Andrea Micheaux
is a partner at Analyse Informatique des Données (AID), a company based in Versailles which specialises in marketing databases and statistical analysis for marketing. She is an authority on database and relationship marketing in France, and lectures on postgraduate marketing and statistics courses in several universities.

Anne Gayet
has a postgraduate degree in statistics and has been a professional statistician in marketing for ten years. At AID she is head of the statistical analysis department, leading a team of analysts on database marketing and geo-marketing projects.

Keywords: relationship marketing, attitudes, database marketing, segmentation, loyalty, CHAID analysis

Andrea Micheaux
AID,
4, rue Henri le Sidaner,
78630 Versailles,
France
Tel: + 33 13923 9345
E-mail: andrea@micheaux.com

Turning a marketing database into a relationship marketing database

Andrea Micheaux and Anne Gayet

Received: 28 February 2001

Abstract
A significant change in the methodology of database marketing is under way. Echoing the ancient cry of the salesforce for data capable of turning customer contacts into sales, relationship marketing managers are beginning to suspect that despite huge volumes of transactional data and the availability of geo-demographic overlays, the variables they most need for marketing may not actually reside in their customer databases.

Introduction
Over the past ten to 15 years, companies and consultants have been enthusiastically data mining for segments of customers with economic leverage. We are capable of predicting churn rates, response rates, proportions of customers who will convert from being mono-product holders to multiple purchasers. We can decile the database according to lifetime value and identify the top strata of customers who account for over half the profits.[1] These customers are targeted for commercial action and tracked through reporting.

The problem is that once the cream has been skimmed from the best segments, cross-sell, churn, upgrade rates and lifetime value strata can remain depressingly stable, to the extent that the no one looks at the flow reports any more. The customers within these value segments are identified and described in terms of their characteristics. Usually, these characteristics are defined by the most discriminating among whichever variables happened to be in the database when the statistical analysis was carried out. Unfortunately, statistically discriminate definitions such as age combined with ‘date of last visit’ and ‘amount spent at first purchase’, or ‘number of calls to the customer service desk over the past three months’ combined with ‘method of payment’ and ‘geographic code’, are not necessarily helpful in determining how to convince customers to move into a ‘better’ or ‘lower-risk’ segment.

The principle of relationship marketing would have it that customers can be effectively stimulated or retained by using pertinent personal information in communication. In other words, the message or contact should be timely and have meaning for the customer in terms of his own lived experience.[2] In situations of high involvement and high risk, which is most often the case in database marketing, pertinence (as an essential element of involvement) contributes to the memorisation of sales arguments.[3] That customers notice and remember a ‘personally meaningful’ message more easily than mass communication is a founding principle of relationship marketing (and of advertising in general). Relationship marketing techniques ought, then, to provide an effective means of getting messages across to customers in segments on the database classified as having potential or being at risk. Sadly, the transactional variables defining these segments cannot be relied upon to provide the substance of a pertinent message or offer. Database marketers already use events in the customer life cycle, or driven by the product life cycle (such as the end of a loan or the mileage of a car), or occasions to contact which are valid within the cycle of the relationship (welcome call, service contact. . .) to endow their communication with meaning. However, by definition these timely actions can only be targeted at those customers falling into an event-driven segment that can be detected on the database. And some people have magic moments more often than others.

Event-driven communication

Clearly, databases are often lacking in the information which is of greatest potential use to relationship marketers: what the customer is looking for in the relationship, whether it makes any sense at all to the customer to enter into a direct relationship with the company, why he/she subscribed or purchased originally, what he/she thinks about cars, insurance or gardening in general (whether he/she even cares about cars, insurance or gardening), and so on.

The current move toward psychographic[4] overlays or data enrichment is evidence that direct and interactive marketers are climbing out of their data mines and widening their perspectives. No longer content with analysing variables to hand, they are looking for information which will not only be useful in discriminating between segments of differing economic potential but which will also be helpful in communicating with these targeted customers on an individual, personal level, conferring meaning and increasing the impact and level of retention of the message.

The dimensions of the customer relationship

What, then, is the 'ideal' content of the relationship marketing database? The authors believe that marketers should evaluate the dimensions of the customer relationship which are relevant and important to their businesses along the lines of the template set out in Figure 1. This gives four broad categories of data: behavioural, attitudinal, potential value and socio-demographic. Within these categories, different types of data contribute to both customer insight for effective communication and the discriminating power of commercial predictive modelling and segmentation. Some of these data categories are well known to database marketers and exist in most marketing databases (solid lines), others are less well known to practitioners and are not currently integrated at individual level in proportions that would allow the use of such data in targeting and personalisation (dotted lines). It is suggested that managers think along these dimensional lines (reviewed in some detail below), both identifying areas which are relevant to their own businesses and which may not be available to marketing, and revisiting even well-known data categories in the light of relationship marketing theory and practice.

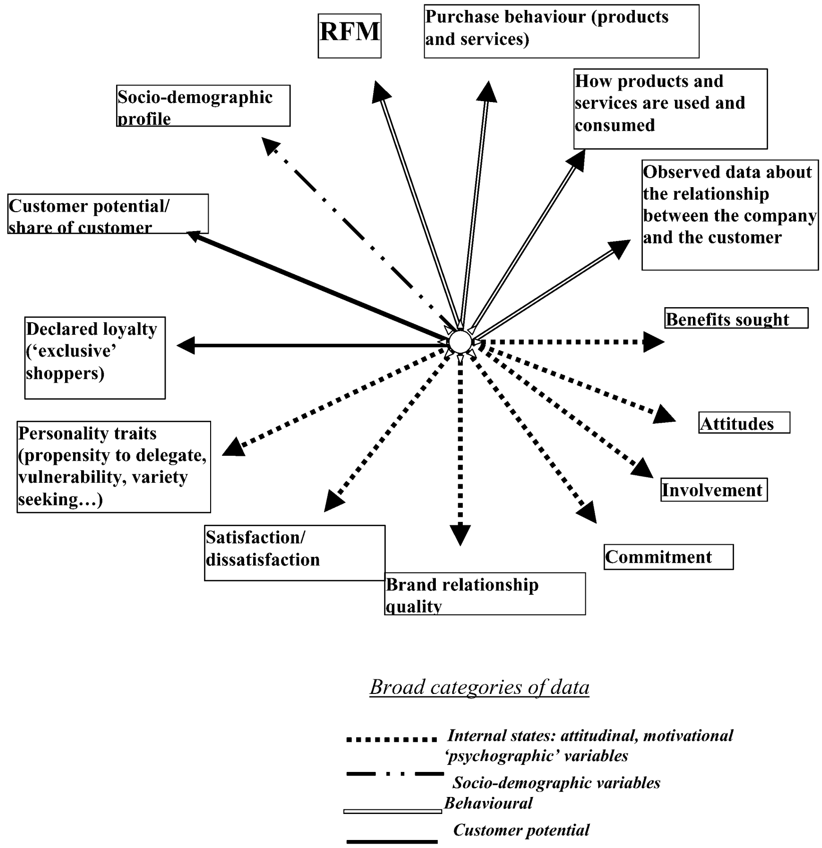

Figure 1: Dimensions of the customer relationship as a template for the content of a relationship marketing database

