

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/109608/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Greene, Max, Morgan, Peter and Foxall, Gordon 2017. NEURAL networks and consumer behavior: NEURAL models, logistic regression, and the behavioral perspective model. *The Behavior Analyst* 40 (2) , pp. 393-418. 10.1007/s40614-017-0105-x

Publishers page: <http://dx.doi.org/10.1007/s40614-017-0105-x>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Neural Networks and Consumer Behavior: Neural models, Logistic Regression, and the Behavioral Perspective Model

Abstract

This paper investigates the ability of connectionist models to explain consumer behavior, focusing on the feedforward neural network model, and explores the possibility of expanding the theoretical framework of the Behavioral Perspective Model to incorporate connectionist constructs. Numerous neural network models of varying complexity are developed to predict consumer loyalty as a crucial aspect of consumer behavior. Their performance is compared with the more traditional logistic regression model and it is found that neural networks offer consistent advantage over logistic regression in the prediction of consumer loyalty. Independently determined Utilitarian and Informational Reinforcement variables are shown to make a noticeable contribution to the explanation of consumer choice. The potential of connectionist models for predicting and explaining consumer behavior is discussed and routes for future research are suggested to investigate the predictive and explanatory capacity of connectionist models, such as neural network models, and for the integration of these into consumer behavior analysis within the theoretical framework of the Behavioral Perspective Model.

One of the most significant recent discussions in academic marketing is the explanation of consumer behavior. It is becoming increasingly difficult to ignore the fact that often research is preoccupied with a product or service, largely ignoring the brand level, and is limited to the act of purchasing, not taking into account consecutive consumption and consequences it generates. Traditionally, quantitative tools such as logistic regression have often been used to model consumer behaviors such as loyalty. However, real consumers are adaptive decision makers and connectionist models such as neural networks which operate in a learning mode ought to more naturally model their behavior. Furthermore, the logistic regression model is nested within the commonly used feedforward neural network (sometimes referred to as a multilayer perceptron) and this offers an excellent opportunity to evaluate any additional capacity that such feedforward networks might have to account for consumer behavior. The Behavioral Perspective Model (BPM) was proposed by Foxall (1990/2004) to provide a behavioral account of consumer behavior, largely drawing upon experimental analysis of behavior (EAB). Based upon radical behaviorism, EAB provides behavior-based explanation of performance response rates through environmental consequential causes. This explanation is contrary to inferred internal causes of cognitive theories, such as attitudes and intentions, which are in fact only behavior precursors and antecedents that do not provide direct causality in the radical behaviorism sense. The aim of this paper is to examine the ability of neural networks to model consumer behavior and, in particular, consumer loyalty.

Recently, researchers in many disciplines have shown an increased interest in applying connectionist concepts to the testing of established theories, and to identify new, promising areas for future research. Hence, the Behavioral Perspective Model (BPM) will be examined employing Feedforward Neural Networks (NN) in an attempt to provide a connectionist dimension to the BPM framework.

The dataset used for the analyses contains 52 consecutive weeks' purchasing of fast moving consumer goods – biscuits – and includes a number of product data categories and purchase event details, along with the demographics of over 1800 individual consumers available in the dataset.

1. The field of Consumer Behavior

Initial justification and motives for the BPM model development were drawn from the relativist perspective (Anderson, 1986), as it was argued that the predominant position of cognitivism may hinder theoretical progress through the suppression of pluralism of perspectives in the field of consumer behavior. Furthermore, propagation and juxtaposition of competing theories would provide alternative theoretical accounts of the subject studied, and result in advancing scientific progress. Consequently, the Behavioral Perspective Model (BPM) was proposed by Foxall (1990) to provide a behavioral account of consumer behavior, largely drawing upon EAB. Based upon radical behaviorism, EAB provides behavior-based explanation of performance response rates through environmental consequential causes. This explanation is contrary to inferred internal causes of cognitive theories, such as attitudes and intentions, which are in fact only behavior precursors and antecedents that do not provide direct causality in the radical behaviorism sense. This discontent with the cognitivist paradigm is explored further in the following section.

Information processing theories of consumer behavior, fundamental to cognitivism, assume consumers to be rational goal-seeking decision makers that rely on intellectual functioning and personal abilities to engage in extensive assessment of alternatives and information processing to achieve their goals. Even though these models, largely derived from cognitive psychology, occupy principal positions in the field of consumer behavior research, they have

been extensively criticized due to insufficient empirical correspondence, high levels of abstraction, and their inability to accurately describe and predict actual consumer behavior (Foxall, 1980, 1984).

Furthermore, many product categories include a number of brands that are practically indistinguishable in terms of their basic attributes, which comprise a repertoire of close substitutes that consumers choose from rather than showing a total loyalty to any one brand. Brand changes may occur often, with consumers making frequent brand substitutions over the sequence of purchases. It has been suggested that information processing, cognitive-based models have been unable to adequately predict these behaviors. To account for those aspects of observed consumer behavior that cognitive models are not always able to accommodate, simple behavioristic models have been proposed.

In the event of moderate brand commitment, consumer experiences with the product during the trial period determine inclusion into a repertoire of products subsequently considered for repeat purchase. A relatively simple and straightforward model proposed by Ehrenberg and Goodhardt (2014) shows repeat purchase as a function of a trial purchase and consumption, where trial is a function of awareness: awareness, trial, and repeat purchase. This suggests that awareness alone could only result in trial, whereas actual consumption could consequently lead to the product being adopted for subsequent repeat purchase.

Central to information processing theories is the assumption of the consumer being a rationally involved decision-maker, which has itself also been questioned. Research conducted shows consumers as exhibiting very restricted inclinations towards information processing, and as performing limited rational evaluation of brands based on their attributes (Foxall, 1984). Moreover, an increase in available information leads to increased consumer satisfaction, yet at the same time to diminished rational decision-making. Evidence suggests that consumers tend to drastically limit their information seeking behaviors, and in many instances purchases may not be preceded by the decision sequence described in information processing modeling at all (Foxall, 2009). Even in situations of repeat purchasing, strong brand attitudes expected to emerge according to information processing theories could not be observed. As a result, uninvolved and uncommitted consumer types emerge.

The following arguments further question the ability of cognitive and other similar theories to provide an accurate account of consumer behavior (Wells, Chang, Oliveira-Castro, & Pallister, 2010). (1) These theories are claimed to be incomplete as they fail to accurately identify factors that account for internal events that cause behaviors, such as environmental precursor. (2) They are fictional in the sense that the internal causes that explain behaviors are inferred from the observations of the very behaviors they are supposed to elucidate; and (3) these theories are unnecessary as behavior could be explained and predicted using simpler behavioral theories (Occam's razor) that offer a more direct approach to obtaining knowledge without relying on the explanatory power of unobservable events and circular reasoning. This is not to say that theorizing is not present in behaviorism – just the theory making that relies on unobservable events. As a result, behaviorist research can often be found in counter-position to cognitive theories, and is aimed at re-examining principal assumptions of cognitivism to further develop the understanding of the field of consumer behavior by adopting a relativist approach (Yan, Foxall, & Doyle, 2012).

Contrary to cognitive and other comparable theories that attempt to explain behaviors through some internal processes, behaviorism avoids any explanation of behavior through mental, neural, and conceptual means, or other hypothetical constructs. Instead, EAB strives to explain observable behavior through contingent, environmental stimuli, following the process

where response frequency is controlled by antecedent signals and consequent stimuli – reinforcers and punishers. The importance of situational variables in determining behavior is particularly emphasized, highlighting events that precede, indicate and follow behavior. This three-term contingency could be exemplified through operant theory ($S^D - R - S^{R/P}$), in which responses (R) are reinforced or punished ($S^{R/P}$) in the presence of discriminative stimuli (S^D). As a result, reinforcing or punishing behavior would result in an increased or decreased response frequency respectively in the future under similar conditions. Extensive research has been able to demonstrate this conceptual framework to be accurate in a wide range of situations.

2. Behavioral Perspective Model

In order to incorporate consumer behavior within the behaviorism doctrine, a model was constructed according to the EAB that considers the arguments pointed out above. Subsequently, the BPM (Foxall, 1990, 2009) proved to be a constructive addition to the field of behavioral economics. The model could be described in the following manner: it depicts the rate of consumer behavior as a function of setting openness where the behavior takes place and Informational and Utilitarian reinforcers are available immediately or potentially in this setting. It is possible to identify in the BPM the three-term contingency discussed above, adapted to operate in the consumer situation and taking consumer learning history into account. As a result, the BPM forms an environmental perspective on consumer behavior, incorporating situational influences into purchasing behavior. From the modeling viewpoint, consumer behavior could be expressed through a consumer's learning history, behavior setting, and resulting consequences of behavior (Figure 1). To provide a better representation of the model, these variables will be further elaborated on below.

