DEPARTEMENT BEDRIJFSECONOMIE

Computer Simulated Shopping Experiments for Analyzing Dynamic Purchasing Patterns: Validation and Guidelines

Katia Campo*
Els Gijsbrechts**
Fabienne Guerra***

WORKING PAPER
98-262
March 1998

D/1998/2263/2

* Postdoctoral Fellow of the Fund for Scientific Research - Flanders (Belgium), and UFSIA, Prinsstraat 13, 2000 Antwerp, Belgium
** Associate Professor, UFSIA, Prinsstraat 13, 200 Antwerp, and CREER, FUCAM, Chaussée de Binche 171, 7000 Mons, Belgium
*** Professor, CREER, FUCAM, Chaussée de Binche 171, 7000 Mons, Belgium

Acknowledgments:
This research was possible thanks to a research grant made available by the Belgian Science Policy Department to the research center CREER (FUCAM) (PAI n° 03-53), and a special research grant from the University of Antwerp, UFSIA.

Introduction

Computer simulated experiments have gained increasing attention from marketing researchers interested in studying dynamic buying behavior. Recent years have witnessed the development of software programs enabling consumers to engage in 'virtual shopping', that is, to repeatedly purchase items from a computer screen reproducing store shelves (the 'virtual store') under highly controlled conditions. Besides simulating the consumers' shopping behavior, these software packages typically allow to collect additional information from consumers through computerized questionnaires. Computer simulated shopping programs have been developed in various degrees of sophistication, and are continuously evolving (Burke et al, 1992, Burke, 1995, Burke, 1996, Cohen and Gadd, 1996). According to Burke (1996), they offer a viable alternative to other marketing research tools currently employed:
"...most marketing research techniques are, by and large, sadly outmoded. The tools most marketers employ are too expensive, vulnerable to observation and manipulation by competitors, contrived and unrealistic, or simply incapable of providing the information managers really need....Virtual shopping simulation can help managers make tactical decisions in areas such as new products and promotions, packaging and merchandising. Ultimately, the tool promises to change the way companies innovate and how they approach a variety of strategic issues that range from entering new markets to responding to a competitor's attack".
Computer simulated shopping experiments could offer 'the best of two worlds' for modeling consumer purchases over time. As far as other laboratory tests are concerned, Computer Simulated Shopping Experiments (CSSE) offer more realism (compared to 'rudimentary' paper and pencil tests) or more flexibility at a much lower cost (compared to simulated test markets in laboratory stores). In comparison with real life test markets or even controlled field experiments, CSSE offer more control over extraneous variables and more flexibility in manipulating marketing variables and treatments, without exposure to competitive reactions or adverse consumer or distributor effects, and this at higher speed and lower cost (Burke, 1996, Cohen and Gadd, 1996, Brucks, 1988).
While recognizing these advantages, though, scientific scrutiny dictates that a critical view is in order. CSSE remain an artificial environment, and this artificial setting entails potential biases and threats to validity that should be carefully examined. Previous research on the issue is relatively scarce, though. An interesting study by Burke et al in 1992, provides a first evaluation of the validity of CSSE for analyzing dynamic purchasing behavior. In a small scale study, the authors examine potential biases in CSSE results at different stages of the buying process, and hypothesize links with experimental features. A recent article by Cohen and Gadd(1996) provides further insights into characteristics that enhance the realism and the attractiveness of virtual shopping programs to respondents[1].
The present paper intends to shed further light on the issues of CSSE validation and design. It builds upon the work by Burke et al, and extends it in several ways, as indicated in the

[1] Note that the study of Cohen and Gadd does not analyze behavior dynamics, but concentrates on experiences of consumers having gone through one virtual shopping trip.

next section.

Previous Research and Purpose of this paper

While a number of authors have used computer simulations to generate dynamic purchasing data, research on the validity of these outcomes is scarce[2]. Burke et al. (1992) report on a first systematic evaluation of the quality and 'realism' of computer simulated multi-period buying data. In their study, Burke et al. compare the data obtained from a sample of 16 respondents participating in a rudimentary and a more realistic laboratory shopping experiment, to their actual purchases under identical marketing conditions, over a seven-month period. The authors postulate a number of threats to external validity, resulting from artificial experimental conditions. They suggest that in the alternative-evaluation stage, respondents' familiarity with the shelves, and the simulation's capability to reproduce cues on which actual consumer decisions in the category are based, will strongly affect the realism of the simulation outcomes. Similar observations are made by Burke(1996), and by Cohen and Gadd(1996), whose pilot study of a 3-D modeled virtual store emphasizes the importance of store familiarity and availability of sufficient product and promotion information. In the purchasing and choice stages, computer simulated results may be distorted by the lack of involvement of other persons in the choice process, the absence of time, budget, or space constraints, and respondents' tendency to exhibit socially desirable behavior. Finally, the time compression factor (simulated purchases are typically made over a short time span) and lack of actual product consumption may affect dynamic purchasing patterns. The authors find that their more sophisticated computer experiment provides significantly better results than the rudimentary test. Yet, even in their more realistic computer experiment, the above mentioned threats affect consumers' purchase quantity, reaction to promotions, and purchase dynamics. Respondents tend to buy larger quantities in the computer experiment compared to actual purchase data, especially when the purchased item is on promotion. Variation in choice behavior is also substantially lower in experimental than in real purchases. Burke et al. conclude:" ...we hope our study can provide some practical guidelines for subsequent larger-scale validation studies...". In a recent HBR article, Burke(1996) briefly refers to an additional validation study of dynamic virtual shopping. In this study, 300 consumers were invited to make 6 shopping trips in a simulated computer store. These purchase data were compared to UPC scanner information obtained from a concurrently running conventional test market. Simple comparisons of consumers' reactions to package size and price manipulations in the virtual store with conventional test market-effects pointed to similar results. Also, simulated market shares for various brands of the tested product categories (cleaning products and health and beauty aids) closely matched scanner market shares. The author concludes that "the performance of the virtual store was particularly impressive given the number of ways in which the real and computer generated stores differed", yet emphasizes the need for "future benchmarking of virtual shopping results against existing methodologies".

[2] Obviously, the literature on the design and validity of laboratory experiments in general is extensive (for a recent overview on choice experiments, see Carson et al, 1996), and provides us with very relevant basic insights.

In this paper, we follow up on the suggestions of Burke(1996), and specifically Burke et al.(1992), to provide additional in-depth evidence on the validity of CSSE, and the interaction between contextual and experimental features and CSSE outcome validity. We extend their work in several ways.

First, the CSSE analyzed in this paper is comparable in nature and degree of sophistication to Burke et al (1992)'s 'realistic computer experiment'. Yet, following up on the findings and suggestions by previous researchers, we altered the experimental setting and added some features to the software that may improve the results.

Second, the study of Burke et al (1992) is a pilot study, comprising an in depth analysis of a small number of households, for which actual and simulated purchases are 'matched'. This paper takes a different approach and opts for a between-subjects design (Carson et al., 1996), as suggested by Burke(1996). More specifically, we collect computer simulated shopping data from a large sample of households, and confront the results with information from an equally large panel receipt data set. The magnitude of our sample base allows us to shed further light on the external validity of CSSE outcomes, including method-specific household selection biases.

Finally, our study intends to offer additional insights, by examining aspects not dealt with in previous research. Our data base helps to clarify the relevance of CSSE data for comparisons of promotion types. Also, we investigate the impact of design variables not explicitly studied so far, such as the number of periods and the completeness of the on-screen assortment on the validity of CSSE data.

Description of the CSSE experiment

The software for the CSSE experiment was developed in the programming language TenCore (Computer Teaching Corporation Champaign, Illinois USA), and runs on any 'normal' PC (386 or higher). The experiment involved four product categories: jam, cereals, paper towel and margarine[3]. The product categories selected contain a 'reasonable' number of items in their assortment, are regularly promoted, and are frequently purchased by a majority of consumers. Given the high purchase frequencies, consumers can be expected to be fairly 'familiar' with items in the categories. Items in the product classes are typically chosen on the basis of 'cues' that can be visualized, such as brand, pack, price, promotion, and product information. In this paper, we report on validation results for the jam category, for which scanner receipt data were made available by the supermarket chain that served as a model for the virtual store development. The program comprises three modules. The 'pre-purchase' module involves some preliminary data collection on respondent's (household's) background characteristics as well as product category information (use rate, experience, and consideration set) and respondents' reactions to propositions from published Variety Seeking Tendency Scales. Including these propositions not only yields valuable information on consumers' variety seeking tendency, but also helps to avoid "mere-measurement" impacts of consideration set elicitations on subsequent item choices. Module two comprises the purchase simulation, where respondents are given the opportunity to purchase items from the screen. Using scanned pictures of real products, actual shelf

[3] Purchase data of the 4 product categories were collected and used for an analysis of individual differences in promotion sensitivity (see Campo, 1997).