Dimensions of the customer relationship as components of the relationship marketing database

RFM data

Transactional recency/frequency/monetary (RFM) data fuelled the development of database marketing and filled customer databases to the extent that they became warehouses. Even highly aggregated behavioural data are an extremely powerful budgeting and relationship management tool. Mail-order companies, retailers, airlines and travel companies, such as La Redoute and Club Mediterranée, use RFM weighting[5] as a budgeting tool to predict future cash flow and fix cross-selling and prospecting objectives. Recency/frequency patterns became the basis for the first loyalty programmes, such as the Air France Club 2000 and the Club Med Millésia programme, rewarding customers with a '11111. . .'[6] frequent purchase pattern over a given period and at a given value. This solely behavioural interpretation of 'loyalty' has contributed to the divorce of the term 'loyalty' from relationship marketing, thus integrating the more qualitative aspects of interpersonal relationships. However, where relevant, RFM indicators should be present in any database, for their usefulness as predictors and because they are themselves endogenous variables.

Product purchase and product use information

Detailed information on products purchased generates huge amounts of data, particularly in retailing or mail order where thousands of product codes are analysed using data-mining techniques to reveal clusters of products which could represent cross-selling or merchandising opportunities (the 'beer and nappies' example, or in the case of Boots, photographic and infants' departments illustrate this principle[7]). In financial service companies, analysis of the order sequence in which accounts are opened is highly conducive to developing ongoing cross-selling operations. Some marketing databases are able to capture information about the customers' experience in using products: computer service hotlines, vehicle servicing and washing-machine repairs are valuable opportunities for database building, transforming the initial sale into a relationship and a glimpse of the customer's 'lived experience' with the product, a recognised part of building a relationship between consumers and brands.[8] Certain products themselves imply loyalty — the current account being a well-known example. In retail databases statistical methods, such as the approach developed by Dominique Crié[9] in France, are used to identify loyalty-generating products in cases where they are not self-evident due to volumes of product codes.

Psychographic Variables

Observed data about the relationship between the company and the customer

More recently, customer relationship management (CRM) systems provide a means of capturing descriptive data about the relationship itself. Call centres, direct marketing campaign management systems, salesforce contact management databases, e-mail, websites and interactive television are means of uploading information useful in measuring the quality of the direct relationship with the client. For example, recording 'commercial pressure',[10] either from the company to the customer or contacts from the customer to the company, measures mailing response, but also identifies 'passive' from 'active' customers and lights up warnings where a customer who has not changed address has received a liberal dose of mailings, but has never responded. These 'observed relationship' variables can be highly predictive: in a model for a bank, 'date of last visit to the branch' was one of the key variables identifying potential savings customers. Similarly, the cumulative amount of time the customer service unit spent on the telephone with a client or the number of contacts made by the client over a given period appear as discriminate variables in attrition models for telecom companies.

Observed relationship variables

Socio-demographic profile data

Whether obtained from lifestyle overlays, collected directly from customers on order forms and insurance contracts or derived from first names,[11] socio-demographic profiling data combine with behavioural information to amplify predictive models, increasing the difference between 'best' and 'worst' performing customer groups through interactions between independent variables. Socio-demographic data often prove to be sufficiently independent of behavioural data to create powerful interactions and provide better models than if, for example, only

Customer potential and attitudinal data most relevant to relationship marketing

RFM indicators were used. In the aforementioned example of the bank, 'date of last visit' actually combined with age of customer and current deposit account balance to reveal the most predictive strata of prospective savings-plan holders. Socio-demographic data have intrinsic virtues for relationship marketing, and particularly for the marketing opportunities arising from the customer life cycle — the most obvious examples are major life transitions such as entry into or changing professional activity, formation of couples or family, retirement and death.[12] In the age of CRM, we should remember how important these life stages are in developing relationships, and pay even more careful attention to their inclusion in the database. However, many companies still overlook the potential of such information for marketing and miss opportunities to collect the data in more pertinent ways.

Appropriate segmentation

Most French companies use the official Institut National de la Statistique et des Etudes Economiques (INSEE) categories when collecting occupation profiles, and while this is statistically relevant, it is often more meaningful to the customer to ask questions more appropriate to the context. For example, a private health-plan company (*mutuelle*) should include such categories as independent tradesman, as in France such people do not have the same insurance cover as employees. A further example of appropriate segmentation in loyalty programme questionnaires is shown by the oil company Total including high-petrol-usage job categories, such as ambulance or taxi driver. Even events which are not direct sales opportunities are useful indicators of changing consumer behaviour. For example, the birth of a child for an auto-centre customer, detectable through the purchase of a child seat, may also precede the purchase of a new car under manufacturer guarantee — bad news for the independent car-maintenance network — or a change in habits when the former home-mechanic no longer has time for DIY car repair.

Customer potential

Customer databases increasingly include variables used to estimate the potential spend of a household or company in a particular market. 'Share of customer'[13] is calculated by comparing the progress of actual purchases against the theoretical budget of the client. In France this is called *taux de nourriture* or 'rate of nourishment'. Insurance companies calculate the theoretical budget for the household by extrapolating from research based on indicators such as place and type of residence, number of children etc. An international transport company based in France asks their customers to declare the approximate numbers of international parcels shipped, and uses the comparison between this declared total and transactional information to segment their customer base and differentiate their commercial actions. Even though the absolute value of such declared data is not necessarily usable for relationship marketing, the relative positions of customers on the resulting scale of penetration are significant. Along the same lines, retailers ask customers how many different stores or types of store they frequent, this information proving more reliable than declared data regarding purchases among rival product brands.[14]

Internal states of the customer

Psychographic variables

These 'psychographic' variables have the major advantage of being useful at face value in marketing communication and product and brand positioning. In integrating these variables into the marketing database and combining them with behavioural, customer potential and socio-demographic variables, they gain predictive power, becoming key components in dual-objective segmentation:[15] identifying segments with economic leverage, where customers are described in terms which can be used in marketing communication. The gardening case study presented in this paper provides an illustration of how customer segments of varying growth rates and spend are defined using attitudes to gardening and gardens, directly actionable in an agency brief.