Behavior setting can be described as consisting of not just physical, but also social environments that provide signaling stimuli for a consumer choice event. Settings, ranging from closed to open, offer varying degrees of suggested consumer responses and levels of control over behavior. For instance, a dental office would suggest a very limited scope of behavior choice, as patients are assumed to follow the established protocol and procedures, whereas going out at the weekend offers a much broader choice of behavior.

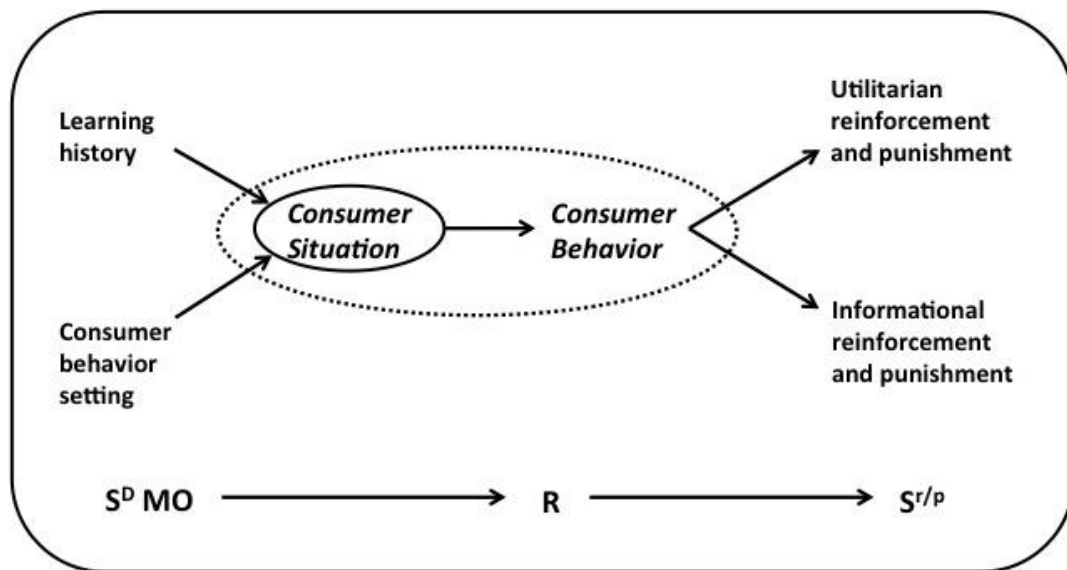


Figure 1. Summative representation of the Behavioral Perspective Model (BPM). *Source:* Foxall, G. R. (2016). *Perspectives on Consumer Choice: From Behavior to Action, From Action to Agency*. London and New York: Palgrave Macmillan.

Learning history also contributes to the consumer situation: it provides a capacity for consumers to interpret the stimuli available in the consumer behavior setting. Referring back to the previous experiences of encountering similar behavior settings, consumers are able to predict possible consequential outcomes of behavior in the current setting. In behaviorist terms, consumers acquire the ability to discriminate between stimuli that, depending on the consumer's behavior, offer one of the three types of consequence available: (1) Utilitarian Reinforcement that refers to the functional benefits that the purchase and consumption of product or service is able to provide, (2) Informational Reinforcement that refers to symbolic consequences of behavior, and (3) aversive outcomes of behavior which are monetary or other costs incurred as a result of the behavior. To exemplify, the Utilitarian Reinforcement of going abroad for vacation would be the health benefits of changing one's usual environment and, depending on the destination, the time to relax in a warmer climate. Informational Reinforcement, in contrast, refers to the social status and other symbolic consequences of travelling, such as the admiration of others. For instance, one might travel to visit a prestige location, or visit friends and family. Aversive outcomes of travel would include the monetary costs of travelling along with time spent planning and choosing the right destination and other details of the trip. It is argued that all products and services include Utilitarian, Informational and aversive consequences of varying intensity (Foxall, 1990, 2009). Much like the scope of behavioral setting, reinforcers operate on a continuum basis from low to high.

Thus, purchase probability is dependent on consequential reinforcing and aversive outcome strengths signaled by aspects of the consumer behavior setting. In relation to this, product and service attributes could be understood as reinforcing and aversive factors, and suppliers (including manufacturers and distributors) aim to modify these factors to make their product or service appear more appealing to consumers. It is difficult to predict whether these planned reinforcers would actually work, and one of the central questions in marketing literature aims to identify what events and to what extent could actually serve as consequent reinforcers or punishers of consumer behaviors.

Following the work of Skinner, operant conditioning has been defined as behavior that is controlled by its consequences, and employs a method of training with Reinforcement that follows a particular schedule. Any procedure that delivers reinforcers following a specific rule could be defined as a Reinforcement schedule. When speaking of any reinforced behavior, an integral part of the analysis is the suggestion of a Reinforcement schedule through which behavior is maintained over time. In a laboratory setting, it is possible for an investigator to control the schedules while response patterns are examined. In operant conditioning, typical Reinforcement schedules are fixed interval (FI) and fixed ratio (FR), and variable interval (VI) and variable ratio (VR). Under FI schedules, Reinforcement is presented every n th amount of time; and after every n th response under FR schedules. Under VI schedules, Reinforcement is presented on average every n th amount of time. Under VR schedules, the number of responses required for Reinforcement varies in each trial. In the real world however, it is rarely possible to classify complex, social, human behavior maintenance according to any definitive schedule. As a result, buying and consumption behavior research rarely follows any strictly-enforced schedule in the experimental laboratory sense. The unit of analysis is also defined using a much broader scope or on a much broader scale, including not only the instance of purchase or consumption, but also pre-purchase and post-purchase responses (Foxall, 1990, 2009).

Consumer behaviors can be classified into four overarching operant classes: accomplishment, pleasure hedonism, accumulation, and maintenance (Figure 2).

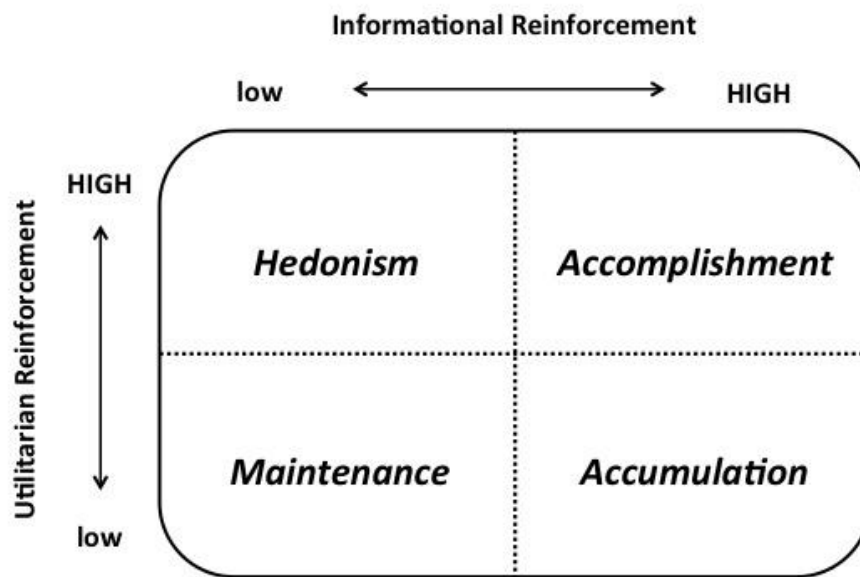


Figure 2. Pattern of reinforcement and operant classes of consumer behavior. *Source:* Foxall, G. R. (2016). *Perspectives on Consumer Choice: From Behavior to Action, From Action to Agency*. London and New York: Palgrave Macmillan.

Behaviors classified as *accomplishment* are maintained by high Utilitarian and Informational levels of Reinforcement and may include behaviors such as acquisition and conspicuous consumption of status symbols and activities that seek sensation and excitement. Behaviors classified as *hedonism* are characterized by high Utilitarian and low Informational levels of Reinforcement and are usually positively reinforced by popular entertainment or negatively reinforced by behaviors such as taking analgesics. Behaviors classified as *accumulation* may include collecting and saving (such as loyalty programs) and are maintained by high Informational and low Utilitarian Reinforcement. Behaviors classified as *maintenance* are necessary to sustain one's social and physiological being and include fulfillment of duties to the society characterized by low levels of Utilitarian and Informational Reinforcement.