layouts are imitated on the screen. To make the task as realistic as possible, the actual choice context (assortment, shelf layout, prices) of a large supermarket chain was reproduced in the computer simulation. Respondents can zoom in on an item, retrieve product information printed on the (actual) product package, or select one or several items for purchase from the fictitious shelf. Each respondent makes purchase decisions (including a 'no-purchase' option) over twelve successive weeks. Product prices are displayed on a price bar below the shelf. In the course of their shopping trips, consumers are confronted with in store-promotions of different types (coupons redeemable on the next purchase, direct price cuts, premiums, or extra amount), or with shelf space adjustments (change in the number of shelf facings). Promotions and shelf space manipulations occur for two test items in each category, and are randomly assigned to periods and consumers. No promotions occur in the first two (initialization) periods, and the promotion frequency is in line with the normal occurrence of in-store promotions in the category. Promotions are visualized by using (realistic) on pack labels in combination with colored shelf tags. Additional cues displayed at the bottom of the screen are the number of available coupons (when a couponed item was purchased on a previous occasion), and the current household inventory level. The latter was updated weekly using information on the household's reported average consumption rate (module 1), and previous purchase decisions. For margarine and paper towel, all items available in the real supermarket outlet are included in the experiment's assortment. For cereals and jam, a limited assortment of 22 items was selected[4], including the most 'popular' product variants (tastes) and major national brands, in addition to distributor brands and generics. Module 3, finally, collects supplementary data on item preferences, product perceptions, and choice tactics used. The program was extensively tested on a sample of housewives to check for clarity of the questions, and visibility as well as realism of the shopping simulation.

In the actual data collection stage[5], shoppers were intercepted in the store and asked to participate in the computer experiment. Interviews were conducted in five outlets of the supermarket chain that had served as a 'model' for the shelves presented in the shopping module. Information on the household's consumption rate - collected in module 1- was used to select two product categories (a hedonic and a functional category) on which the respondent would be interviewed. Respondents were only assigned categories for which their reported household use rate was at least one pack per month. Respondents accompanied by other persons were asked to cooperate in completing the experimental task, as they would do in real life shopping. In total, 1129 respondents were interviewed. Of these, 156 persons did not complete the purchase simulations, because of too low use rates. Of the remaining 973 interviews, 41 were omitted from the analysis because of low quality (based on interviewers' notes made during the purchase simulation, or obvious inconsistencies in the data). The remaining interviews were fairly evenly spread over product categories and promotional conditions, resulting in between 400 and 500 respondents, and between 3000

[4] Out of 29 items for cereals, and 57 items for jam

[5] The interviews took place in spring 1995, and were carried out on Thursday, Friday and Saturday of 16 successive weeks.

and 4000 actual purchase occasions[6], per category.

The CSSE used here is comparable, in degree of sophistication, to the 'realistic' experiment described in Burke et al(1992). Yet, the approach possesses distinct features in the software design and interview setting. First, while Burke et al. seem to offer complete assortments in all product classes (eventually covering several screens), this is not true for two out of four categories in our experiment. As pointed out by Cohen and Gadd(1996), including complete assortments in virtual stores will be practically infeasible for many categories, yet restricting the consumers' choice set may well affect results. Second, a wider range of promotion types are examined than those covered in previous CSSE research. Third, a specific feature of our program is that respondents are offered weekly information on their current inventory level. No budget constraint per shopping trip was imposed since consumers, having to purchase from only two categories per trip, were bound to experience such a constraint as highly artificial[7]. Fourth, besides the possibility to zoom in on item packages (front view), our experiment includes the additional option of consulting detailed information (e.g. on composition and perishability) on items on the shelf, a possibility not offered in many of the previously tested CSSE. Fifth, like Burke et al.'s program, our software is specifically designed to study dynamic behavior in a time compressed, multitask setting. It is not clear how often consumers are given the opportunity to purchase in Burke et al's experiment: our impression is that the timing of screen presentations matched their actual purchase timing. In our experiment, consumers are confronted with the shelves on a weekly basis, and may or may not purchase during that week. Time (week) indicators are provided, and shopping trips are separated from one another using 'distracting' screens showing a store view. Finally, respondents in our study are intercepted in the store they usually shop in, and interviews are immediately conducted within the retail outlet. This guarantees consumer familiarity with the virtual store shelves, and ensures that consumers find themselves in a noisy, stimulating, natural environment - issues revealed to be particularly important in creating an atmosphere of reality (Cohen and Gadd, 1996). As indicated below, these differential features may affect the outcomes studied in earlier research, and open up new possibilities for analysis. The next section provides an overview of our research hypotheses.

Research Hypotheses

As stated earlier, a *first* objective of our research is to shed more light on CSSE's *capability to 'reproduce' true consumer purchasing behavior*. Burke et al.'s (1992) validation study indicated that purchase quantities, choice shares, and choice dynamics more closely reflect actual buying behavior in realistic than in rudimentary computer experiments, but can still

[6] Weeks in which at least one item from the category was bought.

[7] In fact, this expectation was later supported by the findings of Cohen and Gadd (1996), who imposed a fictitious budget constraint based on the household's 'normal spending' in the product category. In a follow-up questionnaire, many respondents indicated that the budget constraint was too restrictive and highly unrealistic.

be biased. In accordance with Burke et al.'s guidelines, we introduced a number of new features into the simulation program that would hopefully improve the realism of respondents' buying decisions. First, inventory information is provided and interviews are conducted in-store, to reduce purchase quantity biases due to lacking budget and time constraints. Second, the simulated shelf layout (item position and number of facings) closely matches actual shelves of the store in which respondents are intercepted and interviewed. The familiarity with CSSE shelves and assortments, combined with the 'realistic' interview environment, is expected to enhance the validity of respondents' choice decisions. Third, to reduce time compression effects, successive simulation weeks are separated by pictures indicating the start of a new shopping trip, and week numbers are displayed on screen. With the same objective of motivating respondents to take dynamic buying behavior characteristics into account, switching is given as an example in the purchase simulation instructions (to clarify the task of imitating actual buying behavior for 12 fictitious weeks). Expected implications for the validity of experimental purchase data are discussed below. Hypotheses 1 to 3 investigate the external validity of experimental shopping trips disregarding promotional effects, that is, when promotions are either absent or separately accounted for (by removing their effect on the test variables). In turn, propositions are formulated on the purchase quantity, brand choice, and purchase dynamics (degree of purchase variation) within a product category. Hypothesis 4 relates to consumers' differential reactions to in store promotions in a CSSE setting, where the effect of promotions on purchase quantity as well as brand choice is considered, and different promotion types are compared.
Our *second* objective is to *link experimental features to external validity of the outcomes*. Hypotheses 5 and 6 deal with issues related to this objective, and concentrate on how the number of shopping trips, and the 'completeness' of the CSSE assortment, affect the consumers' experimental shopping (quantity, choice and variation) behavior. Below, we briefly present the hypotheses and associated comments.

Hypothesis 1:
For non-impulse products, and in the presence of inventory cues, the number of units purchased by consumers in a CSSE is in line with actual purchases.
Hypo 1 sub 1: *Consumers' simulated purchases are close to their reported use rate*

Comments:
Previous research suggests that, given the absence of budget, time and space constraints, laboratory shopping leads to above normal purchase quantities. In our CSSE, conducting the interviews in store implies that some time constraint is active. We further believe that, for 'non-impulse' products (Narasimhan et al, 1996) for which consumers want to permanently have some stock available, providing information on current inventory levels helps make the purchasing task more realistic.
Our hypothesis of realistic purchase quantity decisions is based on the assumption that these are made on a 'stock replenishment basis', and that the inventory adaptation calculated by the CSSE is realistic.

Hypothesis 2: *In the context of CSSE, consumers tend to buy more national brands, and less private labels and generics, than in reality.*
Comments: This hypothesis is in line with previous research on the presence of response bias in interviews, and the social desirability to purchase national brands, rather than private labels or generics. Note that the findings may be country- or even chain-specific, and are based on a lower quality perception for distributor than for main national brands. Evidence on this bias was also reported by Burke et al (1992).

Hypothesis 3: *In frequently purchased product categories with which consumers are highly familiar, the degree of variation in purchases in simulated shopping is comparable to that in actual purchases.*
Comments: Burke et al hypothesize that the time compression factor increases the impact of past purchases on current purchases. They also argue that satiation is less likely to occur since respondents do not actually consume products. Their conclusion is that time compression decreases switching, and reduces the number of different brands bought in the category. Contrary to their hypothesis and findings, we do not postulate such an effect in our experiment. First, we reduced the time compression factor and made purchases more realistic by providing time indicators and distracting screens between purchases. Second, we postulate that stronger purchase event feedback can be positive (more loyalty) as well as negative (stronger variety seeking). While the absence of satiation effects could reduce purchase variation, the lower risk of having to actually consume a bad 'new choice' might have the opposite effect. Overall, we therefore expect the degree of variation (measured as entropy) to be realistic, especially for items for which the full assortment was presented.

