1950s, for example Proctor & Gamble and other companies used a technique of memory-based interviewing in which the interviewer conducted each interview on the respondent's doorstep, following a memorised series of open-ended and probe questions, and recorded the responses from memory away from the interviewee some time after completing the interview. It was felt that this technique, which enabled interviewers to maintain eye contact with interviewees and pay full attention to their answers, might elicit superior data (Squirrel 1996). Although the technique was considered successful at the time, intensive training of interviewers was felt necessary to maximise consistency and accuracy, and it was recognised that the reliability and validity of data collected in this way were difficult to establish. As the proportion of households without telephones declined, memory-based interviewing was largely superseded by telephone interviewing. However, similar problems of consistency, accuracy, reliability, and validity beset mystery customer research, which normally requires techniques similar to the memory-based interviewing of the 1950s.

Conclusions and recommendations

Mystery customer surveys provide an excellent market research technique that has many advantages over more conventional methods. The accuracy (reliability and validity) of mystery customer research is unknown, but surveys that are properly designed and executed are probably at least as reliable and valid as more conventional market research surveys. There are nevertheless potential threats to the accuracy of mystery customer surveys. Some of these potential threats arise from the memory demands that they place on the assessors, who normally record the attainment or non-attainment of various standards that they have observed some time after making the relevant observations. While they are visiting the target premises they may forget to check whether one or more of the specified standards was satisfactorily attained, and after they have retired to fill in their assessment forms they may forget significant details related to the standards that they have checked. Omissions and distortions of memory can arise at all three stages of the memory process: encoding, storage, and retrieval. In the light of this, a review of findings from cognitive psychology suggests a number of steps that could be taken in designing and carrying out mystery customer surveys to minimise errors arising from memory failures.

- In order to reduce the memory burden on assessors, it might be possible to restrict their task to checking the attainment of personal and interactive standards of service delivery that only mystery customers can judge – for example, ‘Was I served within two minutes?’, ‘Did the bank teller smile?’, ‘Were the ashtrays emptied regularly?’. This would leave them free to concentrate on assessments tied in with the transaction and would also reduce the memory demands of their task, thereby helping to minimise errors arising from memory overload. It would relieve the assessors of the burden checking and memorising whether the impersonal and relatively fixed, ‘physical’ standards were satisfactorily attained – for example, ‘Were the lights and ventilation functioning properly?’, ‘Were the toilets in working order?’, ‘Was the company logo prominently displayed?’. These impersonal standards could be left to non-mystery customer assessors who could openly carry clip-boards and would have no need to commit anything to memory. Although this suggestion would involve two assessors making two separate site visits instead of one assessor making one visit, each visit would be shorter and the quality of the reports would be higher, so it is worth considering.
- The second suggestion concerns the recording of observations. It is essential that this should take place during or immediately after the visit to reduce the problems of decay and reconstructive memory distortion. Recording should probably be done in writing rather than via telephone interviews, and the questions on the assessment forms should be carefully designed to give maximal retrieval cues and above all to minimise the use of suggestive or leading questions (e.g. ‘Was the lawn overgrown?’). The best format and wording of the assessments forms seems to be a question on which research is urgently needed.
- It may be possible to reduce memory problems in mystery customer research

by using event recorders. These are small devices that can be carried in one's pocket, and if all the standards require simple binary (Yes/No) answers, for example, then the assessors could simply press one button for Yes and the other for No, and the event recorders would log strings of Yes and No symbols, which could later be interpreted as answers to a series of questions in a known order. The assessors' memory task would then be restricted to remembering what standards to check and in what order to check them.

- Assessors should be encouraged to make their visits at a time of day when they are alert and not tired and when the ambient lighting gives them the best chance of seeing what needs to be seen.

- The training of assessors should include a suggestion that, if they have difficulty remembering certain details while filling in an assessment form, they should try shutting their eyes and vividly imagining themselves back in the place where their observations were made. In addition, as far as possible assessors should attempt to retain a neutral emotional state throughout the assessment visit and when recording the results, although trainers should acknowledge that, in practice, complete emotional neutrality may be difficult to maintain throughout an assessment.