Bundled into the 'psychographic' category are different dimensions of attitudinal, motivational and situational variables long established through market research as different and often independent concepts. As such, the authors believe they should be explored individually as candidates for database variables. Different constructs will come into play in different sectors of activity, and managers should consider how each concept can be transformed into relevant variables for their business. Particularly useful qualitative dimensions include benefits sought, involvement, commitment, satisfaction/dissatisfaction, relationship quality,[16] traits of personality, and the academic definition of attitudes (beliefs and feelings about the product category or brand which affect the way the customer behaves). Numerous publications are available discussing each of these constructs and it is not the objective of this paper to 'reinvent' them, particularly as in many cases teams of academic researchers have spent decades developing scales and models which can be publicly consulted. The difficulty is that where scales exist, they have been developed for research purposes, and even those used by practising market research companies can contain batteries of dozens or even several hundred questions. If we are to use these attitudinal dimensions in CRM, the data must be available across the entire database, or most of it at least. This means we must develop methods of identifying only the most pertinent information — that which is entirely complementary to the data already held, and which will be most efficient both in economic modelling and in providing the content for powerful relationship marketing communication.

Three groups of customers

Integrating 'missing' dimensions

Many companies are still lacking some of the behavioural dimensions in Figure 1, such as observed data about the relationship between the company and the customer, or have not pursued product category data as far as identifying loyalty-generating products. New behavioural indicators such as these can increasingly be found in-house through CRM information flows, but as new data become available, the content of the marketing database is not always revisited. The authors believe managers should evaluate the data they need for effective relationship marketing and be proactive about gleaning these indicators where possible. The majority of missing dimensions, however, consist of customer potential and attitudinal data. These are the very categories most relevant to relationship marketing, both in terms of identifying economic potential

Customer potential and attitudinal most relevant to relationship marketing

and applying the 'share of customer' principles, and in terms of generating meaningful communication at an individual level.

Segmentation based on transactional and demographic data allocates customers into groups with overall similarities in economic terms, but this homogeneity does not apply if the relationship marketing leverage is to operate fully. For example, in a group of customers living in similar houses with gardens, some will enjoy gardening more than others, have different amounts of spare time to devote to their gardens, have different budgets, etc, even though on average the members of this cluster will be much 'closer' to one another in terms of economic profile than they are to a separate group of customers living in apartments.

Low-budget survey

It is not suggested at this stage that existing segmentation schemes, on which budgetary decisions are based, should necessarily be thrown out in favour of new, complex models combining all the dimensions described in Figure 1. First, 'missing' dimensions are usually integrated in stages by direct qualification, psychographic overlays (not yet available in France) and/or modelling, which means that the new schemes would not apply to the entire database. Secondly, not all economic segments will justify investment in 'qualitative' relationship marketing data. Thirdly, most companies engaged in database marketing for several years benefit from a learning curve derived from past investments in marketing action and analysis of flows between segments. Marketers are reluctant to risk taking steps backward. However, the authors do believe that marketers should at least be actively working on adding a 'second layer' to their segmentation schemes. Often the problem is one of breaking up borderline segments in which massive quantities of 'average' customers accumulate, representing a good deal of potential if only they could be stimulated. Relationship marketing techniques ought to be effective in such segments. But databases rich in transactional and demographic data are not fully 'armed' for relationship marketing practice until layers of attitudinal data are integrated.

There is already debate as to the opportunity of, and methods for, integrating attitudinal data into marketing databases. The authors believe each business should make its own decision as to the relevance of acquiring missing dimensions of the customer relationship and the methods to be applied. The case study below results from recent work carried out for a chain of garden centres in France. In this case, certain segments of a customer base were identified as requiring customer insight on an individual level to solve a business problem (generating store traffic among members). The objective is to demonstrate that in cases with clear business objectives, such as this, it is possible to identify key variables that are worth the effort of seeking. Furthermore, these qualitative variables can be employed as a complement to existing relationship dimensions in positioning direct and interactive communication.

Case study

Background

The context of this practical illustration is a customer loyalty scheme (business to 'many consumers') where all new members fill in a questionnaire. This presented a clear opportunity for completing the missing dimensions in Figure 1.

A chain of garden centres in France has been operating a loyalty scheme for more than four years and currently has over 200,000 members. The marketing database has been operational throughout this period, and historical and membership questionnaire data on 100 per cent of members are available.

At the beginning of the third year of the programme, a cluster analysis was carried out using the information from the membership questionnaire and aggregated sales data. The template in Figure 1 was used in deriving the initial segmentation scheme. Raw data were introduced and variables created representing as many of the dimensions in the customer relationship as were available at the time the analysis was carried out. The resulting model formed on socio-demographic, customer potential and behavioural components.

Among different segments produced by the model were three groups of customers, A, B and C, living in houses with gardens.

Three groups of customers

— Group A: The best clients, with high sales levels and a high level of declared loyalty to the store.
— Group B: Medium customers turning out to be good sales performers, some of whom migrate into group A.
— Group C: Many recently joined members, usually of a similar habitat to group B but with a lower level of declared loyalty to the store and fewer and sporadic purchases, despite their membership of the loyalty scheme.

After this segmentation scheme had been in use for about a year, groups B and C, each representing approximately 30 per cent of the client base, prompted a quest for attitudinal data. In theory the customers in both groups had potential for the garden centres, based on their responses to the membership questionnaire.

Despite attempts to attract more business from group C customers, campaign results remained 'RFM' in nature — the best customers continued to buy more and respond better than the lower strata largely independently of mailing content, much to the exasperation of the garden centre marketers. They decided to use a second layer of attitudinal data to break up group C and learn what might be driving the successful groups A and B (Figure 2).

The objective was to identify subgroups with differing economic leverage, using variables which would also be useful in differentiating the tone and content of direct marketing campaigns. In a market invaded by 'me-too' promotions and rival loyalty schemes, personalised communication in phase with the customers' own view of gardening, using attitudinal variables to which competitors do not have access, seemed a reasonable tactic.

The initial data-gathering survey

In addition to qualitative research, Figure 1 was again used as the inspiration for designing a questionnaire which would produce attitudinal and motivational variables for analysis in combination with existing data.