Hypothesis 4: *Increased visibility of the presence of a promotion causes consumers to react more strongly to sales promotion offers in CSSE than in reality, resulting in larger quantities and more pronounced switching to promoted items.*
Comments: Previous research suggests that consumers react more strongly to promotions in a laboratory setting because of higher promotion visibility, and less environmental 'distractions' than in an actual store. In our experiment, promotions were presented in a fairly realistic manner, and interviews were conducted in the real store environment. Yet, the shelves presented on screen for the jam category were a simplification of reality, which could artificially enhance the attention drawing capability of promotions. We therefore expect a more than realistic impact of promotions.

Hypothesis 5: *For non-impulse products, CSSE can yield valid information on successive shopping trips. Consumers are willing and able to 'reproduce' their actual purchases (quantity and choice decisions) over a reasonable number of subsequent periods.*
Comments: A crucial question in analyzing dynamic behavior in a laboratory setting concerns the number of successive shopping trips that can be presented to the consumer before serious testing or maturation effects (e.g. boredom and fatigue, lack of willingness to spend more time) jeopardize the quality of the answers (Carson et al, 1996). Similarly, one may wonder whether initialization effects occur, that make information on early purchases unusable. In our experiment, we believe that, since consumers were already somewhat familiarized with the pictures in the first module, and could closely examine items before purchase, only the first week would lead to an exceptional sales level (stock

building). Our field experience further suggests that, except for a few cases, most consumers took their task seriously, and were willing to carry the experiment through to the end[8].

Hypothesis 6: *For non-impulse products, the absence of preferred items on the computer screen does not affect the quantity purchased in the product category, but leads to a choice bias in favor of major national brands.*
Comments: Hypotheses 1 to 5 are expected to hold for consumers who find their regularly purchased products on the computer screen, and are therefore able to realistically reproduce their real life purchasing patterns (Cohen and Gadd, 1996). In many CSSE, it will be practically infeasible to include all the items in the category in the virtual store. Given that consumers were not given the possibility to shop elsewhere, and to the extent that jam is not an impulsively bought product, we do not expect assortment reductions to affect purchase quantity. Yet, in line with previous research, we believe that consumers who do not find their regular product on the shelf, tend to switch to a 'well known product'. In our laboratory setting, this tendency will be reinforced by the response bias referred to in hypothesis 2, and the presence of only brand, price, 'objective' product information, and promotion cues.

Methodology

To validate our hypotheses, we confront the CSSE outcomes with results obtained with actual scanner (panel) receipt data. Note that these data are widely used in practice to establish purchase levels, market shares and promotional reactions. Receipt data are available for 2 outlets of the considered retail chain, over 16 subsequent weeks, for all store customers possessing a chain loyalty card. In addition, information is collected on various types of in store promotions during the period for which receipt data are available. In comparing the experimental with the receipt data, we correct as much as possible for 'systematic' sample or 'selection' differences, and for differences in in-store conditions.

Matching Samples and Store Conditions

Inhabitants in the trading area of the outlets from which receipt data are available, have a socio-demographic profile that closely matches that of the outlets where CSSE interviews were conducted. In the receipt data panel, we retain consumers who purchase at least one pack in the category per month: this minimum use rate is similar to the one used for screening CSSE respondents. For the jam category, this leaves us with information on 486 households in the experimental data set, and 283 households from the receipt panel data . Other potential differences between the experiment and real life setting concern the frequency with which consumers are confronted with the shelves (weekly visits in the experiment), and the width of the assortment presented (experiment comprises only subset of items). Information on scanner panel members' store visit frequency, and on whether

[8] Respondents who did not take the task seriously, or who indicated that they had difficulties in imitating real buying behavior because of insufficient shopping experience, were omitted from the analysis (the 41 "omitted" respondents mentioned earlier).

CSSE consumers find their regularly purchased products on the computer shelves (see control question in module 3) will be helpful in detecting eventual differences in behavior arising from experimental store visit and assortment conditions.
As outlined above, the artificial shelves presented on screen comprise only a subset of the items sold in the chain's outlets: for jam, 22 out of 57 items are available on screen, accounting for about 55% of the real category volume. The shelf layout is a reproduction of real store planograms for the category. While regular price levels in our experiment match those in the receipt data set, promotional conditions are different between the two data sets.

Purchase Quantity and Choice Models

To assess the external validity of purchase quantity and choice decisions, we estimate for each product category and data base: (i) a purchase quantity model, linking the number of units bought by a consumer in a given category and week, to household and marketing variables, and (ii) a stochastic choice model, explaining the choice probability of any item in the assortment, given a category purchase, as a function of item characteristics, previous choice decisions, and promotion variables. Estimating these models will allow us to 'partial out' promotional impacts when comparing purchase quantities, choices, and purchase variation, between data sets.

In each data set, weekly *purchase quantity* probabilities per consumer are modeled as a Poisson process with the following structure[9]:

$$P(y_{ht}=n_h)=\frac{\exp(n_h.\lambda_{ht}).\exp(-\exp(\lambda_{ht}))}{n_h!} \qquad \textbf{(1)}$$

where $P(y_{ht}=n_h)$ represents the probability that household h purchases n_h units in week t. Similar to Bucklin and Gupta (1992) and Gupta (1988), we specify the expected weekly quantity, $\exp(\lambda_{ht})$, as a function of household use rate, inventory level, and promotions in the category[10]:

$$\lambda_{h,t}=\alpha_{s,h}+\beta_1.US_h+\beta_2.Inv_{h,t}+\sum_k.\beta^k_{pro}.Prom^k_{h,t} \qquad \textbf{(2)}$$

where
$\alpha_{s,h}$ = a segment specific constant.

---

[9] In accordance with Dillon and Gupta (1996), we use an exponential function for the expected purchase rate to ensure positive values.

[10] While the 'use rate estimate' allows to capture differences between households, the inventory variable mainly accounts for intertemporal variations in λ_{ht} .

The Poisson model is estimated using a latent class approach, where segment-specific -s account for unexplained heterogeneity between households.
US_h= household use rate. US_h is either measured as self-reported use rate (experimental data set), or estimated from the household's purchases during a four week initialization period (receipt data set).
$Inv_{h,t}$ = a dynamically updated inventory level estimate, representing the number of units the household has in stock at the beginning of week t. Inventory levels are weekly updated, based on information on inventory and purchases in the previous week, and data on weekly use rates of the household:

$$\begin{aligned} Inv_{h,\,t} &= Inv_{h,\,t-1} + N^{h}_{t-1} - US_h \\ N^{h}_{t-1} &= \textit{number of units bought by h in } t-1 \end{aligned} \tag{3}$$

$Prom^k_{ht}$ = the level of the k'th promotion variable, observed by household h in period t. Promotion conditions and hence promotion variables differ between the experimental and scanning data sets.

In the experimental data set, promotions take the form of price discounts, coupons, bonus packs, premiums and increases in shelf facings. These promotional conditions occur for two items: one private label item (PL), and one national brand item (NB). The following promotion variables are therefore included in the experimental purchase quantity model:

$Price(j)_t$: a dummy variable equal to 1 if a price cut was offered for item j in period t, and equal to 0 otherwise (j=PL, NB).
$Coup(j)_t$: a dummy variable equal to 1 if a coupon was offered for item j in period t, and equal to 0 otherwise (j=PL, NB).
$Bonus(j)_t$: a dummy variable equal to 1 if an extra amount (bonus pack) was offered for item j in period t, and equal to 0 otherwise (j=PL, NB).
$Gift(j)_t$: a dummy variable equal to 1 if a premium was offered for item j in period t, and equal to 0 otherwise (j=PL, NB).
$Shelf(j)_t$: shelf space dummy variable equal to 1 if item j received additional shelf facings in period t, and equal to 0 otherwise (j=PL, NB).

Given the nature of the experiment, all these promotions are 'in-store'.

In the scanner receipt data set, no coupon, gift, or shelf space promotions are encountered. A limited number of quantity discounts occur during the observation period: some of these correspond to 'bonus packs', others to 'one-free-with-several' promotions. Quantity discounts are always 'featured' (announced in the retailers' folder), they occur only for national brands, and apply to all items of the promoted brand simultaneously. Promotion variables were specified in a similar way as for the experimental data set, with the exception that 'count' rather than 'dummy' variables were used in cases where several items of the same type (PL, NB) were on promotion simultaneously. The bonus pack variable $Bonus(NB)_t$, for example, now quantifies the number of national brand items on which a bonus pack is offered in that week, while $OFS(NB)_t$ indicates the number of national brand items enjoying a one-free-with-several promotion. $QUAN(NB)_t$ is a summary variable, representing the number of items on quantity (either bonus pack or one-free-with-several) discount. Featured price discounts occur for the Private label only: the variable $PriceF(PL)_t$ indicates the number of private label items on price feature in week t. Finally, the scanning data set comprises non-displayed and non-featured price cuts for some national brands, these are measured by $Price(NB)_t$ (the number of national brand items on discount without feature or display in week t).