As a result of the three dimensions of the theory – Informational Reinforcement, Utilitarian Reinforcement, and behavior setting scope – eight environmentally-located contingency categories emerge (Figure 3):

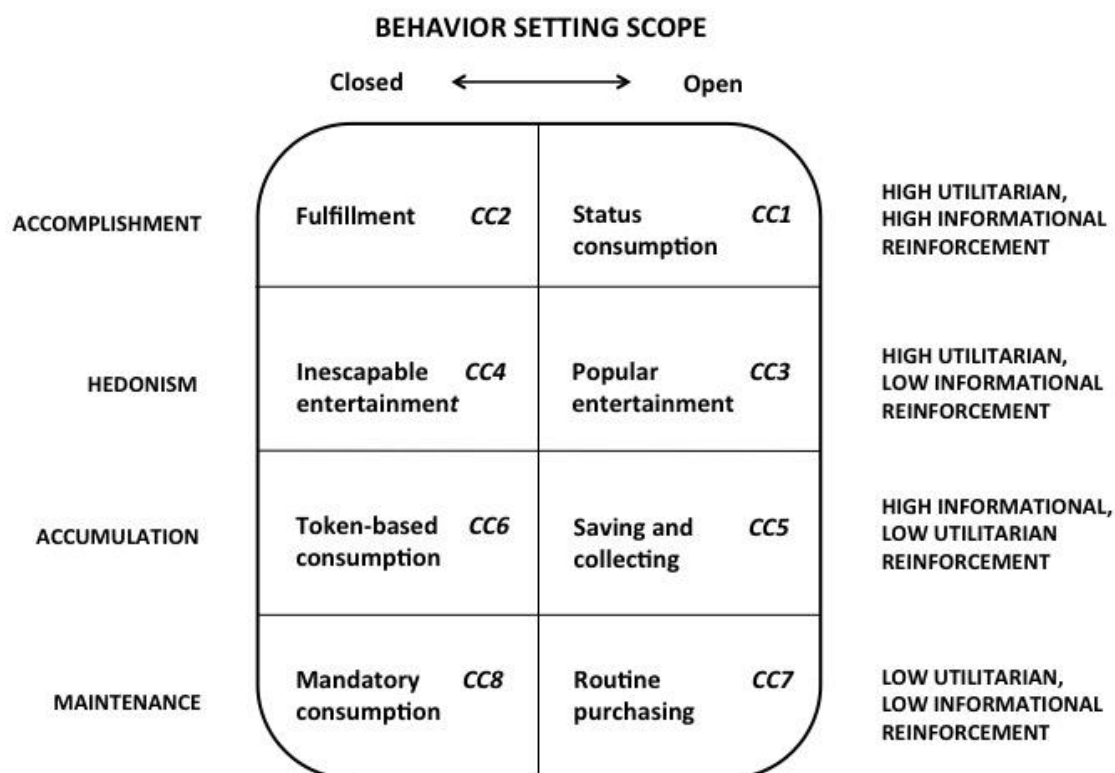


Figure 3. The Behavioral Perspective Model (BPM) Contingency Matrix. *Source:* Foxall, G. R. (2016). *Perspectives on Consumer Choice: From Behavior to Action, From Action to Agency*. London and New York: Palgrave Macmillan.

Several interesting patterns can be derived from the examination of these categories. A general relationship between Reinforcement and setting could be suggested: behavior setting influence declines as Reinforcement variables influence increases, maintaining behavior on high variable schedules. A more detailed discussion and interpretation of the four operant classes and contingency categories can be found in Foxall (2010).

Over the last decades, the BPM has proved to be a useful tool in explaining consumer behavior. In the following section, recent research areas that employ BPM are examined.

3. BPM research overview

Since the BPM was first proposed, the framework has been useful in examining several strings of research in consumer brand choice (Foxall & James, 2001, 2003; Foxall, Oliveira-Castro, James, Yani-de Soriano, & Sigurdsson, 2006; Foxall, Oliveira-Castro, & Schrezenmaier, 2004; Foxall, Wells, Chang, & Oliveira-Castro, 2010; Foxall, Yan, Oliveira-Castro, & Wells, 2011; Oliveira-Castro, Diogo, Foxall, & Schrezenmaier, 2005; Oliveira-Castro, Foxall, & James, 2008; Oliveira-Castro, Foxall, & Schrezenmaier, 2005; Oliveira-Castro, Foxall, & Wells, 2010; Oliveira-Castro, Foxall, Yan, & Wells, 2011; Wells & Foxall, 2011). The following sections will review a number of studies that reflect on the usefulness of the matching law, operant conditioning, and behaviorist perspectives as applied to consumer behavior.

3.1. The matching law and substitutability of brands

Choice is interpreted by behaviorists as the relative rate at which behavior is performed, rather than a single event. Choice is a distribution of behavior over time, a proportion of choosing one thing over the other (Baum, 1974; Herrnstein, Rachlin, & Laibson, 1997). Contrary to cognitivist explanation, the behaviorists' explanation of choice is done through environmental events that increase (Reinforcement) or decrease (punishment) the probability of repeat behavior, and not through mental constructs. Therefore, the behavioral analysis of choice includes the analysis of alternatives to identify the Reinforcement configuration that maintains it.

In the context of consumer behavior, research into choice could be said to follow Herrnstein's (1961) influential experimental work with pigeons, where he discovered what is referred to as matching law. The matching law defines the predisposition of choice preference to follow the reward the alternative provides. This could be illustrated with a simple example of two alternative choices, with one of the alternatives offering a reward twice the amount of the other. According to Herrnstein's (1961) matching law the alternative offering the higher reward, twice the amount, will attract a choice frequency twice as high as the other alternative.

Even though a number of limitations have been acknowledged elsewhere (Baum, 1974; Donahoe & Palmer, 1994; Herrnstein et al., 1997), the law defines choice in terms of response strength – which fits the field of research in operant choice. The law provides a quantification for choice behavior (Herrnstein, 1970), and offers a predictive explanation of choice. Expressed arithmetically, the matching relation in terms of ratios takes the following form:

$$\frac{B_1}{B_2} = b \left(\frac{r_1}{r_2} \right)^s$$

where B_1 and B_2 are the frequencies of responding to the alternatives 1 and 2, and r_1 and r_2 are the numbers of Reinforcements. Parameters b and s are determined empirically and represent bias and sensitivity to changes in Reinforcement ratio (Baum, 1974, 1979).

Alternatively, it is possible to interpret exponent s as a measure of substitutability between reinforcers (Rachlin, Kagel, & Battalio, 1980), where s equal to 1.0 signifies perfect substitutability. As a result, matching law could be useful in interpreting substitutability between different products of brands.

Applying this construct to consumer data, Foxall (1999, 2017) provided a basis for gathering empirical evidence for Ehrenberg's theoretical account of sequential patterns of consumer brand choice and multi-brand purchasing. Among other things, the analysis confirms the explanation of multi-brand purchasing through brand similarity, where functionally comparable brands are substitutes for each other, reducing consumer loyalty rates. Multi-brand purchasers prefer a selected range of brands, suggesting brand indifference and multiple systematic patterns of brand preference. Underlying behavioral mechanisms of consumer choice are further investigated on an individual consumer level by Foxall and James (2001) using a small consumer sample. Results suggest that individual consumer brand choice follows theories of matching (Herrnstein, 1997) and maximization (Kagel, Battalio, & Green, 1995), displaying predicted sensitivity patterns for brands. Underlying mechanisms of choice however, follow neither theory entirely but rather involve the maximization of Utilitarian and Informational rewards that products provide. Due to small price difference of seemingly analogous, competing products, there is little expectation in marketing literature

for individual consumers to maximize, giving emphasis to parameters of the marketing mix that determine consumer choice. In contrast, results of empirical investigations illustrate the relationship between the price and the quantity bought through sensitivity and bias measures. Relative prices are demonstrated to correspond with the total Utilitarian and Informational Reinforcement the consumer receives (Foxall & James, 2003).

A more extensive, 80-consumer panel data was used in further analyses following the same line of research into multi-brand purchasing (Foxall et al., 2004; Foxall & Schrezenmaier, 2003). Sensitivity and bias (s and b parameters of Equation 2.1) were shown to be very close to 1.0 suggesting perfect matching: brands that comprise the consumer preferred product subset tend to act as substitutes. Analyses of additional product categories support multi-brand choice, where individual consumers decide on brands from their preferred subset in no particular order, exhibiting both maximization and matching. A small group of consumers however, choose particular brands exclusively. Some are price-insensitive and prefer only high end (prestigious) brands, maximizing Informational Reinforcement. The behavior of others is particularly price-sensitive and elastic, as they opt for the cheapest brands, maximizing Utilitarian Reinforcement. All other consumers adopt behavior that entails higher product diversity (Foxall & Schrezenmaier, 2003). Further investigation reveals that consumers acquire their preferred subset of brands guided by the Utilitarian and Informational Reinforcement level offered by the brands (Foxall et al., 2004). This is significant for marketers as it opens the discussion for customer segmentation based on clearly distinct consumer categories, based on interconnections of Utilitarian and Informational Reinforcement levels. Different consumer categories were shown to provide varying levels of reaction to price alterations. Price elasticities could be further segregated into intra-brand elasticities that represent a response to the aversive consequences of giving up money, and inter-brand elasticities: Informational and Utilitarian (Foxall et al., 2004). As a result, choice patterns could be established around the avoidance of aversive consequences and the maximization of Utilitarian and Informational Reinforcement.