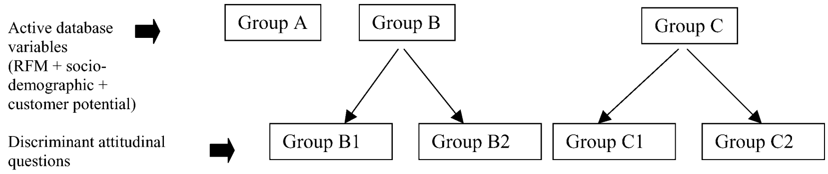

Figure 2: Breakdown of groups B and C

Low-budget survey

A low-budget survey was carried out using a self-administered questionnaire containing 17 questions with batteries of items, mailed to 3,300 customers randomly selected from the targeted groups A, B and C (C customers were filtered to include members who had made at least one transaction). Quantities were weighted, as it was expected that the less loyal B and C customers would not respond as well as the A group. The response rates actually differed less than was expected (Table 1): the Bs had almost as much to say as the core group A. Response was motivated by a gift of a plant to be collected at the store, and customers were promised a 'gardener's profile' which was sent a few weeks later. Subsequent customer feedback showed this initiative to be highly appreciated, and 626 of the 658 responses returned were usable. These responses were merged with database data. (By research standards 626 is a relatively large population, although by database marketing standards it is minute. Budget and time constraints meant it was not possible to gather more information at this stage in this case. Attention was given to significance testing.)

Predictive or descriptive analysis?

Highly personalised relationship marketing is expensive, and the authors do not believe it should be applied regardless across an entire customer base. In this case the researchers were looking for variables which would be useful for relationship marketing communication and which would also allow the recognition of groups of customers with differing economic potential. This objective placed the study within the scope of predictive modelling. The choice of the dependent variable(s) was therefore an important step in the process.

The dependent variables

The questionnaire response was merged with the marketing database data taken at two different dates, T0 and T1. This allowed customers to be profiled one month before the survey was carried out, and again five months later.

The following dependent variables were calculated:

— amount of customer spend at T0

Table 1: Response to attitudinal questionnaire

	Mailed	Responded	Response rate (%)
Group A	605	150	24.8
Group B	1,567	331	21.1
Group C	1,132	177	15.6
Total	3,304	658	19.9

— amount of customer spend at T1
— evolution in spend between T0 and T1.

'Spend' was defined as average monthly spend: the total amount spent by the customer in the (sliding) 12 months preceding T0 or T1, divided by 12. A threshold of N000FF (the actual sales threshold cannot be mentioned for confidentiality reasons) observed in RFM and segment-to-segment flow analysis was used as being a significant divide between profitable and less-profitable customers.

In this case, T0 was the end of October and T1 the end of April — significant dates for garden centres, as spring sales are well under way by the end of April. The sliding 12-month calculation cancelled the effect of seasonality in sales growth between T0 and T1, except for recent customers (six to 12 months' presence at T0). These recent customers represented 82 per cent of group C and 54 per cent of group B. This partly explains the strong increase in average sales per customer, as some of these households would have been registering their (traditionally stronger) spring purchases for the first time in April.

As could be expected, the survey itself had a positive effect on the performance of the respondents, particularly among group C customers, who seem to have benefited from this attention (Tables 2 and 3). At T0, no significant difference was found in sales performance between targeted customers, non-targeted customers and questionnaire respondents. Six months later, twice as many respondent group C customers had moved over the profit threshold than non-respondent C customers. A 5 per cent increase in the percentage of respondent group B customers versus non-respondents was observed. This common empirical finding is good news for marketers hesitant about questioning their customers. The attitudinal groups which emerged from the analysis were effective at T1 and T0; the impact of the survey on behaviour did not affect the validity of the results.

Identifying the key questions to be retained in the modified membership questionnaire

CHAID tree analysis

CHAID tree analysis was used to 'explain' the dependent (behavioural) variables by the questionnaire response.

The tree analysing sales at T1 against the questionnaire variables proved more discriminate than the analysis at T0 (Tables 4 and 5). This is

Different types of gardeners revealed by tree analyses

Table 2: Performance at T0 for mailed and non-mailed customers

	Group A % customers over N000FF sales threshold at T0	Group B % customers over N000FF sales threshold at T0	Group C % customers over N000FF sales threshold at T0
Did not receive questionnaire	99.6	62.6	22.7
Mailed, did not respond	99.6	60.5	21.4
Responded	100.0	61.7	24.3
Total over threshold	99.6	62.6	22.6

Note: Group A customers were over the threshold by definition. The results were significance tested at 5 per cent and 10 per cent; no significant differences were found.

Table 3: Performance at T1 for mailed and non-mailed customers

	Group A % customers over N000FF sales threshold at T1	Group B % customers over N000FF sales threshold at T1	Group C % customers over N000FF sales threshold at T1
Did not receive questionnaire	97.8	65.9	20.1
Mailed, did not respond	96.9	63.0	30.2
Responded	98.7	69.4	41.6
Total over threshold	97.7	65.9	22.6

Note: All differences are significant for group C at 5 per cent. For group B the difference between non-mailed customers and mailed non-responders is significant at 10 per cent, and the difference between mailed non-responders and mailed responders is significant at 10 per cent.

Table 4: Statistical testing of the comparative performance of CHAID analysis carried out on competing dependent variables

Chi-square test	Model 1: Customer above N000FF sales at T0 CHAID tree with 10 cells DF = 9	Model 2: Customer above N000FF sales at T1 CHAID tree with 10 cells DF = 9
'Good' customer at T0 (above N000FF)	Chi-square = 266 $P \approx 0$	Chi-square = 243 $P \approx 0$
'Good' customer at T1 (above N000FF)	Chi-square = 158 $P \approx 0$	Chi-square = 179 $P \approx 0$

Note: Critical value for this test at 1 per cent = 21.7. The chi-square statistic measures the degree of dependency between the 'good' customer variable and the way the tree divides up the population. Model 1 does prove better at identifying 'good' customers at T0 than model 2: 266 > 243. Similarly, model 2 is better than model 1 in describing 'good' customers at T1: 179 > 158.

Table 5: Fisher's test

Fisher's test	Model 1: Customer above N000FF sales at T0 CHAID tree with 10 cells DF = (9; 616)	Model 2: Customer above N000FF sales at T1 CHAID tree with 10 cells DF = (9; 616)
Durable involvement in gardening	$F = 10.6$ $P \approx 0$	$F = 16.7$ $P \approx 0$
Sales per customer at T0	$F = 99.9$ $P \approx 0$	$F = 99.7$ $P \approx 0$
Sales per customer at T1	$F = 75.1$ $P \approx 0$	$F = 75.7$ $P \approx 0$
Change per customer in sales T1–T0	$F = 1.8$ $P = 6.5\%$	$F = 2.3$ $P = 1.5\%$

Note: Critical value for this test at 1 per cent = 2.4. As these are numeric variables, Fisher's test can be used to synthesise the degree to which the tree explains variance. Model 2 is better than model 1 at explaining durable involvement in gardening and change in sales performance from T0 to T1. Models 1 and 2 are equivalent in explaining sales at T0 and at T1.

probably due to the recent clients whose latent differences in purchase behaviour were being revealed by April.

The tree shown in Figure 3 illustrates how discriminate questions were selected for inclusion in a remodelled membership questionnaire (variable names have been changed).