- Assessors should be warned about the problem of social pressure and the tendency many people have to prefer giving favourable reports rather than unfavourable ones, especially if the people working in the target establishments seem pleasant or easy to empathise with. They should also be encouraged to assess each establishment objectively on its own merits rather than consciously or unthinkingly making direct comparisons between different establishments.

- The standards that form the basis of mystery customer surveys should be as objective as possible. For example, 'Was I served within two minutes?' is completely objective, but 'Was the bar tidy?' or 'Was the shop tidy?' requires a subjective and debatable judgement, which is likely to undermine the reliability and validity of a survey, and the same applies to questions regarding cleanliness, friendliness and courtesy of staff, appropriateness of music, and so forth. The client company should be asked wherever possible to specify exactly what they mean by 'tidy', 'clean', and so on, in order to enable the definition of objective standards.

- Video recordings of interactions between mystery customers and service providers may be useful in the training of assessors. Ideally, recordings should be taken from hidden cameras (in briefcases, for example), so that from the point of view of the service providers the interactions are not out of the ordinary, but this may often be impractical. Training sessions using video recordings of service encounters in conjunction with mystery customer reports of these encounters would be an efficient method of providing trainees with feedback. In practice, video recordings of a few typical service encounters, including common problems and difficult distinctions, may be useful for training future mystery customers and establishing common standards.

- On the whole, women are likely to provide more accurate mystery customer reports than men, and for some surveys it may be best to use women assessors

only. However, women customers are also likely to be treated differently from men (see above), so it may often be undesirable to exclude male assessors. The ages of assessors are also likely to affect the results of mystery customer surveys, and on the whole young adults are likely to be most reliable, but for the same reason mentioned in relation to gender, it may be inappropriate in some surveys to exclude older assessors. These factors should all be considered carefully in designing surveys.

- Buyers and users of mystery customer research should establish a 'best practice' protocol for conducting mystery customer surveys and should then stick to it rigorously (Dawson & Hillier 1995). Changes in procedure can have unpredictable and unknown effects on the validity and reliability of the findings.
- Further research is required into the optimal design of assessment forms for recording observations, the effects of gender, age, and other demographic factors on the reliability of assessment, and most importantly of all, on the reliability and validity of mystery customer surveys in general. It may be worth exploring the use of video recordings of service encounters, referred to earlier, to investigate the reliability and validity of mystery customer reports.

References

- ANDERSON, M. J., PETROS, T. P. & BECKWITH, B. E. (1991). Individual differences in the effects of time of day on long-term memory access. *American Journal of Psychology*, **104**, pp 241-255.
- ASCH, S. E. (1956). Studies of independence and conformity: a minority of one against a unanimous majority. *Psychological Monographs*, **70**, 9, Whole No. 416.
- BAKER, C. H. (1961). Maintaining the level of vigilance by means of knowledge of results about a secondary vigilance task. *Ergonomics*, **4**, pp 311-316.
- BARTLETT, F. C. (1932). *Remembering: a study in experimental and social psychology*. Cambridge: Cambridge University Press.
- BEKERIAN, D. A. & BOWERS, J. M. (1983). Eyewitness memory: were we misled? *Journal of Experimental Psychology: Learning, Memory and Cognition*, **9**, pp 139-145.
- BOON, J. C. W. & DAVIES, G. M. (1993). The influence of biographical information on event memory: when it's not what you know but whom you know. *Journal of General Psychology*, **120**, 517-530.
- BROWN, M. G. (1990). How to guarantee poor quality service. *Journal for Quality & Participation*, December, pp 6-11.
- BURKE, A., HEUER, F. & REISBERG, D. (1992). Remembering emotional events. *Memory and Cognition*, **20**, pp 277-290.
- CECI, S. J., ROSS, D. F. & TOGLIA, M. P. (1987). Age differences in suggestibility: narrowing the uncertainties. In S. J. Ceci, M. P. Toglia & D.F. Ross (Eds.), *Children's eyewitness memory* (pp 79-91). New York: Springer-Verlag.
- CHRISTAANSEN, S. & LOFTUS, E. F. (1987). Memory for traumatic events. *Applied Cognitive Psychology*, **1**, pp 225-239.
- DAWSON, J. & HILLIER, J. (1995). Competitor mystery shopping: methodological considerations and implications for the MRS and Code of Conduct. *Journal of the Market Research Society*, **37**, pp 417-427.
- DOOB, A. & KIRSCHENBAUM, H. (1973). Bias in police lineups: partial remembering. *Journal of Police Science and Administration*, **1**, pp 287-293.