Household *item choices* conditional on purchase are estimated using a simple MNL type model. Our specification is based on the structure suggested by Guadagni and Little(1983), subject to some minor modifications. Following a suggestion by Bucklin and Gupta(1992), we replace the original dynamic 'loyalty' variables in the G&L model by static measures of household preferences, and supplement them with 'last purchase variables' (as specified in equation 5). This better allows to distinguish between intrinsic preference heterogeneity, and state dependence. Relevant item characteristics refer to brand and taste, no size differences are encountered. We further supplement Guadagni and Little's model with an interaction effect between the household's tendency to seek variation in behavior (see equation 5) and the past purchase variable, thereby allowing some households to exhibit reinforcement behavior, and others to show a preference for switching. The specification for the choice model is given in equation 4.

$$p^{h}_{j,t}=\frac{\exp(v^{h}_{j,t}+u^{h}_{j,t})}{\sum_{l}(\exp(v^{h}_{l,t}+u^{h}_{l,t})} \qquad (4)$$

where
$P^{h}_{j,t}$ = probability that household h chooses item j on occasion t
$u^{h}_{j,t}$ = random component of utility of item j for household h on occasion t,
$v^{h}_{j,t}$ = systematic component of utility of item j for household h on occasion t,

The systematic utility is further given by:

$$\begin{aligned} v^{h}_{j,t}=&\sum_{b} D_{j,b}*(\alpha_{1,b}+\alpha_{2,b}.Loyb^{h}_{j,t})+\sum_{v} D_{j,v}*(\alpha_{1,v}+\alpha_{2,v}.Loyv^{h}_{j,t})+\\ &\gamma_{1}.Lastp^{h}_{j,t}+\gamma_{2}.Lastp^{h}_{j,t}.VST^{h}+\sum_{k}.\beta^{k}_{pro}.Prom^{k}_{j,t} \end{aligned} \qquad (5)$$

where
$D_{j,b}$ ($D_{j,v}$) are attribute dummy variables, equal to 1 if item j belongs to brand b (variant v), and equal to 0 otherwise.
$\alpha_{1,b}$ ($\alpha_{1,v}$) are brand-(variant-) specific intercepts (see Fader and Hardie 1996).
$Loyb^{h}$ ($Loyv^{h}$) are static loyalty variables indicating the fraction of consumer purchases in the initialization period accounted for by brand b (variant v),
$Lastp^{h}_{j,t}$ indicates the difference between item j and the item(s) purchased on the previous occasion[1], the difference is equal to zero if the items have the same brand and variant (taste), .5 if either the brand or the variant is different, and 1 if both differ.
$Prom_{kht}$ = the level of the k'th promotion variable encountered by household h on purchase occasion t.
VST^{h}= household h's tendency to seek variation in buying behavior,

[11] Simultaneous purchases of 2 (or more) different items were split up into 2 (or more) single purchase observations for estimation, using an approach similar to Papatla and Krishnamurthi (1992). When more than 1 item was purchased on the previous purchase occasion, the variable $Lastp^{h}_{j,t}$ was set equal to the average difference between j and the previously bought items.

approximated by the entropy measure computed over a 4 week initialization period. The entropy measure was rescaled to account for differences in assortment size and purchase quantity (determining the opportunities for variation in buying behavior; see Campo 1997). After rescaling, the entropy lies between 0 (no variation in behavior) and 1 (maximum variation in behavior).

$$VST^{h} = \frac{ent^{h}}{entmax^{h}} \tag{6}$$

$$ent^{h} = \sum_{j} -share^{h}_{j} * \ln(share^{h}_{j})$$

$$entmax^{h} = \ln(M^{h})$$

M^{h} = $\min(J,Q^{h})$
J = *number of items in the assortment*
Q^{h} = *overall purchase quantity of h during the initialization period*
$share^{h}_{j}$ = *share of item j in household h's purchases during the initialization period*

While this model remains fairly simple, it provides a good description of household behavior, and serves our purpose of removing major immediate promotion impacts.

Estimation Results

The purchase quantity model is estimated using maximum likelihood procedures, for a varying number of segments or latent classes. Estimation for the experimental data set is based on 11 weeks, week 1 being an initialization week for the lagged promotion effects. For the scanning receipt data; weeks 1 to 4 serve as initialization weeks over which the household's use rate in the category is calculated, and model estimation is based on purchases from weeks 5 to 16. Table 1 reports goodness-of-fit statistics for different numbers of segments. As can be seen from this table, the optimum number of latent classes based on the Consistent Akaike Information Criterion (CAIC; see e.g. Dillon and Gupta 1996) equals 7 for the experimental data set and 8 for the scanning data. Estimation results for the 7- and 8-segment solution are provided in Tables 2 and 3. Both models reveal a highly significant positive effect for use rate, and negative impact for inventory level. Estimated promotion coefficients are either insignificant, or have the expected sign: a more detailed discussion of promotion effects is provided when discussing hypothesis 4. To get a better grasp on the models' descriptive validity, expected purchase quantities based on the model are compared to observed quantities in each week. The results are provided in figure 1: they indicate that both models fit the real data well.

<insert tables 1, 2 and 3 here>
<insert figure 1 here>

Choice models are estimated by maximum likelihood procedures, using information from weeks 5 to 12 (for the experimental data set), and from weeks 5 to 16 (for the scanner data set). Weeks 1 to 4 again serve as initialization weeks, to compute the household's intrinsic preferences (loyalty variables) and variety seeking tendency. Estimation results are provided in Tables 4 and 5, Figure 2 summarizes observed and predicted market shares for the different brands in both data sets. To test whether the MNL model's IIA assumption is justified, a procedure suggested by Fader and Hardie (1996) was applied. Results demonstrated that both data sets comply with the IIA assumption.

<insert tables 4 and 5 here>
<insert figure 2 here>

Hypothesis Tests

Hypothesis 1

Figure 3 provides a histogram of consumers' self-reported use rates, and observed average weekly purchases during the 11 last experimental weeks. Week 1 is left out of this analysis as it is considered exceptional (starting inventory is zero for all consumers), and since the purchase quantity model is estimated for weeks 2 to 12. Next to observed experimental purchases, which may be affected by promotional activities in the experiment, we calculate expected weekly purchase quantities per consumer in the absence of promotions. These expected quantities are obtained from the purchase quantity model through 'dynamic' simulation, taking into account that consumers who buy less in a given week when the promotion is absent, will end up with a lower inventory level at the beginning of the following week, and hence are more likely to buy larger quantities then. For purposes of comparison, we also include expected quantities from the Poisson model when promotions are present. Figure 3 shows that, while the distribution of 'observed' quantities is somewhat more irregular than that of reported use rates and of predicted purchase rates with the promotions, they seem to follow a similar pattern. Simulated quantities under the no promotion condition lead to an expected shift towards lower quantities, but the impact seems small.
To formally test hypothesis 1 sub 1, we perform a Wilcoxon Matched-Pairs Signed-Ranks test on reported versus observed use rates in the experiment. The resulting test value of -.575 ($p=.565$) clearly indicates that the Null Hypothesis that both variables have the same distribution cannot be rejected: hence we accept our hypothesis 1 sub 1; and conclude that within the context of a CSSE, consumers' experimental purchase quantities are in line with their self-reported use rates.

<insert figure 3 here>

To check whether experimental purchases do not only match what consumers 'announce', but also what they buy in real life settings, we compare the weekly experimental purchase rates with those obtained from the scanner panel data set. For purposes of comparison we dynamically simulate, for the receipt data set, expected weekly purchases in the absence of promotions on the basis of the estimated purchase quantity model. Figure 4 shows a histogram of weekly purchases per consumer as they are (i) observed in the experiment, (ii) simulated in the experiment under the no-promotion condition, (ii) observed from the scanner panel, and (iv) simulated for the scanner panel in the absence of promotions.

Correcting for promotions leads to an important shift in the scanner panel data set, so that a 'reasonable' comparison is that between weekly experimental and scanning quantities in the absence of promotions. From figure 4, we observe that the experiment has a higher frequency for larger weekly quantities per consumer (1 or 2 units per week) compared to the scanning data set, while the proportion of buyers adopting 1 unit every 3 weeks (use rate .33) or every 4 weeks (use rate .25) is more important in the scanner panel than among CSSE consumers. A nonparametric test on the difference between the quantity distributions indicates that the deviations are significant (Man-Whitney $z=-8.765$; $p=0.00$). Table 6, which summarizes means and standard deviations for the various weekly purchasing rates, provides further support for this finding. We therefore cannot accept our hypothesis that experimental purchase rates coincide with real life purchase frequencies.
One reason behind this deviation may be the difference in 'shopping frequency' between experimental consumers (who are placed in front of the shelves every week) and scanner panel consumers. To get a better grasp on the importance of this factor, we reconsider the weekly purchase quantities for a subset of scanner panel consumers who visited the chain at least 16 times during the 16 week data period. 201 out of 286 households fulfill this requirement, their histogram of weekly purchase rates is provided in figure 5. From this figure, it is clear that while concentrating on frequent scanner panel shoppers somewhat reduces the difference between scanner and experimental purchase quantities, a significant bias remains (Mann-Whitney $z=-7.567$; $p=.00$). A plausible explanation for the observed purchase quantity differences is that consumers have a tendency to overestimate their category use rate. Having specified a use rate, they tend to engage in experimental purchases consistent with this estimate (they are encouraged to do so since their inventory level throughout the experiment is based on the use rate estimate), hence the higher experimental purchase rates. Under these assumptions, a meaningful approach for future CSSE would be to estimate respondents' use rates on the basis of their past observed purchase behavior, and to rely upon this information for experimental inventory updates instead of upon their self-reported use rates.