The choice of questions for the remodelled membership questionnaire focused only on those variables with significant evidence of a relationship to economic performance ($P \leqslant 5$ per cent). Further analysis (not shown) was carried out to investigate the attitudinal characteristics of group A. As this was the group of best customers, they were all over the sales threshold — none had slipped under during the five months after the survey.

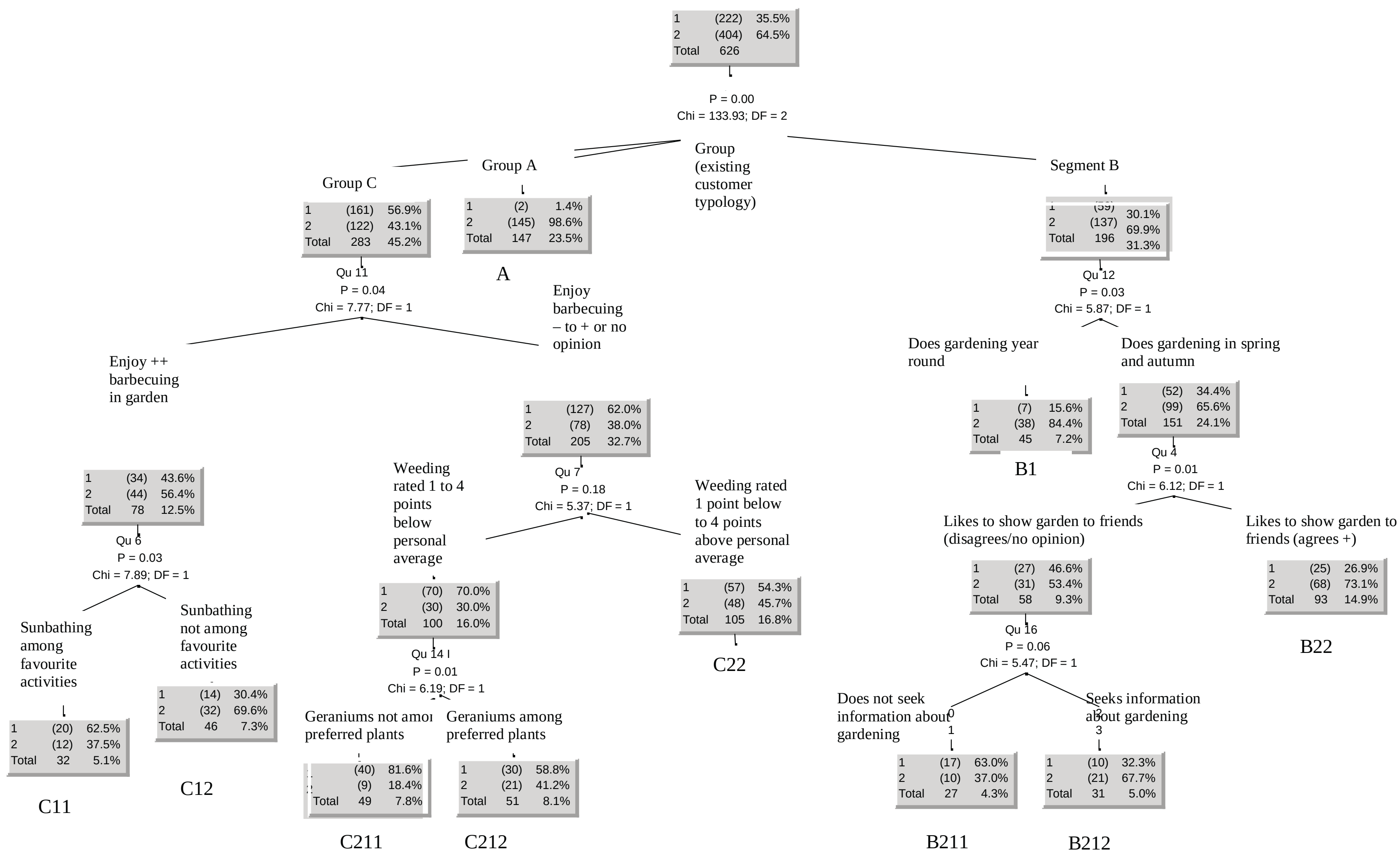

Figure 3: CHAID treee analysis of response to attitudinal questionnaire — Customers under sales threshold of N000FF at T1 (category 1) versus customers over N000FF sales threshold at T1 (category 2). Overall, 64.5% of respondents were over the sales threshold at T1 (end of April).

Analyses in which several trees were produced and used to select discriminating questions showed significant differences in sales performance in terms of purely attitudinal questions directly actionable in direct marketing. In other words, in the high-involvement context of gardening and in the particular context of this chain of garden centres, different attitudes about gardening are linked to behaviour.

For example, in Figure 3 cell C11 has only 37.5 per cent of customers over the N000FF threshold, compared to 64.5 per cent average for the population as a whole and 43.1 per cent for all group C respondents. Using the variable names shown (for illustration purposes), C11 is defined as group C customers who responded to the statement in question 11, 'I especially enjoy barbecuing in the garden', with a high score (6 or 7 on a 'disagree–agree' scale of 1–7) *and* chose sunbathing from a list of favourite activities in the garden. Among the segment C customers who disagree, moderately agree with or have no opinion regarding barbecuing (the other side of the question 11 branch), those who rated weeding lower than other gardening tasks for enjoyment *and did not* choose geraniums from a list of preferred plants had the lowest performance: only 18.4 per cent above the sales threshold, cell C211. On the other side of the tree, among group B customers, year-round gardeners are more profitable customers than spring-and-autumn-only gardeners.

While these attitudinal variables are not needed to identify customers of differing economic value (the existing segmentation and RFM criteria fulfil this role), they have evident face value in developing meaningful communication to stimulate the different subgroups identified in groups B and C.

The tree analyses revealed very different types of gardeners, among which were found:

Different types of gardeners revealed by tree analyses

- keen, year-round, all-weather gardeners with high levels of personal involvement in the activity of gardening, deriving therapeutic benefit from even arduous gardening tasks;
- pleasure-oriented hedonists taking the fresh air, enjoying the spectacle of their garden, sharing the results of their creation with friends and family, and finding pleasure in some gardening activities while rejecting others;
- sleepers, eaters and sunbathers, who do not necessarily like gardening but enjoy the result.

These groups were derived from clear-cut combinations of discriminate questionnaire variables. More insight into the tree cells was gained by cross-tabbing them with other variables of interest (Table 6); significant differences were found which helped the researchers to make recommendations as to how each subgroup should be approached in terms of communication. The following were looked at in particular:

- sales performance at T0 and T1;
- growth in sales from T0 to T1;
- declared loyalty to the store (at the time of joining the programme);
- durable involvement[17] in gardening (using an academic six-item scale developed by Strazzieri[18] and included in the questionnaire).