- DOUGHERTY, B. (1987). Service with a scowl. *Rydge's*, **60**, 3, pp 77-81.
- EBBINGHAUS, H. (1885). *Über das Gedächtnis*. Leipzig: Dunker. [Trs. H. Ruyer & C. E. Bussenius, *Memory*. New York: Teachers College Press, 1913]
- EICH, E. (1995). Mood as a mediator of place-dependent memory. *Journal of Experimental Psychology: General*, **124**, pp 293-308.
- EISMAN, R. (1993). Going undercover. *Incentive*, **167**, 5, 70-71.
- FRENCH, C. C. & COLMAN, A. M. (Eds.). (1995). *Cognitive psychology*. London: Longman.
- GALIN, A. & BENOLIEL, B. (1990). Does the way you dress affect your performance rating? *Personnel*, **67**, 8, pp 49-52.
- GUASTELLO, S. J., TRAUT, M. & KORIENEK, G. (1989). Verbal versus pictorial representations of objects in a human-computer interface. *International Journal of Man-Machine Studies*, **31**, pp 99-120.
- GUERRIEN, A., LECONTE-LAMBERT, C. & LECONTE, P. (1993). Time of day effects on attention and memory deficiency: is chronopsychology a method for studying the functioning of the human subject? *Psychologica Belgica*, **33**, pp 159-170.
- HALL, D. F., LOFTUS, E. F. & TOUSIGNANT, J. P. (1987). Post-event information and changes in recollection for a natural event. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: psychological perspectives* (pp 124-142). Cambridge: Cambridge University Press.
- HALL, E. J. (1993). Smiling, deferring and flirting: doing gender by giving good service. *Work & Occupations*, **20**, 4, pp 452-471.
- HEUER, F. & REISBERG, D. (1990). Vivid memories of emotional events: the accuracy of remembered minutiae. *Memory and Cognition*, **18**, pp 496-506.
- HURST, S. (1993). Revealed at source. *Managing Service Quality*, January, pp 449-451.
- JOBY, J. (1992). Patient satisfaction: the impact of past experience. *Journal of Health Care Marketing*, **12**, 3, pp 56-64.
- JONES, G. (1993). Measuring customer service. *Association of British Market Research Companies Bulletin*, **38**, March, pp 1-3.
- LEEDS, B. (1992). 'Mystery shopping' offers clues to quality service. *Bank Marketing*, **24**, 11, pp 24-26.
- LEIRER, O. VON, TANKE, E. D. & MORROW, D. G. (1994). Time of day and naturalistic prospective memory. *Experimental Ageing Research*, **20**, pp 127-134.
- LIPTON, J. (1987). On the psychology of eyewitness testimony. *Journal of Applied Psychology*, **62**, pp 90-95.
- LOFTUS, E. F., LEVIDAU, B. & DUENSING, S. (1992). Who remembers best? Individual differences in memory for events that occurred in a science museum. *Applied Cognitive Psychology*, **6**, pp 93-107.
- LOGIE, R., BADDELEY, A. & WOODHEAD, M. (1989). Face recognition, pose and ecological validity. *Applied Cognitive Psychology*, **1**, pp 53-70.
- LOGIE, R., WRIGHT, R. & DECKER, S. (1992). Recognition memory performance and residential burglary. *Applied Cognitive Psychology*, **6**, pp 109-123.
- MACRAE, C. N., HEWSTONE, M. & GRIFFITHS, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, **23**, pp 77-87.
- MARQUIS, K. H., MARSHALL, J. & OSKAMP, S. (1972). Testimony validity as a function of question form, atmosphere, and item difficulty. *Journal of Applied Social Psychology*, **2**, pp 167-186.
- MARSHALL, J. (1966). *Law and psychology in conflict*. New York: Bobbs-Merrill.