<insert figures 4 and 5 here>
<insert table 6 here>

Hypothesis 2

In comparing market share levels across data sets, we have to account for differences in promotional conditions, as well as for the fact (which is typical for CSSE) that our experimental assortment is incomplete: two national brands are not included in the experiment, while for other national brands and the private label some taste variants are missing. To make the results more comparable, we therefore predict market shares for both data sets in the absence of promotions, and rescale the (predicted) scanner data market shares of items included in the experiment so that they sum to one. Figure 6 provides a summary of these calculations. It reports, for the brands and items included in the experiment: (i) the observed market shares in the experimental data set, (ii) predicted shares in the experiment under the no-promotion condition, (iii) observed scanner data shares of items available in the experiment (rescaled so that they sum to one), and (iv) predicted scanner data shares in the absence of promotions, for the items available in the experiment, rescaled to sum to

one.
Some interesting patterns emerge from figure 6. For two of the national brands (NB1 and NB3), experimental shares significantly[12] and systematically exceed true shares: a finding in line with our hypothesis. For the generic product, observed experimental share is significantly below observed scanning share, and predicted experimental share without promotions is dramatically below predicted scanner share. This observation, too, is consistent with our expectations. Yet, the findings for the Private Label do not coincide with the hypothesized pattern. After correcting for promotional differences, Private Label share in the experiment is comparable to the market portion it captures in the scanner panel data set. Finally, experimental share for one National Brand (NB2) is in line with the rescaled scanning data share.
We conclude from this discussion that results provide partial support for hypothesis 2. While the experimental data set systematically underestimates the true share of generic products, this does not hold for the private label. A plausible explanation is, that customers of the analyzed supermarket chain have come to perceive private labels as valuable and socially acceptable alternatives, in contrast to generics which are still attributed a negative sign value. As demonstrated by Dhar and Hoch (1997), performance of store brands varies widely among retailers and product categories, depending on factors such as the quality and breadth of private label offerings, the retailer's overall pricing strategy, and degree of promotional support for private labels. The realism of private labels' choice shares in CSSE's may therefore depend on these factors and associated buyer perceptions, yielding biased estimates for some retail chains and/or product categories but not for others.

<insert figure 6 here>

Hypothesis 3

To test for differences in purchase variation between experimental and real life settings, we compute the level of entropy in the purchase history of CSSE and scanner panel households. Entropy is a widely accepted indicator of purchase variation (see, e.g., Van Trijp and Steenkamp, 1990). Since jam items differ in brand and taste, both attributes are accounted for in our entropy computations. Table 7 reports the average entropy measure, and its standard deviation, across households in both data sets. As assortment depth and purchase quantities differ across data sets, 'absolute' entropy levels are not really comparable. We therefore calculate household's relative entropy levels (see equation 6), the mean and standard deviations of which are also reported in Table 7. Clearly, entropy levels are highly comparable across data sets. A nonparametric test reveals that while absolute entropies differ (Man-Whitney $z=-5.14$, $p=0$), there is no significant deviation between relative entropy distributions in both data sets ($z=-1.14$, $p=.254$): we therefore accept hypothesis 3[13].

---

[12] Based on a t-test of the significance between two estimated proportions. All differences reported as significant exceed at least the 5% confidence level.

[13] The degree of purchase variation was also found to vary substantially across product categories. For the 2 hedonic products cereals and jam, respondents displayed a much higher degree of variation in choice

<insert table 7 here>

Hypothesis 4

In the purchase quantity model, the parameter β_{pro} associated with a promotion indicates the shift in λ_{ht} when an additional item in the category is on promotion. It can be shown that $\exp(\beta_{pro})$ quantifies the ratio of expected weekly purchase quantity with and without a(n) (additional) promotion action. Table 8 summarizes the results for various promotion types in the two data sets.[14] The promotion parameters in the item choice model indicate the (additive) increase in an item's utility when promoted, they are also reported in Table 8[15]

<insert table 8 here>

When comparing the promotion parameters, it has to be taken into account that the experimental and real promotion conditions differ on important points. Scanner receipt and associated promotional data were made available shortly after the interviews had been conducted, and revealed the following differences in promotion conditions. First, while the CSSE promotions are necessarily in-store, most merchandising activities in the scanning data set refer to featured promotions, announced in the retailer's leaflet. Second, the scanner data set covers only two of the four types of promotion instruments analyzed in the experiment: while price cuts and quantity discounts are offered, no coupons or gifts are found in the scanner data for jam, and there is no information on possible shelf rearrangements. Third, when a brand is on promotion in the scanner data set, this promotion usually applies to most of its variants (tastes) simultaneously, which is not the case in the experimental data set. Even though we measure the impact of promotions in both data sets at the item level (so that our measures are comparable), this difference in promotion practice may affect the estimated outcomes. Finally, while some (actual) promotions apply to national brands, others affect private label items - a difference that should also be accounted for when comparing the results.

Even though the difference in promotion conditions hampers the comparison between data

behavior than for the two functional products margarine and paper towel (a difference that remains significant, even after individual differences in variety seeking tendency have been controlled for). These results are in line with theoretical expectations and reported empirical findings (see e.g. Van Trijp et al. 1996), providing further support for hypothesis 3.

[14] In the experimental data set, promotions do not occur for different items of a brand simultaneously, and the promotion variables therefore take the form of dummies indicating whether or not the item is on promotion. In the scanner data set, the promotion variables 'count' the number of items of the promoted brand to which the offer applies (since these items are either promoted together or not at all, it is not possible to estimate the effects for the various items separately).

[15] Given that the choice model has a similar structure (functional form, measurement of variables) in both data sets, these coefficients can roughly be compared between the experimental and the scanning setting. Alternatively, one can compute - given the estimated promotion coefficient- the quantity $(1-p^h_{it})*(\exp(\beta_{pro})-1)/(1+p^h_{it}*(\exp(\beta_{pro})-1))$, which is the % change in choice probability p^h_{it} resulting from a promotion for i.).

sets, some interesting patterns emerge. First, comparable promotions in the scanner data set seem to be less effective than experimental promotions, and this both in terms of purchase quantity and choice. More specifically, the non-featured price cut for the national brand has a larger impact in the experimental than in the scanning setting. Even featured promotions in the scanning data set are less effective than (non-featured!) experimental promotions: this is true for the private label price cut as well as for the national brand quantity discount. One reason might be that scanning promotions on an item always coincide with promotions of other items of the same brand, which reduces their impact. On the other hand, it seems like the on-shelf promotions in the experiment act more like 'displayed' promotions. Though no specific display materials are used to support them, the fact that CSSE consumers are placed in a front view of the (reduced) shelf, instead of walking through the aisle in a noisy store environment, may lift promotions up to a 'display' status.

When viewed as displayed promotions, the CSSE promotional effects are quite in line with expectations from the literature. The relative effectiveness of price cuts and quantity discounts observed in the CSSE matches the findings in the scanning data set. Also, the differential reactions to experimental national brands vs private labels (national brand promotions generating stronger response) correspond to prior expectations, and to the findings in the scanning data set.

In conclusion, these preliminary results roughly point in the direction of hypothesis 4: promotions in the context of CSSE lead to more pronounced reactions than featured or unannounced promotions in a real life environment, and are more in line with response to displayed promotional activities. Relatively speaking, the experimental promotion effects provide a realistic picture of reactions to different instruments and brand types.

Hypothesis 5

To assess whether systematic differences in quantity and choice appear over the 12 subsequent weeks, we re-estimate the purchase quantity and choice models for the experimental data set including weekly dummy variables (and selecting week 5 as the reference week).

In the purchase quantity model, only weeks 1 and 2 lead to statistically significant deviations from week 5. These deviations could be interpreted as 'initialization' effects, resulting from the fact that for all consumers, inventories are set to zero in the beginning of the experiment. The pattern of weekly parameters over the subsequent weeks displayed in figure 7 'confirms' this interpretation of an initialization effect that dies out: there is no evidence of a trend in the quantities bought throughout the experimental history, nor is there a systematic end-of-period effect.

In the choice model, the majority of weekly parameters are not significantly different from zero. From the 154 parameters estimated, only 5 are significant at the 10% significance level. From this we can conclude that there is no systematic tendency for items to be more or less often bought in early versus late weeks of the experimental period.