Table 6: Cross-tabulation with other variables of interest

Cell labels	Average sales per customer at T0 (indexed)	Average sales per customer at T1 (indexed)	Growth in sales T0–T1	Durable involvement score	% customers declaring they do all their shopping at this chain of garden centres
C11	26	57	89	3.2	12.5
C12	27	90	106	3.8	13.0
C211	22	44	144	5.2	8.2
C212	24	61	144	5.2	15.7
C22	28	74	144	5.2	23.8
B1	244	189	156	5.6	57.8
B211	135	73	108	3.9	44.4
B212	206	161	144	5.2	32.3
B22	211	145	131	4.7	54.8
Total	100	100	100	4.8	30.5

Note: Cell labels refer to end cells in Figure 3.

The differences between the cells for groups B and C can be seen in Figure 4. For the higher-spending B cells, personal involvement in gardening rises with sales and growth in sales, whether or not they originally declared that they did all their garden shopping at the store. For C customers, a gap appears between involvement and declared loyalty. The gap is widest among low-spending cells C211 and C212. Put simply, the more involved in gardening customers are, the more they spend at the store, unless they go elsewhere.

What did the results suggest to the garden centre marketers?

The authors' vision of the relationship between involvement and sales among C customers is poor, due to the high proportion of non-exclusive customers. However, once the more highly involved C customers have become more loyal to the chain, a greater share of their gardening spend will be captured and their level of involvement in gardening can be expected to correlate positively with sales (although this cannot be proven from the present dataset). In operational terms, this means that the cells in Figure 4 representing the widest gap between involvement and declared loyalty (high involvement, low loyalty) should represent the strongest potential. The strategic marketing implications for this chain of garden centres are quite clear. The first battle to be fought among C customers falling into high involvement/low declared loyalty cells C211, C212 and C22 is to draw them away from competitors using arguments derived from the customer insight provided by the attitudinal variables defining these cells.

A closer look at the characteristics of the C cells with the widest difference between involvement and loyalty (C211 and C212) revealed these loyalty card holders as enjoying some gardening activities for pleasure and relaxation while rejecting others (variable names have been changed in Figure 3 and other trees, not shown here, were used in completing gardener type definitions). Such customers are likely to pay less attention to messages created for devoted year-round gardeners than they would to a more pleasure- and result-oriented discourse. Up until this point, it had not been possible to differentiate communication along these lines.

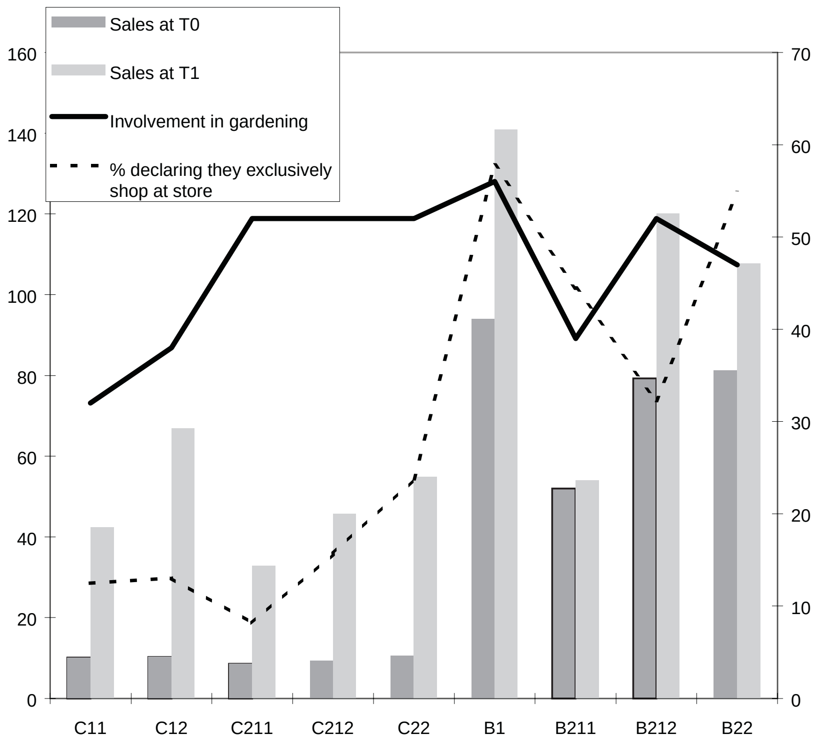

Figure 4: Involvement and sales in groups B and C
Note: Among all customers claiming to shop exclusively at the store, there is a significant positive correlation between spending at T1 and involvement: 0.25 (probability 0 per cent). Among all customers having declared when they joined the loyalty programme that they did not exclusively shop at the garden centres, the coefficient is not significant (0.09, probability 6.6 per cent). Data from competitive garden centres would be needed to establish a relationship between overall sales and involvement in gardening among non-exclusive shoppers.
How trustworthy is customers' declared loyalty? Declared loyalty to a store has been shown to be better linked to behaviour than declared loyalty to a brand,[19] particularly in cases of high involvement, as here. However, it is reasonable to assume that the more recent the information about loyalty, the clearer the relationship between involvement and sales would appear among shoppers claiming to be exclusive. This could explain why the correlation between sales and involvement held out in group B, which contains far less recent members, whether or not the customers claimed they were exclusive. However, the data to hand did not enable further investigations into this relationship between sales, declared exclusivity, length of membership and involvement, and such investigation had not been the original intention.

'Second layer' attitudinal segmentation to personalise messages

Since this work, the membership questionnaire has been modified to include the discriminate attitudinal questions. All new customers, many of whom fall into group C, can be allocated into a 'second layer' attitudinal segmentation. This second layer is used in personalising the content and tone of messages so that relationship marketing leverage can operate.

Individual answers to the attitudinal questions can now be stored on the database and used to position the direct marketing and interactive messages differently and more meaningfully to the customer. In this case, the relationship marketing lever is being used not across the entire base but in selected cells, with the objective of increasing share of customer among those in recent, high involvement/low declared loyalty groups.

In addition, garden centre management used results to compare their own brand positioning against that of competitors, and found that customers belonging to different attitudinal groups are exposed to different competitive brands and different levels of competitive pressure. For example, among keen, year-round gardeners and hedonists, competition is more limited than for results-oriented sleepers, eaters and sunbathers (types C11, C12 and B211 in Figure 3), who are exposed to massive competition from DIY stores and hypermarkets. This approach can be panned out locally, calculating competition density per customer according to which attitudinal group he belongs — a useful by-product of this work for branch operations.