- MILES, L. (1993). Rise of the mystery shopper. *Marketing*, July 29, pp 19-20.
- PARASURAMAN, R., WARM, J. S. & DEMBER, W. N. (1987). Vigilance: taxonomy and utility. In L. S. Mark, J. S. Warm, & R. L. Huston (Eds.), *Ergonomics and human factors* (pp 11-31). New York: Springer-Verlag.
- PENROD, S., LOFTUS, E. F. & WINKLER, J. (1982). The reliability of eyewitness testimony: a psychological perspective. In N. L. Kerr & R. M. Bray (Eds.), *The psychology of the courtroom* (pp 119-169). Cambridge: Cambridge University Press.
- PICK, S., REYES, J., ALVAREZ, M., COHEN, S., CRAIGE, J. & TROYA, A. (1996). AIDS prevention training for pharmacy workers in Mexico City. *AIDS Care: Psychological and Socio-Medical Aspects of AIDS/HIV*, 8, pp 55-69.
- REEDER, K. P. & LOGUE, P. E. (1995). Information processing time adds useful information to traditional recall measures. *Archives of Clinical Neuropsychology*, 10, pp 89-99.
- SCHLOSSBERG, H. (1991). There's no mystery in how to retain customers. *Marketing News*, 25, 3, 10.
- SHEPHERD, J. W. & ELLIS, H. D. (1973). The effect of attractiveness on recognition memory for faces. *American Journal of Psychology*, 86, pp 627-633.
- SPENCER, J. & FLIN, R. (1993). *The evidence of children: the law and the psychology* (2nd ed.). London: Blackstone.
- SQUIRREL, N. (1996). Personal communication, 26 September.
- STEAD, B. A. & ZINKHAN, G. M. (1986). Service priority in department stores: the effects of customer gender and dress. *Sex Roles*, 15, pp 601-611.
- STERNBERG, R. J. (1996). *Cognition*. New York: Harcourt Brace.
- TOUSIGNANT, J. P., HALL, D. & LOFTUS, E. F. (1986). Discrepancy detection and vulnerability to misleading post-event information. *Memory and Cognition*, 14, pp 329-338.
- TULVING, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.
- WALL, P. (1965). *Eyewitness identification in criminal cases*. New York: Charles C. Thomas.
- WHIPPLE, G. M. (1909). The observer as reporter: a survey of the psychology of testimony. *Psychological Bulletin*, 6, pp 153-170.
- WOLFENBERGER, B. (1990). The rise of the mystery shopper. *New England Business*, 12, 12, pp 78-79.
- YARMEY, A. D. (1986). Verbal, visual and voice identification of a rape suspect under different levels of illumination. *Journal of Applied Psychology*, 71, pp 363-370.
- YARMEY, A. D. & KENT, J. (1980). Eyewitness identification by elderly and young adults. *Law and Human Behaviour*, 4, pp 359-371.
- YUILLE, J. C. (1984). Research and teaching with police: a Canadian example. *International Journal of Applied Psychology*, 33, pp 5-23.
- ZINKHAN, G. M. & STOIADIN, L. F. (1984). Impact of sex role stereotypes on service priority in department stores. *Journal of Applied Psychology*, 69, pp 691-693.

Preparation of this report was supported by Grants K38 and M71 from BEM Ltd. The authors are grateful to Gareth G. Jones for drawing their attention to this area of research.