<insert figure 7 here>

From the above discussion, we conclude that hypothesis 5 can be accepted. Apart from exceptionally high inventory build-up purchases in the first week, consumers behave in a realistic manner throughout the observation period: their purchase quantities and choices do not systematically change as the experiment proceeds.

Hypothesis 6

Of the 486 respondents for which experimental information is obtained, 438 report finding their preferred item on the simulated shelf. To check for biases induced by the reduced experimental assortment, we compare the weekly purchase rate of the 438 respondents who do find their preferred products on the computer screen, with that of the remaining respondents.
With an average of .83 (standard deviation .60), the weekly purchase quantity of those who can locate their preferred item is not substantially higher than that of the remaining respondents (for which the average amounts to .70, with a standard deviation of .50). Since not all respondents are confronted with the same promotion scenarios, we also compare the simulated weekly quantities in the absence of promotions: these amount to .79 (SD: .56) for the first group of 438 consumers, and to .66 (SD.47) for the remaining 48 respondents. A nonparametric test on the comparability of the purchase rates in both subgroups suggests that the difference is at best marginally significant, for observed as well as for simulated average weekly quantities (for observed quantities, Mann-Whitney $z=-1.378$, $p=.168$; the Mann-Whitney test statistic for simulated no-promotion quantities $=-1.621$, $p=.105$). Whether purchase rate differences are a result of assortment reductions, can be further explored by analyzing deviations between actual and self-reported purchase rates for the subsample of respondents that did not find their favorite item on the CSSE shelves (cf Hypothesis 1). If respondents who cannot find their preferred item on screen have a tendency to buy less, this should be reflected in observed experimental quantities below self-reported use rates for this group. A Wilcoxon Matched-Pairs Signed Ranks test on reported versus observed weekly purchase quantities indicates that the distribution is not significantly different (z-statistic: -1.103, $p=.27$). We conclude that not finding their preferred items on screen does not induce respondents to buy substantially less, yet our results suggest that some caution is in order.

In a similar vein, we compare the market share of the various brands among respondents who do find their preferred item, and others. Table 9 reports the observed market shares for the experimental brands in both respondent groups. Based on test of equality of proportions, we find that for 2 of the national brands (NB1 and NB3), a significant difference is found (for the private label, the difference is 'marginally' insignificant). Yet, the difference does not go in the hypothesized direction: contrary to expectations, we observe from table 9 that the smaller national brand NB3 enjoys an increase, and the larger one NB1 a decrease. Hence, we conclude that the second part of hypothesis 6 cannot be accepted: though some differences in share are observed, there does not seem to be a 'systematic pattern' in these shifts.

<insert table 9 here>

Implications

The current study has important implications for marketing research. As suggested by a number of recent articles, computer simulated shopping experiments open up a whole new and exciting range of research possibilities. The present research sheds further light on what uses can be made of the outcomes of these types of experiments, and where the pitfalls are.

First, our analysis yields additional insights into the 'realism' of experimentally obtained purchase quantities. It suggests that for non-impulse products, obtaining prior information on use rates, and providing consumers with dynamically adapted cues on their 'remaining' inventory level on each shopping trip, streamlines their experimental purchases in accordance with these use rates. Important to note is that experimental purchases still exceed true purchase quantities. As hypothesized, choices in a CSSE context are biased in the direction of less generic brand purchases. For private labels, though, no such bias is observed. A potential explanation, which is in line with the view of chain managers, is that consumers have come to judge these private label items as 'valuable' alternatives for choice that are accepted by their peers. We emphasize that the result may be specific to this study, because of the image of the chain analyzed, and the status of the consumers (in both data sets, we recruited consumers that are relatively 'chain loyal'). Our findings further reveal that consumers' purchase variation is not systematically affected by the experimental nature of the setting, provided that an effort is made to reduce time compression. The fact that consumers were familiar with the assortment presented in the experiment, and were given a number of cues on item characteristics, may also explain this result. In terms of promotional impact, our study provides some interesting indications. From the study results, we postulate that experimental promotions tend to be more effective than (non-displayed) in store promotions (even though there is no clear display material involved, our experimental promotions act like 'displayed' promotions). Our findings also suggest that CSSE yield valid information on purchase behavior over successive periods. Throughout the 12 week horizon covered by the experiment, consumers kept on buying in a similar (realistic) manner. Yet, it should be noted that some starting up effects in the first week or weeks are inevitable, rendering these weeks less useful for analysis. Also, from consumer reactions, we feel that prolonging the experiment over more than 12 weeks is not advisable. Not only would this reduce response rates due to excessive time requirements, it would render the exercise too boring for consumers and therefore jeopardize response quality in later periods. We therefore posit that while relevant dynamic information can be obtained, the length of the observation period must remain limited. Finally, the absence of favored items on the screen does not reduce purchase quantities. While consumers who cannot locate their preferred items on the screen exhibit different choice probabilities for the remaining items than other respondents, there is no clear pattern in these observed market share deviations.

In summary, our research suggests that for non-impulse products, CSSE can yield meaningful insights into consumers' dynamic purchasing strategies and promotion reactions. These results are obtained for a 'realistic' computer experiment that, while being clearly more sophisticated than paper and pencil tests, is far from being a 'futuristic' gadget. To the extent that results from this type of experiment are valid and useful this will certainly also be true for more advanced virtual shopping programs bound to be developed in the future. Clearly, our analysis is only of limited scope: it is confined to only one product category, and based on a comparison of two consumer panels. Even though both samples were carefully matched, sample sizes are important and the statistical tests robust, confirmation from additional studies might be useful. Also, the study of promotional effects and assortment reduction clearly needs further development. Future research may concentrate on these issues.

References

Brucks, M. (1988), "Search Monitor: An Approach for Computer-Controlled Experiments Involving Consumer Information Search", *Journal of Consumer Research*, 117-121.

Bucklin, R. and S. Gupta (1992), "Brand Choice, Purchase Incidence and Segmentation: an integrated approach", *Journal of Marketing Research*, 29 (May), 201-215.

Burke, R. (1995), "Virtual Shopping", *OR/MS today*, August, 28-34.

____ (1996), "Virtual Shopping: Breakthrough in Marketing Research", *Harvard Business Review*, March-April, 120-131.

____, B. Harlam, B. Kahn and L. Lodish (1992), "Comparing Dynamic consumer Choice in Real and Computer-simulated Environments", *Journal of Consumer Research*, June, 71-82.

Campo, K. (1997), *Variety Seeking and the Sensitivity to In-Store Promotions*, unpublished doctoral dissertation, University of Antwerp - U.F.S.I.A.

Carson, R.; J. Louviere, D. Anderson, P. Arabie, D. Bunch, D. Hensher, R. Johnson, W. Kuhfeld, D. Steinberg, J. Swait, H. Timmermans and J. Wiley (1994), "Experimental Analysis of Choice", *Marketing Letters*, 5 (4), 351-368.

Cohen S. and M. Gadd (1996), "Virtual Reality Shopping Simulation for the Modern Marketer", *Marketing and Research Today*, February, 18-26.

Dhar, S.K. and S.J.Hoch (1997), "Why Store Brand Penetration Varies by Retailer", *Marketing Science*, 16(3), 208-227.

Dillon, W. and S. Gupta (1996), "A Segment Level Model for Category Volume and Choice", *Marketing Science*, 15 (1), 38-59.

Fader, P.S. and B.G.S.Hardie (1996), "Modeling Consumer Choice among SKU's", *Journal of Marketing Research*, 33(November), 442-452)

Narasimhan, C.; S. Neslin and S. Sen (1996), "Promotional Elasticities and Category Characteristics", *Journal of Marketing*, 60, April, 17-30.

Neslin, S. and L. Schneider(1996), "Consumer Inventory Sensitivity and the Postpromotion Dip", *Marketing Letters*, 7 (1), 77-94.

Papatla, P.and L.Krishnamurthi (1992), "A Probit Model of Choice Dynamics", *Marketing Science* , 11(2), 189-206.

Van Trijp, H. and J.B. Steenkamp (1990), "An Investigation into the Validity of Measures for Variation in Consumption Used in Economics and Marketing", *European Review of Agricultural Economics*, 17(1), 19-41.

Van Trijp, H.; W. Hoyer and J. Inman (1996), "Why switch? Product Category level Explanations for True Variety-Seeking Behavior", *Journal of Marketing Research*, 33, August, 281-292.