These findings were further confirmed by the later study (Oliveira-Castro et al., 2010) that employed the AC Nielsen Homescan™ panel dataset that includes more than 1500 British consumers purchasing four grocery product categories for 52 weeks. BPM proposes the combination of the behavioral economic tools such as the matching law analysis of brand choice (where amount of money spent is expressed as a function of the quantity and Informational and Utilitarian levels of the brand bought) with the Utilitarian, Informational and aversive consequences. The study suggests that this combination provides a useful framework to study consumer behavior (Oliveira-Castro et al., 2010).

3.2. Brand repertoire

Levels of Utilitarian and Informational Reinforcement could be useful in classifying brands into distinct categories when analyzing consumer brand repertoire – a set of preferred brands that consumers tend to buy. To examine the relationship between repertoire and Reinforcement levels, the same 80-consumer panel data was used to develop Utilitarian and Informational rankings for each brand (Foxall et al., 2004): two levels of Utilitarian and three levels of Informational benefit. Using this classification, consumer purchasing patterns were analyzed. Results indicate that most consumers (over 70%) buy brands within the same Informational level and the same Utilitarian level, suggesting consumer brand repertoires are associated with the level of Informational and Utilitarian benefits the brand is able to provide.

To further explore whether it is possible to successfully use the features of the BPM for market segmentation, the purchasing patterns of consumers in the UK biscuits market were

explored using the same large AC Nielsen Homescan™ panel dataset. (The same dataset employed in this paper.) Segments explored included the six segments used by Foxall et al. (2004) supplemented by the segments derived from the demographic variables. It was established that all consumers are sensitive to price changes, and are more sensitive to intra-brand (price) changes over changes in Informational and Utilitarian Reinforcement, supporting the previous findings (Foxall et al., 2004; Oliveira-Castro, Foxall, et al., 2005). Segmentation based on single demographic variables was found to be useful, but effects of combined demographic variables on segmentation remain to be explored.

The same data was used to describe brand substitutability and to identify potential product clusters within the same product category (Foxall et al., 2010). It is generally expected that brands within the same product category will act as perfect substitutes, whereas results actually show that subcategories perform as separate products. Thus, the application of behavioral economics' methods to the analysis of consumer choice suggests that the matching law provides a functional and quantifiable classification technique able to differentiate between brands and products useful in the marketing field.

4. Connectionism

Old behavioral economics relied heavily on the insights of Simon (1982). Despite his pioneering work on the serial symbol processing hypothesis in cognitive psychology and artificial intelligence, Simon's contributions (1982, 1987) are rather outdated in the face of the current focus on parallelism and connectionism (Sent, 2004). Connectionism is a philosophical framework that models mental or behavioral phenomena as the emergent processes of interconnected networks of simple units. Some of the more commonly encountered forms of connectionism rely on the use of neural network models. Central to connectionism is the principle that it is possible to describe mental processes by interconnected networks of simple uniform units that represent neurons and connections that represent synapses. Most networks tend to change over time, and incorporate a concept of activation. A computational unit within the network has a numerical activation value, which could represent the probability of that neuron producing a response. Spreading activation models allow extending activation to the other interconnected units over time, and are a common feature of the NN modeling discussed later.

If one accepts the underlying assumptions of connectionism that suggest that the study of mental processes is the study of neural systems, this acceptance establishes the link between the connectionist framework and neuroscience. NN models offer a relative level of biological realism, as the models are based on the architecture of the brain and were originally designed to model brain functionality. As a result, some neural network researchers use NNs to model the biological neural systems. A clear link between neural activity and cognition is an attractive aspect of NN models (Smolensky, 1995). One field, for example, that has embraced NN models in recent years is psycholinguistics. The field of animal learning and cognition and the field of ecological modeling have also been increasingly receptive to the possibilities that connectionist models are able to provide.

In the field of neuroscience, neural coding is concerned with the underlying mechanisms that explain how information is represented in the brain by neuron networks. It is believed that both digital and analog information could be encoded by the neurons. The aim of neural coding is to explain the relationship between the stimulus and the neuronal response, as well as the electrical activity between the neurons.

4.1. The connectionist model features

The most common connectionist models used today are neural networks. They operate under the assumptions that it is possible to describe the mental state as a multi-dimensional vector containing the numeric activation values for the computational units within the network, and that the gradually modified connection strength (weights) creates memory. Variations in the models come from the interpretation of the neurons, the activation function, and the learning algorithm employed to train the network.

The importance of learning and training the models is usually emphasized by the connectionist researchers, and numerous complex learning algorithms to train NN models have been devised. When talking about NN models, learning takes the form of gradually modifying the connection weights. This is generally accomplished through the application of mathematical and statistical algorithms that determine the change in connection weights. Gradient descent over an error surface defined by the weights matrix is a common strategy in connectionist learning. One of the most popular gradient descent algorithms employed in connectionist models is backpropagation, which involves weights adjustment by the partial derivative of the error surface. The mathematical framework that is the foundation of most connectionist models today was proposed as part of the parallel, distributed, processing approach (Rumelhart & McClelland, 1987) that emphasized neural processing nonlinearity.

Even though the relation of the neural network models and the biological architecture of the human brain is debated, as little is known about the actual functionality of the brain, NNs have traditionally been seen as simplified neural processing models. The degree of complexity and individual properties that computational units should have to accurately mimic the functionality of the brain for representative purposes is yet to be determined. From the computational view however, contrary to the traditionally inclined algorithms predominant in computer technologies that follows a sequential processing and instructions execution in an automated predefined manner, neural networks attempt to model the information processing in a way similar to biological systems that rely on parallel nonlinear processing and pattern recognition. As a result, the very core of a neural network is not just an algorithm tasked with sequential execution of predetermined commands but rather a very complex statistical processor.

4.2. Artificial neural networks

Even though computational models based on NNs were developed many decades ago (Hebb, 1949), technological and computer science advances in recent decades are facilitating the growing interest that researchers express towards using the NNs to study a number of diverse phenomena in statistics, cognitive psychology and artificial intelligence (Ripley, 1996). Originally developed for representational purposes to model the functionality of the human brain (Bishop, 1995), NNs have since lost that as a primary function and are increasingly utilized as a method of analysis in predictive modeling and forecasting (Adya & Collopy, 1998).

Inspired by structural and functional features of biological neural networks (non-linear distributed information processing), NNs normally comprise a group of simple processing units, or artificial neurons, interconnected by synapses, and are able to display a complex global behavior determined by connections between the processing units. Whereas real neurons send information along axons and dendrites using electrochemical pulses the body of the neuron integrates the incoming excitatory and inhibitory dendritic signals and fires if their resultant exceeds a threshold, McCullough and Pitts (**ref**) described a mathematical abstraction of a biological neuron where input values may be positive (excitatory) or negative

(inhibitory) and their sum is subject to an activation function before the neuronal unit outputs a signal (usually 0 or 1 or -1 to +1 or ranging between the two. Information is processed employing the connectionist approach as follows: functions are performed in parallel by the units, rather than clearly assigning subtasks to various unit groups. In most cases NNs are adaptive systems, able to adjust their structure by fine-tuning the strengths (weights) of the connections in the network according to external or internal information flow – typically during the training stages (Haykin, 1994). Today, NNs are often used as statistical techniques designed to find patterns in data or to model intricate relationships between dependent and independent variables. Often the neural network is emulated by software rather than realized in hardware form.

The interconnections between the neurons in the different layers form the network part of NN models. An example of a simple one-layer feedforward neural network is shown in Figure 4. The first layer contains the input neurons, which send the data by means of synaptic connections to the second layer of neurons, and by the means of more synaptic connections, to the third layer of output neurons. This is the most commonly encountered and relatively simple architecture as it contains only one intermediary hidden layer (although even simpler input-output NNs with no hidden layers are possible) and no skip layer connections (connections that would in the Figure 4 go from the input layer straight to the output layer bypassing the hidden layer), and more complex architectures would include more layers and an increased number of neurons within each layer. The synaptic connections hold the weights values that are used in the computations.