Database marketing implications and limits

This and similar cases show that combinations of attitudinal variables can define groups of differing economic potential. It may seem a foregone conclusion that customers more interested in sunbathing than weeding are less likely to be high spenders at garden centres, but the essential difference between database marketing and a traditional survey is that by including these questions on a membership form the marketer can differentiate between such customers. This information is particularly valuable regarding new members of a loyalty programme, before predictive behavioural data are available, and among the lower RFM strata. Among the higher-spending, more loyal customers, significant attitudinal differences were also found. For some customers the act of gardening itself is important, while others invest in the 'architecture' of the garden. Each of these types is capable of high spending, and while these customers remain active there may be no operational need to differentiate communication to cater for these differences. On the other hand, knowing who these customers are is a strategic weapon when it comes to warding off attacks from competition.

Although links were found between attitudes towards gardening and purchase behaviour in this case study, the authors do not claim a cause-effect relationship. The apparent better performance of more highly involved gardeners is probably due to the current positioning of this chain: high-performing customers could well be among those whose attitudes and values are in synergy with the projected image. Although this makes the task of conversion more difficult among customers whose attitudes are furthest from the current image, one of the major advantages of direct marketing is, of course, that it should technically be possible to modify the stores' positioning to coincide with the ideal projected role the customer would have them fulfil. In the present case, management is more inclined to drop promotional efforts among the groups farthest away from the projected image, which are also those exposed to the most intensive competition.

Direct qualification is possible more often than is assumed

An obvious limit to the use of attitudinal data in direct and interactive marketing is the question of obtaining them across the board. Presented here is a case where a membership questionnaire has been optimised to allow direct qualification of all new members. The authors believe that direct qualification is possible more often than marketers assume. In this case the number of discriminating attitudinal questions actually turned out to be less than the 'ten key questions' marketing managers envisaged, which facilitates data collection. Five or six questions (some of which contain items) will fit into the available space in a fulfilment questionnaire, an order form or a customer contact script.

Regrettably, many marketers are still printing hundreds of thousands of membership applications or subscription forms to be completed by entire populations of customers, with little consideration for the potential relationship marketing power of these questions. Space is wasted with such questions as first names and dates of birth of children, which are not justified by the commercial context, or in collecting socio/professional data which are not directly actionable. A simple parallel is that of credit scoring: once it was discovered that certain geodemographic questions had to be asked in order to do business, the credit companies went about getting the answers. Relationship marketing is no different: if the very premise of efficient direct communication is that pertinent content and attitudinal (or any other) data do prove vital in generating meaning, then it is worth making the effort to collect them.

Of course, there will be cases where it is not possible to collect attitudinal data, or where it is considered too expensive due to data-entry costs or lack of simple direct contact with end customers. In this case study, data will be collected for new customers, but how can existing customer profiles be enriched? Certain segments have been identified as priorities and incentivised questionnaires will be sent out, but not everyone will respond. Thus, modelling and psychographic data overlays should be considered. In France, psychographic overlays are not yet available, so the question is then whether it is possible to model the attitudinal groups from the database.

The dataset obtained from the survey described in this paper did not allow the successful modelling of the attitudinal cells from transactional data. The closest the authors were able to approach modelling, eg levels of involvement, was links to the geodemographic questions from the original membership questionnaire. Further attempts will be made when sufficient numbers of customers become available. However, whether a model potentially exists or not, attitudes cannot be modelled from behaviour among customers for whom little or no behavioural data are available (eg low spenders, recent customers). Relationships have been found between attitudes towards a brand or store and observed purchase behaviour[20] in situations of high involvement and high perceived risk, where competing products and brands are perceived as being different and where customers are relatively sure about their opinions. The authors believe that geodemographic variables will prove more reliable in projecting attitudinal groups than behavioural database variables, especially as this is the principle on which psychographic overlays currently operate.[21]

A good deal more work needs to be done on understanding the relationships between the dimensions of the customer relationship outlined in Figure 1. By definition, if behaviour and, say, personality traits are independent dimensions, they are orthogonal — one cannot be projected from the other.

Behavioural variables are powerful descriptors and predictors of economic value

The authors believe these dimensions are complementary and serve different purposes in segmentation and relationship marketing practice. Behavioural variables are powerful descriptors and predictors of economic value. Customer potential indicators, generally based on geodemographic indicators (or the equivalent in B2B, such as number of employees, activity. . .), are useful in deciding which segments to stimulate. Internal psychographic variables are, in the authors' opinion, central to empathy, the nirvana of relationship marketing, where the customer's own perspective of the brand or product category is incorporated into the communication to maximise meaning, pertinence, memorisation and, ultimately, affinity with the brand. If the objective of the relationship marketing database is to allow marketers to optimise their actions, then it should cover these dimensions and we should understand how they work together, how they should be weighted, which indicators lie behind them, and in what circumstances and combinations they should be used.

We must also recognise the limits of the marketing database, however 'rich', and face several difficulties.

First, relationship marketing operates at an individual level, whereas many marketing databases are structured at household level. This means building and qualifying an individual level where this does not exist, or flushing out an embryonic individual level. In the case of business-to-business databases, where individual contact levels already exist, people move from one company and one job to another. They may not stay in one position long enough to qualify for a database, whereas a sales representative will use the psychographic profile which he has spontaneously built in his head. In France and Belgium, companies are developing tools to enable a salesperson to build an 'objective' psychographic profile of a face-to-face or telephone contact and use this to adapt a sales pitch.[22]

Variables should be chosen for their validity over time

Secondly, occasions to update declarative information are rare. For this reason variables should be chosen for their validity over time. For example, in the test questionnaire, the scale chosen was that developed by Strazzieri,[23] which measures durable involvement, as opposed to other scales which do not take into account the time factor. The case study also questions the stability of 'declared loyalty' over time. On the other hand, personality traits have been shown to be durable and many of these are significant to brand/consumer relationship marketing — for example, propensity to delegate in the case of a car-servicing network, or vulnerability in the case of insurance. Even with careful selection of qualitative data, socio-demographic profiles and attitudes change at major life-cycle transition stages (formation of couple, children, moving house, changing job, divorce, retirement), which in itself justifies tracking these events in the database to generate a new questionnaire.

Thirdly, asking relevant questions with the intent of improving the quality of communication and service contributes to the reinforcement of the relationship itself. Answering questions makes the customer think about the answers, and also about the brand. On the other hand, asking questions creates expectations: only companies ready to use the information in improving the quality of the relationship through more pertinent communication and better customer service should venture into direct data enrichment.