Table 1: Goodness of Fit for the Poisson model, 2 to 7 segments

# of segments	Scanning		Experiment	
	LL	CAIC	LL	CAIC
6	-3120	6314	-5139	10408
7	-3116	6312	-5096	10329
8	-3103	6293	-5086	10335
9	-3102	6298		

Table 2: Estimation results for the Purchase Quantity Model in the Experimental data set (7 segments)

Variable	Parameter	t-statistic
US_h	.635	23.47
$Inv_{h,t}$	-.452	-30.49
$Price(PL)_t$	.038	.39
$Price(NB)_t$	.163	1.75
$Coup(PL)_t$	.150	1.68
$Coup(NB)_t$	.252	3.19
$Bonus(PL)_t$	.056	.54
$Bonus(NB)_t$	.155	1.60
$Gift(PL)_t$	.081	.81
$Gift(NB)_t$	-.038	-.39
$Shelf(PL)_t$	.062	.88
$Shelf(NB)_t$	.081	1.17
LL (number of obs)	-5096.3 (5346)	

Table 3: Estimation Results for the Purchase Quantity Model in the Scanner data set

Variable	Parameter	t-statistic
US_h	.284	9.33
$Inv_{h,t}$	-.217	-16.01
$PriceF(PL)_t$	.130	15.58
$Quan(NB)_t$	.048	2.13
$Price(NB)_t$	.023	1.14
LL (number of obs)	-3103.4 (3396)	

Table 4: Estimation Results for the Choice Model in the Experimental Data Set

Variable	Parameter	t-statistic
Brand and taste dummies		
NB1	-3.590	-18.710
NB2	-3.124	-13.824
NB3	-4.787	-17.869
PL	-3.283	-17.588
Generic	-3.605	-19.368
taste 1	-.249	-1.555
taste 2	1.171	3.199
taste 3	-.531	-3.295
taste 4	-.541	-2.143
taste 5	-.116	-.730
taste 6	1.380	4.213
Loyalty variables		
NB1	4.946	19.917
NB2	3.801	14.052
NB3	5.298	8.183
PL	4.491	16.964
Generic	5.728	16.129
taste 1	3.750	13.426
taste 2	2.706	.737
taste 3	4.900	14.135
taste 4	8.021	8.760
taste 5	3.400	13.612
taste 6	2.937	2.297
taste 7	3.752	15.309
Purchase event feedback		
$Lastp^{h}_{j,t}$	.056	17.741
$Lastp^{h}_{j,t}$ * $Entr^{h}$	.006	3.284
Promotion variables		
$Price(PL)_t$	2.026	7.865
$Price(NB)_t$	2.768	12.983
$Coup(PL)_t$	1.819	6.678
$Coup(NB)_t$	3.054	15.180
$Bonus(PL)_t$	1.757	6.479
$Bonus(NB)_t$	2.781	11.446
$Gift(PL)_t$	1.268	4.943
$Gift(NB)_t$	2.013	9.856
$Shelf(PL)_t$	.832	4.174
$Shelf(NB)_t$	.606	3.603
LL (number of obs.)	-4101 (2045)	

Table 5: Estimation results for the Choice Model in the Scanner Panel Data Set

Variable	Parameter	t-statistic
Brand and taste dummies		
NB1	-2.635	-5.110
NB2	-3.113	-4.494
NB3	-5.408	-2.599
NB4	-3.400	-3.895
NB5	-.860	-.960
PL	-3.009	-5.177
Generic	-2.902	-5.139
taste 1	-.235	-.569
taste 3	-.065	-.168
taste 5	-.214	-.510
taste 8	-.675	-1.563
taste 9	.733	1.890
Loyalty variabels		
NB1	2.060	3.637
NB2	2.290	3.190
NB3	4.687	1.131
NB4	4.602	2.847
NB5	-1.004	-.985
PL	2.957	4.678
Generic	3.343	4.002
taste 1	4.180	4.927
taste 3	2.614	3.986
taste 5	2.007	2.214
taste 7	1.707	2.984
taste 8	4.723	4.707
taste 9	1.610	3.521
Purchase event feedback		
$Lastp^h_{j,t}$	-.080	-10.194
$Lastp^h_{j,t}$ * $Entr^h$	.011	1.535
Promotion variables		
$PriceF(PL)_t$	1.745	2.718
$Price(NB)_t$	.480	.749
$Bonus(NB)_t$	1.692	1.715
$OFS(NB)_t$	1.003	1.783
LL (number of obs.)	-917 (559)	

Table 6: Means (Standard Deviations) of weekly purchases in the experiment and scanner data set

	Experiment	Scanning
observed	.817 (.591)	.526 (.347)
simulated, with promotions	.791 (.548)	.565 (.403)
simulated, without promotions	.763 (.543)	.484 (.345)

Table 7: Mean (Standard Deviation) of Entropy in the experiment and scanner data set

	Experiment	Scanning
absolute entropy	1.18 (.45)	1.00 (.39)
relative entropy	.57 (23)	.54 (.21)

Table 8: Promotion effect for various brand types and instruments, in the experiment and scanner data set.

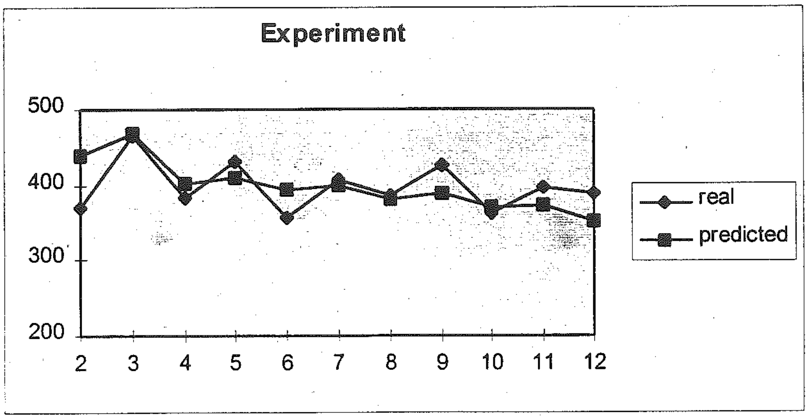
	Experiment		Scanning	
	Purchase Quantity exp(βpro)	Choice βpro	Purchase Quantity exp(βpro)	Choice βpro
Featured Price Cut				
National Brand	-	-	-	-
Private Label	-	-	1.14	1.74
Non-featured Price Cut				
National Brand	1.18	2.77	1.02	ns
Private Label	ns	2.03	-	-
Featured Quantity Discount				
National Brand	-	-	1.05	1.00 or 1.70
Private Label	-	-	-	-
Non-featured Quantity Discount				
National Brand	1.17	2.78	-	-
Private Label	ns	1.76	-	-
Non-featured Coupon				
National Brand	1.29	3.05	-	-
Private Label	1.16	1.82	-	-
Non-featured Gift				
National Brand	ns	2.01	-	-
Private Label	ns	1.27	-	-

ns=not significant, - = not applicable

Table 9: Observed Experimental Market Shares for respondents who find their preferred items, and others

	Respondents who find their preferred items (438 resp)	Others (48 resp)
NB1	0.355	0.289
NB2	0.234	0.257
NB3	0.033	0.067
Private Label	0.271	0.289
Generic	0.108	0.098

Figure 1 : Real versus Predicted Purchase Quantities



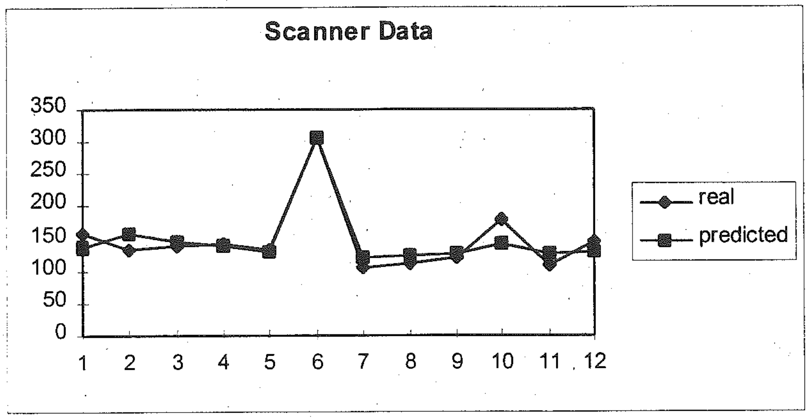
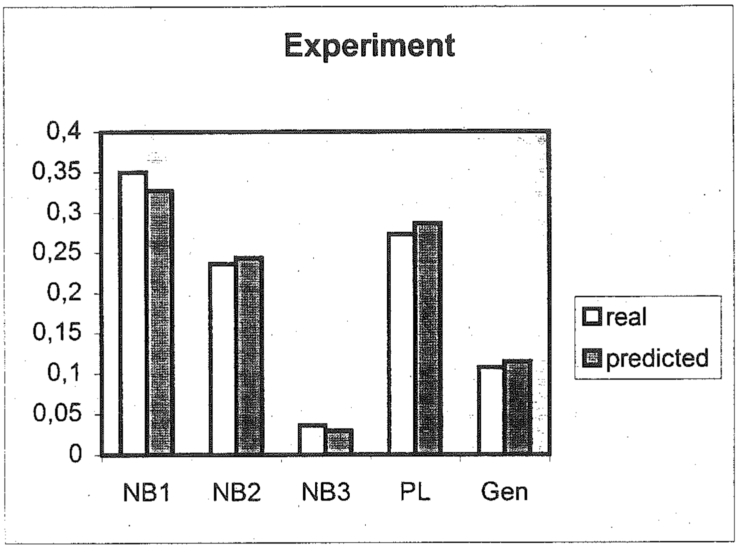