Three factors define the type of NN model: (1) the pattern of interconnection between the neurons, (2) the learning mechanism employed to update the weights, and (3) the activation function to convert the weighted input to its output activation.

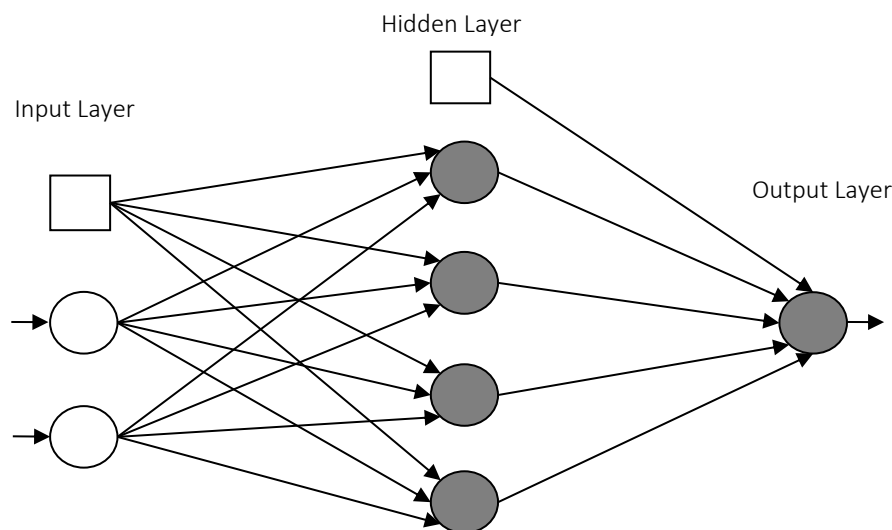


Figure 4. Feedforward NN architecture that includes 3 layers: 2 nodes in the input layer, 4 nodes in the hidden intermediary layer, and 1 output node. The activation function is usually logistic. Unshaded nodes are linear, shaded nodes apply activation function.

The use of bias nodes is very common in feedforward neural networks. These represent a constant value of unity which is fed in to a neuron to change the threshold level at which the neuron fires. A typical and the most commonly used activation function is the logistic function the output of which is the open interval (0,1).

$$Output = \frac{1}{1 + \exp(-Total\ Input)}$$

where Total Input = Bias Weight + Weight₁Input₁ + Weight₂Input₂ + ... + Weight_nInput_n

The bias weights act in the same manner as the intercept in a regression equation and are merely added to the total input from the neurons in the previous layer.

From the mathematical perspective, a network function of a neuron $f(x)$ could be decomposed into $g_i(x)$ functions, which are possible to decompose further (Figure 5). This is possible to interpret in a number of ways. First is the functional view, where the input x is transformed into a three-dimensional vector h , which is transformed into a two-dimensional vector g , which is transformed into f . Second is the probabilistic view, where the $F = f(G)$ depends on $G = g(H)$, which depends on $H = h(X)$, which in turn depends on x . As seen in Figure 5, this network structure defines the individual layer components as remaining independent of each other, allowing parallelism in the architecture implementation.

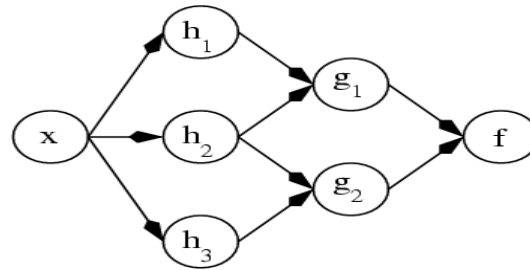


Figure 5. NN dependency graph represents a network structure, with arrows depicting between-variable dependencies.

One of the major research directions in the field aims to establish NN models as a powerful and versatile method of analysis – often employing comparative design contrasting neural networks with other traditionally employed methods (Bishop, 1995). As a result, it is often reported that neural networks not only perform as well as other methods considered, but also often outperform traditionally employed approaches tasked amongst other things with segmentation and targeting (Adya & Collopy, 1998). As it seems to be the case in consumer behavior literature that ongoing research is largely concerned with identifying underlying patterns involving stimuli, it is only natural to attempt examination of consumer behavior with NNs (Curry & Moutinho, 1993).

4.3. NNs and Consumer Behavior

Research on the application of neural networks to the analysis and modeling of the consumer response to advertising stimuli was published by Curry and Moutinho (1993) where a comprehensive discussion of theoretical implications of neural networks is followed by practical application considerations. The authors suggest expert systems as one of the possible applications, but caution about limitations and potential overoptimistic notions in the field. Alternative artificial intelligence based application is suggested: neural networks. A typical NN input-output structure supplemented by a number of intermediary hidden layers brings certain advantages through a more sophisticated platform for modeling consumer behavior as intermediary levels have a tendency to be linked with important conceptual phenomena predisposed to indirect measurement. Another imperative for the consumer behavior concept of NNs is that models are trained: either through a supervised learning process where example connections of input and output pairs are fed into the model, or otherwise through relying on clustering methods in unsupervised learning (Curry &

Moutinho, 1993). The advantageous ability to extrapolate rules from training sample data puts NNs in a superior position compared to rule-based arrangements common in expert systems. However, this ability places heightened importance on the selection of cases for the training sample as selection procedures would eventually impact the model performance. The authors conclude by suggesting that neural networks are particularly appropriate in tasks that involve a concept of cognitive behavior or pattern identification that is similar to the examination of consumer economics. In order to consider the application of artificial neural networks to a dataset composed of fast-moving consumer goods, similar to data used by Foxall and colleagues (Foxall, 2003; Foxall et al., 2004; Foxall & Schrezenmaier, 2003), a number of relevant articles are reviewed in the following sections.

Van Wezel and Baets (1995) test the predictive performance of neural networks and compare it with traditional techniques in their paper on evaluating market response through the examination of variables on fast-moving consumer goods. They suggest a number of different choices to tackle the complex market response estimation task, including more commonly employed statistical models such as multiple linear regression and multiplicative model, and compare their predictive power with what authors call the best known type: the back-propagation neural network approach. The innate configuration of neural networks does not require any prior knowledge about the model structure as it is established through training, and therefore does not require any assumptions about the input and output relationship (Van Wezel & Baets, 1995). This ability (network structure does not need to be predetermined), also suggested by Curry and Moutinho (1993), provides a powerful modeling arrangement.

Some of the problems with neural networks were discussed as well, such as overfitting, when the model fit to the training set is so high that the model does not perform well with external data. Using comparative analysis, models were evaluated producing the result showing neural networks outperform other traditional methods in all the cases tested. As a result of this outperformance, it is then suggested that neural networks, if applied correctly, should be a good alternative to the market response models commonly used. However, the NN model is often viewed as a “black box”, where theoretical interpretation of the process might pose a difficulty (Van Wezel & Baets, 1995). It is also important to remember that attempts to explain complex phenomena with comparatively simple techniques such as linear regressions could be oversimplifying the interpretation, and neural networks could provide a preferential option in this matter. This suggestion was also expressed by Curry and Moutinho (1993). Van Wezel and Baets (1995) suggest a possible extension of research into the use of recurrent neural networks to model market behaviors, as such networks are capable of working with effects that are not immediately occurring.

Another study reports the findings of two experiments into a comparison of neural networks with discriminant analysis and logistic regression in terms of their ability to predict consumer choice (West, Brockett, & Golden, 1997). It is argued that even though neural networks are built to quantitatively imitate the neurophysiological structures and decision-making ability of the human brain, they nevertheless, express resemblance to linear modeling from a statistical perspective. It is also suggested that neural networks may be useful in predicting consumer choice. It is again argued that application of neural networks to study consumer behavior choice poses benefits unmatched by other statistical methodologies, such as their ability to detect nonlinear and noncompensatory processes without prior supposition of parametric relationships between variables such as product attributes and consumer behaviors, already allegedly suggested by others (Curry & Moutinho, 1993; Van Wezel & Baets, 1995). Through empirical work, neural networks models were shown to consistently outperform traditional statistical approaches in predicting the outcome of noncompensatory

rule. The robustness of neural networks has also been discussed, and issue of overfitting addressed through the use of a validation sample in determining training termination. One can conclude then by stating that the neural network exhibits exceptional predictive capabilities in comparison to traditional analytical approaches, and offers great usefulness in predicting consumer choice based on product attributes, assuming that the main goal in consumer research is to predict behavior (West et al., 1997).