Conclusions

Many companies now have open-structured data warehouses shared by different departments. Database and direct marketing have largely contributed to this move, merging multisource transactional data to incorporate the customers' point of view long before interactive marketing and CRM came of age. Now that computer systems experts have established power and a lifetime occupation in managing customer information flows, it is surely time for the marketer to reconsider the adequacy of the actual data. In addition to looking inwards, 'gathering as much data as possible in the hope that a meaningful picture will emerge',[24] marketing professionals should be looking up from the depths of their data mineshafts and outwards towards their customers for variables which can be used pertinently, both in modelling and in communication. The authors have proposed a template for rethinking the content of a marketing database which they have found useful both in benchmarking existing data sources and in generating new ideas for data enrichment in meetings with marketing and sales teams. They have attempted to show through a case study that it is possible to identify discriminate questions which can be used both in segmentation and in positioning relationship marketing communication. Finally, they have called for direct and interactive marketing practitioners to contribute to the advancement of relationship marketing theory and practice through serious investigation of the interactions between the variables underlying the dimensions of the customer relationship. The authors believe that in the interactive and relationship marketing age, the focus for marketers should be on the quality and pertinence of customer data rather than on systems and functionalities. The customers hold the key to which data can reveal the meaning of their relationships with brands. The current database marketing task lies in detecting, gleaning, capturing and using this information for truly operational relationship marketing, not wholesale and across the board, but within targeted segments of economic potential.

Quality and pertinence of customer data rather than on systems and functionalities

Acknowledgment

The authors would like to thank the reviewers for their helpful comments, which have been taken into account in the final version of this paper.

References

1. In practice, the '80:20' rule is more often found to be a '15:35:50' rule, where the top 15 per cent generate 50 per cent of revenue, the next 35 per cent generate 35 per cent of revenue, and the remaining 50 per cent contribute 15 per cent of revenue or less; this rule is common in retail, for example.

2. Hirschman, E. and Holbrook, M. (1982) 'Hedonic consumption: Emerging concepts, methods and propositions', *Journal of Marketing*, Vol. 46, No. 3, pp. 92–101; Holbrook, M. and Hirschman, E. (1982) 'The experiential aspects of consumptions: Fantasies, feelings and fun', *Journal of Consumer Research*, Vol. 9, September, pp. 132–140; Fournier, S. (1991) 'Meaning-based framework for the study of consumer-object relations', *Advances in Consumer Research*, Vol. 18, pp. 736–742.

3. Petty, R. E., Cacioppo, J. T. and Schumann, D. (1983) 'Central and peripheral routes to advertising effectiveness: The moderating role of involvement', *Journal of Consumer Research*, Vol. 10, No. 2, pp. 135–146.

4. Conference on 'The Marriage of Market Research and Database Marketing', Henry Stewart Conference Studies/*Journal of Database Marketing*, London, 21 June 2000; Webber, R. and Sleight, P. (1999) 'Fusion of market research and database marketing', *Interactive Marketing*, Vol. 1, No. 1, pp. 9–22.

5. Hughes, A. M. (1995) ‘Making a database pay off using recency frequency and monetary analysis’, *Journal of Database Marketing*, Vol. 3, No. 1, pp. 77–89; Kuijlen, T. and Paas, L. (1997) ‘Refining RFM variables using ownership patterns and purchase reasons’, *Journal of Database Marketing*, Vol. 5, No. 3, pp. 221–230.

6. In RFM profiling terminology a ‘11111...’ recency/frequency profile means the individual or household has made a purchase during each period taken into account, such as each year, each season, each month. . . In a 10011 profile the zeros represent inactive customers. Usually the first figure to the left represents the most recent period.

7. Rayner, S. (1998) *Customer Loyalty Schemes, Effective Implementation and Management*, Financial Times Retail and Consumer, Financial Times Business Limited.

8. Fournier, S. (1998) ‘Consumers and their brands — Developing relationship theory in consumer research’, *Journal of Consumer Research*, Vol. 24, pp. 348–373.

9. Crié, D. (1999) ‘Les produits fidélisants dans la relation client-fournisseur’ (Loyalty generating products in the supplier-customer relationship), thesis, Université des Sciences et Technologies de Lille, Institut d’Administration des Entreprises.

10. Commercial pressure is defined as the sum of contacts made, by type of contact and in total over a sliding period such as the past year or the total life of the customer.

11. Desplanques, G. (1986) ‘Economie et statistique’, No. 184, January, INSEE.

12. Micheaux, A. (1994, 1997) *Marketing de Bases de Données*, Les Editions d’Organisation, Paris, p. 56; 191–200.

13. Malthouse, E. C. and Wang, P. (1999) ‘Database segmentation using share of customer’, *Journal of Database Marketing*, Vol. 6, No. 3, pp. 239–252.

14. Dubois, B. and Qaghebeur, A. (1997) ‘Les consommateurs font-ils ce qu’ils disent?’, *Actes Association Française de Marketing*, No. 13, May, pp. 891–919.

15. Forsyth, J., Gupta, S., Haldar, S., Kaul, A. and Kettle, K. (1999) ‘A segmentation you can act on’, *Journal of Customer Relationship Management*, September/October, pp. 145–153.

16. Fournier, ref. 8 above.

17. This is an academic definition of involvement, without the perceived risk factor, comprising pertinence, interest and attractiveness (of gardening to the respondent). Involvement is considered a motivational variable. The scale used consists of six items evaluated on a seven-point Likert scale, resulting in a measure of durable involvement in gardening for each person, with a minimum score of 1 and a maximum of 7. Average involvement scores were calculated for each of the cells in each of the CHAID trees produced. This information proved invaluable in interpreting the results and developing copy strategy.

18. Strazzieri, A. (1994) ‘Mesurer l’implication durable vis-à-vis d’un produit indépendamment du risque perçu’, *Recherche et Applications en Marketing*, Vol. 9, No. 1.

19. Dubois and Qaghebeur, ref. 14 above.

20. *Ibid*.; Poubanne, Y. and Chandon, J.-L. (2000) ‘Attitudes envers les marques et achats passés, force de la relation et rôle de l’implication’, in Proceedings of Association Française du Marketing, 16ème Congrès International, Montréal, 18–20 May, pp. 313–326; Infoworks (1999) ‘Attitude-based targeting: The myth and the reality’, *In the Works*, Vol. 1, No. 1, htm infoworks-chicago.com.

21. Webber, R. (2001) ‘Integrating psychographic variables: The marriage between research and database marketing’, in Proceedings of Semaine de la Relation Client, séminaire P3, bases de données, Paris, La Défense, Exposium, 26 January.

22. Thonard, A. and Marchal, S. (2001) ‘Le client est-il réellement roi?’, in Proceedings of Semaine de la Relation Client, séminaire P3, bases de données, Paris, La Défense, Exposium, 26 January.

23. Strazzieri, ref. 18 above.

24. Bray, P. (2000) ‘A jargonaut’s guide to the world of CRM’, reprinted from the *Daily Telegraph* in *Direct Marketing International*, June.