Figure 2 : Real versus Predicted Market Shares



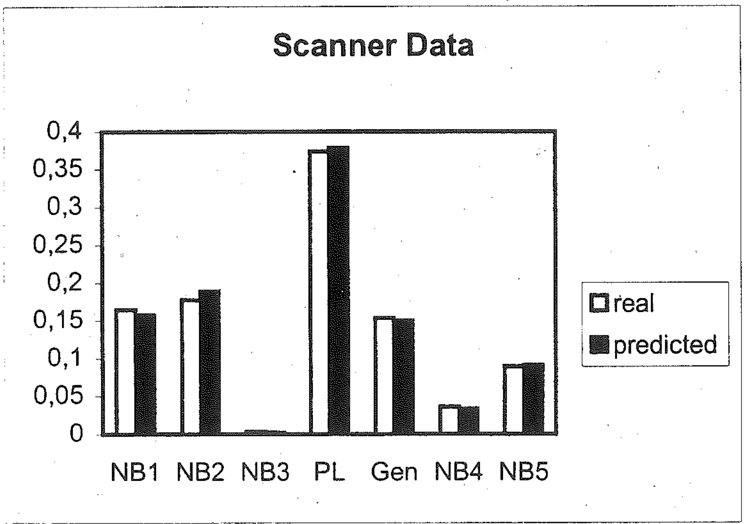
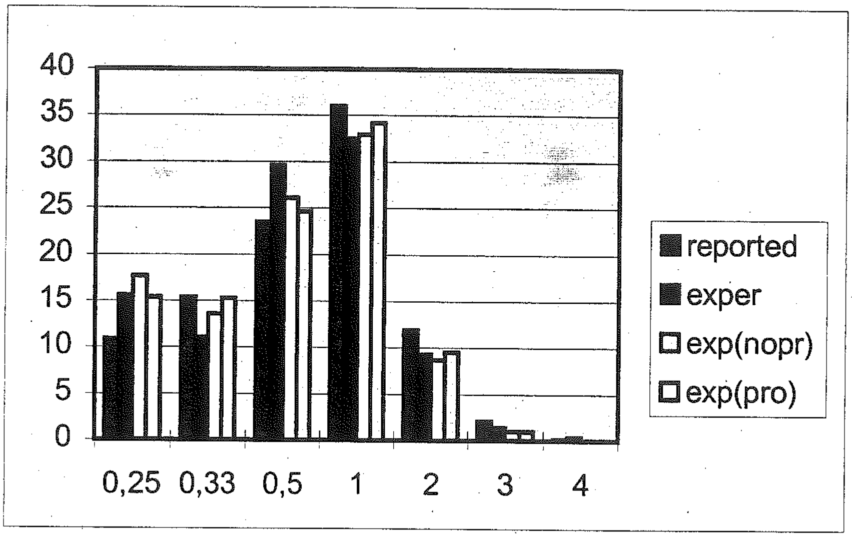
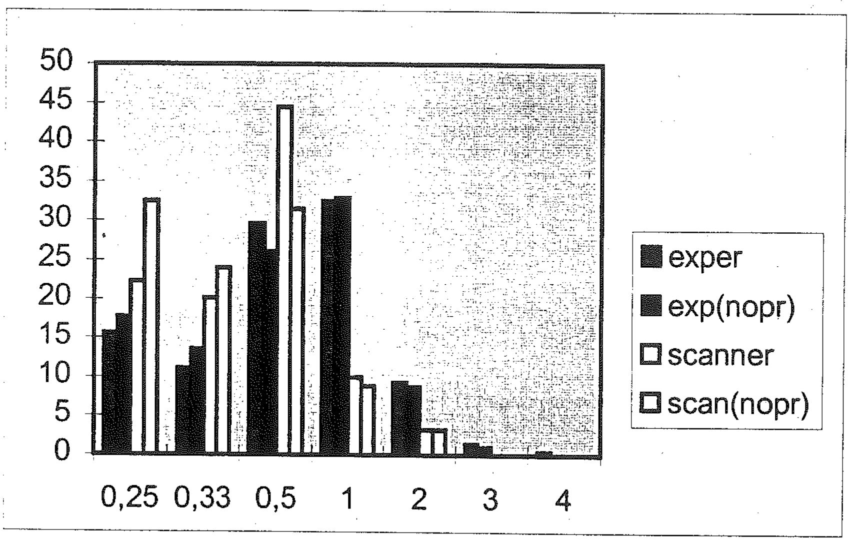


Figure 3 : Reported versus Observed Purchase Quantities (experiment)



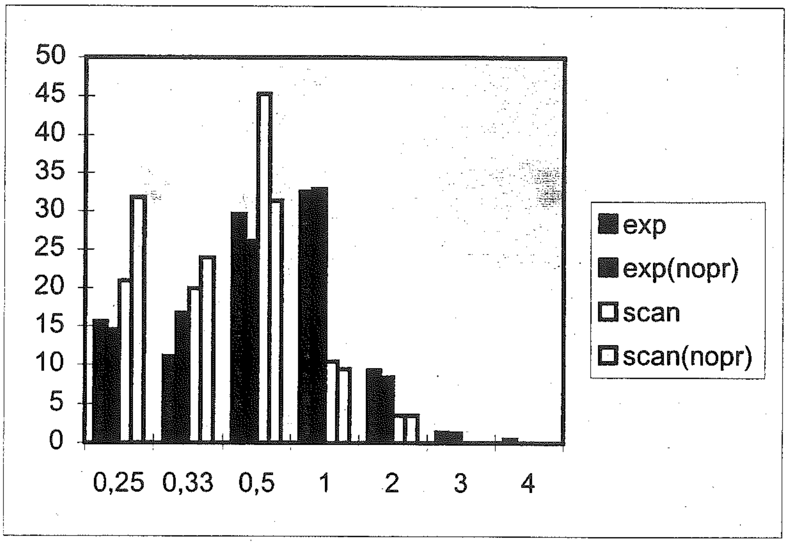
Reported : reported purchase quantity per week
Exper : observed purchase quantity per week (experiment)
Exp(nopr) : simulated purchase quantity, no promotion
Exp(pro) : simultated purchase quantity, with promotion

Figure 4 : Weekly Purchase Quantities in Experiment and Scanner Data Set



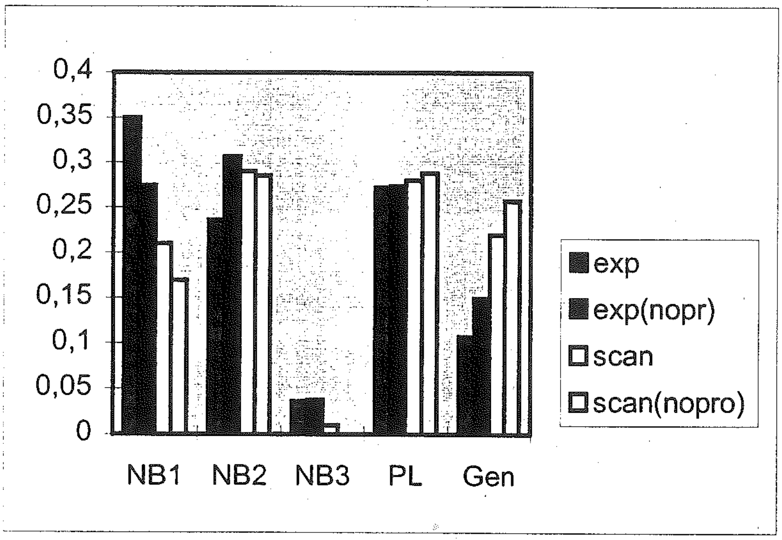
Exper : observed weekly purchase quantities, experiment
Exp(nopr) : simulated weekly purchase quantities, experiment, no promotion
Scanner : observed weekly purchase quantities, scanner data
Scan(nopr): simulated weekly purchase quantities, scanner data, no promotion

Figure 5 : Experimental vs Scanning Purchase Quantities, For scanning consumers with visit frequency>15



Exp : observed purchase quantities, experiment
Exp(nopr) :simulated purchase quantities, experiment, no promotion
Scan : observed purchase quantities, scanner data (visit freq.>15)
Scan(nopr): simulated purchase quantities, scanner data (visit freq.>15), no promotion

Figure 6 : Experimental vs Scanning Market Shares For Items Available in the Experiment



Exp : observed market share, experiment
Exp(nopr) : simulated market share, experiment, no promotion
Scan : observed market share (rescaled), scanner data
Scan(nopr): simulated market share (rescaled), scanner data, no promotion

Figure 7 : Purchase Quantity Pattern over Subsequent Weeks