In the analysis of supermarket shopping behavior, neural networks were used to predict customer satisfaction, number of trips to the supermarket, and the amount spent (Davies, Goode, Moutinho, & Ogbonna, 2001). Advantages of using the neural network in this type of analytical work are stated by authors as follows: the neural network's learning capacity that allows sophisticated approximation which does not require the researcher to specify underlying relationships prior to research, and values of hidden nodes that could be interpreted as unobservable consumer behavior variables. Davies et al. (2001) proceed by building a number of neural networks and found that broad product range and quality exhibits the highest influence on customer satisfaction. They also found that customers with higher income were among those most satisfied, as such customers could take full advantage of choices offered and could travel longer distances to reach those supermarkets with higher available selections (Davies et al., 2001). Other shoppers were found to be more concerned with reasonable prices, and store atmosphere, which could have important managerial implications and considerations such as staff training programs. It seems that customer dissatisfaction comes from a feeling of restriction in their choice, either through limited range of choices or restricted purchasing power, and these could often be interconnected. Authors cautioned that customer satisfaction should not be linked with spending, as the model suggested that only disposable income impacts spending directly, with other factors playing a small part (Davies et al., 2001).

One of the more recent articles discusses the introduction and application of neural networks in consumer behavior settings, and explores attitudes towards a web site (Moore, Beauchamp, Barnes, & Stammerjohan, 2007). The fact that choice is based on noncompensatory rules is noted by the authors, along with the nonlinearity of the preference concept. Also noted are the statistical approximation approaches such as linear modeling, which is frequently applied in attempts to determine consumer attitudes. Moore et al. (2007) also state that referential decision-making contributes to new methods such as neural networks being considered. Through a demonstration of a neural networks model, it is suggested that the process of decision-making could be successfully examined with the use of neural networks considering a number of environmental and personal influences. Innovative consumer behavior theory could benefit from neural networks application as well (Moore et al., 2007). Practical beneficial outcomes of an effective neural networks predictive model could include better inventory management and merchandising, improved online experiences for customers, successful employee profiling and training, and strategic decision-making.

5. Methodology

5.1. Research questions and hypotheses

Consumer behavior as a field of study benefits from the contributions from a number of interrelated disciplines, including economics, marketing, sociology, philosophy, and psychology (Bashford, 2009; Calder & Tybout, 1987; Holbrook, 1987; McKee, 1984; Pachauri, 2002). Highly quantitative research is common in the field (for example Cornwell et al., 2005; Cunningham, Young, Moonkyu, & Ulaga, 2006; Güneren & Öztüren, 2008; Lu

Hsu & Han-Peng, 2008; van Kenhove, Vermeir, & Verniers, 2001; Watson & Wright, 2000), and the present research paper is no exception.

Here we deal with discrete choice problem where examples of such problems could be a choice between two product categories or brands. Models employed here could be subdivided into two distinguishable types. On the one hand, models could be characterized as highly quantitative, linear, and require a number of assumptions to be met (e.g. normal distribution) in order to perform properly. Logistic regression has been identified as particularly suitable for the type of a problem such as the one in our case, and is indeed popular in similar research (Adya & Collopy, 1998). On the other hand NN models are also highly quantitative and computationally demanding, but are non-linear and yet do not require a predetermined structure. Once logit models are developed to the highest potential, simple NN models without hidden layers are trained to be compared with the results of the logit models. Since the two-layer neural network where inputs are connected directly to the output is equivalent to a logit model linear in its independent variables and with no interaction terms, practically there should be no difference between the fit. This is a useful initial test of the method. Then we can proceed to more complex nonlinear NNs with a hidden layer having 1, 2, 3, etc. to provide greater and greater capacity for nonlinearity.

As indicated above, the main research questions are concerned with examining the predictive power of different methods and models that could be useful in explaining consumer choice. One dimension that has kept the interest of consumer behavior researchers over decades is consumer loyalty (Oliver, 1999) and we use loyalty as a typical example of consumer behavior in order to see if .

- a) Consumer behavior models based on NNs can provide better predictive power than those based on traditional techniques such as logistic regression and
- b) Consumer behavior models based on NNs provide better explanatory power than those based on traditional techniques such as logistic regression.

In addition to the analyses of predictive power of different models in contrast to NNs, this research aims to examine possible ways in which NNs could be useful to extend the application of the BPM. This will be further elaborated on in the discussion section below.

This section deals with the specifics of the research undertaken. The sample is described in detail, experimental research design explained and justified, and statistical techniques discussed.

5.2. Data Set

The dataset was acquired from AC Nielsen Homescan™ panel that comprises 15000 UK households representing the British population. Grocery purchases are recorded by participants that use hand-held barcode scanners to collect information on all their everyday purchases of fast-moving consumer goods. The product subset used in this research contains data on biscuits for the time segment of 52 consecutive weeks, 76,683 cases, 1847 individual consumers, and 14 variables (and is a part of a larger dataset employed for example by Oliveira-Castro et al., 2008). General demographic variables are included (age, social class, working status), as well as product specific information by date such as counts, weight, and quantities of product purchased, as well as the brand and type of biscuits, and the name of the supermarket.

5.3. Variables

The process of identifying variables that carry predictive power for the dependent variable studied with traditional methods could be tedious and time-consuming. Models need to be tested using changes in AIC, adjusted R^2 or some other model fitting criterion to identify whether adding or removing variables improves the model. Once you start looking into interactions, the choice of what to include and not include in the model increases exponentially. This is one of the drawbacks of traditionally employed methods, as they require a predetermined structure. NNs on the other hand do not involve any predetermined structure for the modeling process: the model is constructed using a complete dataset available and is able to identify through the process of model training the best predictive variables and interactions within the data capable of improving the model predictive capacity.

Some data manipulations were required to define a dependent variable for the analyses described in the following section. As previously discussed, choice is a probabilistic value in behaviorist terms, and needs to be defined as a proportion of instances of choosing one product over the other in a given time frame. In the context of the BPM and theoretical framework adopted for this research, *loyalty* could serve as a dependent variable in developing predictive models. Developed models would then be useful in offering an insight into what factors influence the product loyalty in a consumer choice situation and to what extent, and be able to predict changes to the product loyalty from the changes in the independent variable values. Models could also aid in developing a descriptive account of the consumer product loyalty phenomenon.

Informational and Utilitarian Reinforcement data has been acquired from previous studies on matching (for example Foxall et al., 2010) and integrated as separate variables within the dataset. As a result, two additional variables have been appended to the dataset on a transaction level to reflect the Informational and Utilitarian reinforcers each brand offers: each case in the dataset that represents a brand purchasing decision benefits from Utilitarian and Informational Reinforcement parameters to be used as independent variables.

5.4. Software

The statistical software used herein are SPSS 17.0 (SPSS-Inc., 2007), and R version 2.11.1 (R-Development-Core-Team, 2010).

6. Analysis and results

6.1. Statistical analysis

The dataset includes a number of variables that describe demographic characteristics of consumers, offer brand and product information, along with quantities purchased and money spent. The dataset provides information on a transaction level, meaning each transaction is recorded as a separate case in the dataset – contrary to customer level data where each case represents an individual consumer, and shows all transactions for that customer.

Initial data manipulations involved cleaning the dataset and transforming it purely superficially without any adjustments to the information it contains – to assure the data transfer between different software packages would not be an issue. After initial exploratory analyses, some of the cases have been removed from the dataset, and dataset has been amended to better suit the purpose of present research. Thus, only consumers with 7 or more transactions remain in the dataset used for all further analyses. As a result, the individual consumer number decreased to 1,594 and the dataset contains 75,563 cases. It may be advantageous to preserve the dataset in its original state as much as possible in other

situations, but those consumers who do not necessarily fit the established research criteria and would only create additional noise without offsetting this with sufficient contribution to the explanation of the phenomena at this stage were purged for all the consecutive analyses discussed in later sections (Oliveira-Castro et al., 2008).

For the dependent variable, some data transformations have been carried out. For each individual consumer, all purchases have been analyzed to determine consumers' preferred product type. This is defined as a product category that was purchased the most within a 52-week period, i.e., the highest amount of money spent to purchase the product of that category. Once that is established, it is possible to identify the consumer loyalty to that particular category, which provides a proportionate value between 0 and 1 and explains the probability of a consumer purchasing a product of a preferred category compared to all other products purchased within the 52 week period. To illustrate, the following example is offered:

Consumer A has a total spend of 377.23 within 52 weeks on all types of biscuits, where 186.83 are spent on BISC_CHOC_COUNTLINES. This then provides a loyalty value of $186.83 / 377.23 = 0.495268$, which means that nearly half the time consumer prefers chocolate countlines to any other type of biscuits.

As the study is mainly concerned with the classification problem here dealing with the binary output variable using logit and NN models, *loyalty* dependent variable could be recoded into binary using a median split. Not an optimal decision from the statistical point of view (for extended discussion see for example Aiken, West, & Reno, 1991), but nevertheless the split is necessary. This procedure is carried out on the consumer level (as opposed to transaction level), where each of the consumers is assigned into a high and low loyalty category using a median split, resulting in a different number of transactions in each loyalty category depending on the individual consumer purchasing frequency.

In addition, some of the variables such as brand names and store identifiers are recoded to reduce the excessively wide range and improve computational functionality (until this was performed, models would consistently crash as the processing limit of resources available was quickly reached).

6.2. Comparison between the Neural Network and the Logistic Regression

Logistic Regression was compared with a number of single hidden layer (3-layer) connectionist network models, where number of neurons in the hidden layer was varied from 1 to 100. Figure 6 shows Receiver Operating Characteristic (ROC) curves for a sequence of neural networks with 1, 2, 3, and so on up to 100 hidden nodes, demonstrating how the expansion of the network hidden layer yields smaller and smaller classification error. It is also clear that connectionist models show superior performance over the logistic regression model, with larger connectionist models greatly outperforming logistic regression results.

Results are consistent across multiple iterations of the test: 2-fold validation, entire procedure replicated 10 times, over 2000 models developed and assessed.

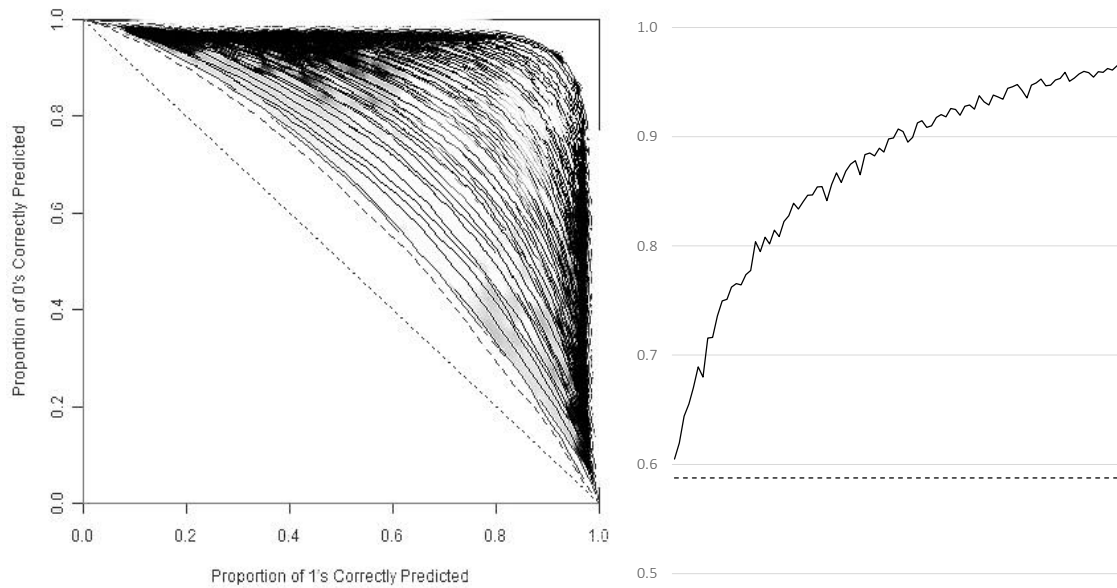


Figure 6. Left: ROC curves for a logistic regression model (dash line) and a sequence of neural networks (solid lines) with 1, 2, 3, and so on up to 100 hidden nodes in a single hidden layer. Right: progressive increase in the area under the curve for the same connectionist (solid line) and logistic regression (dash line).

6.3. Informational and Utilitarian reinforcers

As both reinforcers and the group information that each brand is assigned to, according to the Utilitarian and Informational Reinforcement levels, were included as independent variables and inputs into the models, it is possible to exclude them and examine the effects this exclusion would have on the overall model performance. To do so, NNs models developed in the previous stages are compared with the analogous connectionist models that do not contain Informational and Utilitarian Reinforcement variables. Employing the same testing procedure, connectionist models that include and exclude Informational and Utilitarian Reinforcement variables are compared. As a result, models that exclude utilitarian and informational reinforcement are smaller, but lack the predictive capacity that Informational and Utilitarian Reinforcement variables are likely to offer. Results are consistent across multiple iterations of the test: 2-fold validation, entire procedure replicated 10 times, over 2000 connectionist models developed and assessed.

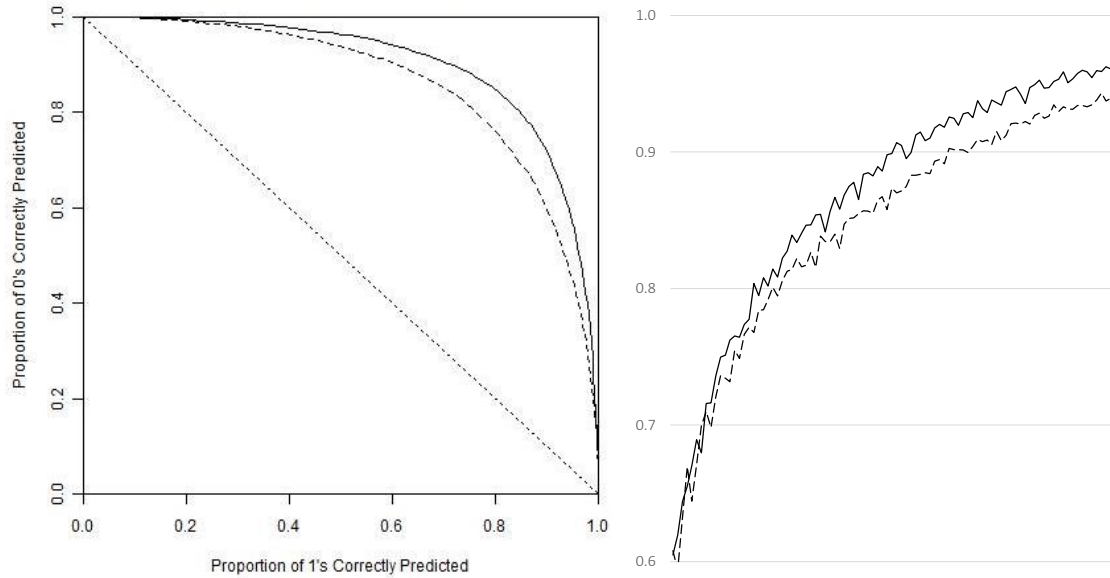


Figure 7. Left: Example shows contribution of Informational and Utilitarian Reinforcement variables using 2 connectionist models (46 hidden nodes) – one includes utilitarian and informational reinforcement (solid line) vs one that excludes them (dash line). Right: Progressive increase in the area under the curve for a sequence of neural networks with 1, 2, 3, and so on up to 100 hidden nodes in a single hidden layer for connectionist models that include (solid line) and exclude (dash line) utilitarian and informational reinforcement variables.

As a result, Figure 7 (right) shows progressive model performance (ROC area under the curve, y-axis) depending on the number of neurons (1-100, x-axis). Models compared are identical with the only difference being the inclusion (solid line) and exclusion (dash line) of the Informational and Utilitarian Reinforcement variables. The highest improvement that NNs models with the Informational and Utilitarian Reinforcement variables demonstrated was 0.040 (model with the 46 hidden nodes as shown on the left in Figure 7), with average improvement of 0.021 across all the models.

These findings demonstrate the usefulness of the Informational and Utilitarian Reinforcement variables as employed in *connectionist* models. Further tests would be required however to verify out of sample performance and model ability to predict new data.

7. Discussion

7.1. Model performance and nonlinearity analysis

NNs models performed considerably better than logit models once a hidden layer was incorporated into the models. The network structure that incorporates hidden nodes between inputs and outputs allows exploration of nonlinear relationships. It is clear from the results, that when consumer data and consumer behavior is the field of study, nonlinearity provides a substantial advantage over the linear models. Relatively weak performance levels of logit models shown could be due to a number of factors. It is possible that the data does not contain the variables vital for the prediction of consumer loyalty information, or that the variables that are readily available for marketing researchers (and therefore most frequently collected and analyzed) do not contain sufficient predictive power. Another possibility is that the relations of independent variables with the dependent variable that describes consumer loyalty are not linear. If this is indeed the case with the insufficiently predictive data, there is

not much that can be done. ‘Rubbish in – rubbish out’ as they say, meaning no good models could be built using bad data. If however, the problem lies in the nonlinearity, more appropriate methods of analysis would be able to extract the relations from the data. As results show, this is the case with the dataset used here, as NNs methods were able to extract a lot more information useful in prediction analyses. It is, of course easy to include powers for any numerical independent variables as well as two-term, three-term interactions, for logistic (or any other) regression models. However, with even a modest number of independent variables, this process rapidly becomes unwieldy, whereas the NN models with hidden layers and the logistic activation function automatically includes interactions as well as the nonlinear element. If the connection weights are such that their summed inputs are small and the neurons are operating around the near-linear middle section of the activation then the NN can even simulate an overall linear function as can a feedforward NN with skip-layer connections (Curry & Morgan, 2003).

The NNs model complexity tests showed some promising results as well. While working with the smaller dataset, sufficiently sophisticated NNs models are capable of learning the entire dataset with the appropriate training and therefore make perfect predictions. The dataset used here though is sufficiently large to avoid such issues, and allows the testing of networks that are particularly complex. Results obtained employing such test design provide information on the effects of model size on overall model performance. The performance of NNs models containing a number of hidden nodes that range from 1 to 100 (and several models that incorporate even higher numbers of hidden neurons as described above,) compared with the performance of logit models, show continuous improvement in the NNs model ability as the model size increases – and as a result, shows the ability of the model to account for the nonlinear relations within the data. From this, it should be safe to suggest that NNs models are particularly suitable for the analysis of consumer behavior data.

It is expected that at some point NNs models’ performance will flatten out as model size is considered in the comparative analysis and large models are penalized. The consumer behavior dataset used here however, may be large enough to allow for bigger NNs models to improve continuously, extracting even more information every time to increase model performance. The NNs model developed here contains a single hidden layer with a number of neurons that ranges from 1 to 100 to examine to what extent an increase in the nonlinear capacity of the model improves the predictive power. It is however, possible in addition, by increasing the number of neurons within a hidden layer, to also include multiple hidden layers, thus increasing model complexity even further by adjusting a number of neurons within each hidden layer. This is often an unnecessary step as the NNs models with multiple hidden layers are prone to overfitting, but with a sufficiently large dataset such as ours it could be worth exploring. The problem of overfitting and model generalizability is a potential area of future research, to build upon the findings offered here. To do so, the single hidden layer NNs model size needs to be increased by increasing the number of hidden neurons to the point where an increase in the model performance due to the model size is not sufficient to cover the size penalty applied to the model by the comparative mechanism. As a result, the optimal, single layer NNs model could be identified from the computational standpoint. It is then necessary to evaluate the performance of the models on out of sample data to assess their generalizability and their ability to make predictions using new data and identify the optimal model architecture using these criteria as well, consequently comparing such models with the computationally optimal model. Certain early stopping mechanisms could be examined as that could help avoid overfitting issues during the model training stages.

Depending on the primary model application, be that either predictive capacity or explanatory power, a number of strategies could be implemented to improve model performance. The following sections focus on NNs models predictive and explanatory abilities in greater detail.

7.2. Informational and Utilitarian Reinforcement

Results indicate that models indeed demonstrate improved performance with the Informational and Utilitarian Reinforcement as independent variables, corroborating some of the most recent findings (Yan et al., 2012). This however, needs to be further explored, as improved model performance needs to be evaluated relative to model size, as larger models are expected to be able to better fit the data.

It is worth mentioning, however, that it is expected that brand information will be modified during the analyses, as it was here. For example brand name variables have been transformed here to include only the top brands and not the complete set comprising thousands of different brand names, to alleviate the computational strain and pace the processing resources. During this data transformation it is clear that some information will be lost. Informational and Utilitarian Reinforcement variables however are numeric, and need not be transformed. As a result, they are able to preserve the original level of information predictive ability on the transaction level, which might be lost during the transformation of other non-numeric variables (i.e. brand name) to better fit the statistical analyses. These variables could also be obtained independently in addition to main marketing data sourced elsewhere, supplementing the dataset with a relatively low effort (through a survey of an organization's current customers, for example, or through focus groups).

7.3. Theoretical implications

Once convinced by the predictive capacity of NNs when used with consumer data and by the appropriateness of using nonlinear modeling techniques, it is important to discuss what this means in terms of theory.

NN modeling is based on the connectionist theoretical framework where simple computational units put together are capable of displaying high performance that is unattainable individually. The inherent mechanisms of activation within the connectionist models are rather simple. The fact that original NNs models were developed to imitate the processing capacity of the brain should suggest the connectionism would be particularly useful and appropriate to study human behavior. The work done in the field of cognitive psychology could provide supporting evidence for such claims.

While studying animal behavior and cognition, the connectionist framework has been shown to be particularly appropriate in explaining certain aspects of discrimination learning. Pearce (1994, 2002) developed a comprehensive model based on the work of Herrnstein (1970) that could be used to predict animal behavioral responses to outside stimuli in a quantitative manner. It was shown how such behavior could be elegantly explained through connectionism on a biological neurons and synapses level inside the brain. Many others also noted the appropriateness of NNs and connectionism to study human behavior (Curry & Moutinho, 1993).

Another field that embraces NNs models as powerful tools able not only of explanation, but also of new promising lines of research discovery, is psycholinguistics. Acquisition and development of language is extremely complex and follows a long learning process. NNs models are developed through the process that is also often referred to as the learning process. It is not surprising then that psycholinguists increasingly turn to NNs models to help explain language acquisition.

A somewhat different yet very successful application of NNs models could be observed in engineering. Here NNs-based models have been proven to be increasingly successful at computationally demanding tasks such as automated face recognition (Er, Wu, Lu, & Toh, 2002; Lawrence, Giles, Tsoi, & Back, 2002; Rowley, Baluja, & Kanade, 2002).

8. Conclusion

This paper has given an account of and the reasons for the widespread use of NNs models to study a wide range of phenomena with great success. It is argued here that connectionism shows excellent promise for explaining consumer behavior. The predictive abilities of NNs models in explaining consumer choice have been investigated, and the usefulness of the connectionist framework to the BPM discussed.

This study set out to determine whether NNs models could be useful in explaining consumer behavior following the established theoretical framework of the BPM. In the course of research, a large number of NNs models (2000 models, number of nodes within the hidden layer ranging from 0 to 100) of varying complexity have been developed and assessed. This was done by comparing the NNs models with the traditional methods of analysis such as logistic regression, and through a comparison of NNs models with each other in the test that examined the predictive power and contribution value of Informational and Utilitarian Reinforcement variables. Returning to the hypothesis posed at the beginning of this study, it is now possible to state that NN models showed a better performance than the traditional methods of statistical analysis (logit) did.

This study has shown that NNs models offer the capacity to help develop the understanding of consumer behavior in the future. These findings suggest that in general the complex nonlinear nature of consumer behavior data could be analyzed with the parallel connectionist models relatively successfully. One of the more significant findings to emerge from this study is that the logistic regression that may very well be the preferred method of analysis in marketing literature was greatly outperformed even by the simplest of NNs models. The second major finding was that the performance of more complex NNs models just kept improving as the additional neurons were added into the models. This is likely to be explained by the relatively large dataset employed here that allowed each successively more complex model to find more significant relations within the data that contributed to the explanation of consumer choice. The relevance of Utilitarian and Informational Reinforcement variables in predicting consumer behavior is clearly supported by the current findings. A number of models of varying complexity (2000 models, 1 to 100 hidden nodes) have been developed to examine the Utilitarian and Informational Reinforcement variable contribution, and results have shown that the models that included Reinforcement variables consistently produced better NNs models as compared with the NNs models that excluded the Informational and Utilitarian Reinforcement variables. Utilitarian and Informational Reinforcement emerged as reliable predictors of consumer choice.

The evidence from this study suggests that consumer data contains nonlinear relations between the variables normally considered by marketing researchers (demographics, product details, consumer situation information). NNs models therefore could be increasingly useful in working and modeling such data. The results of this research support the idea that the proven framework of BPM could be considerably extended with the application of connectionist constructs to help explain consumer behavior and consumer choice. The interdisciplinary nature of connectionism also complements the complexity of consumer behavior research that often draws upon different disciplines such as psychology, economics, marketing, and other to develop a complete account of consumer behavior. In general,

therefore, it seems that connectionist models are able to account for complexities within the data that linear models are unable to do. Not only is the predictive power that NN models are able to provide in many cases superior to traditionally employed statistical methods such as logistic regression, but also the explanatory power that NN models offer employing a number of algorithms greatly surpasses that of traditionally employed methods. Taken together, these results suggest that connectionism is one string of research that could be a logical continuation for the BPM framework promising good new findings in the future.

This research will serve as a base for future studies into the application of connectionist models to consumer behavior data. These findings enhance our understanding of the consumer situation and provide an alternative approach to examining the decision-making process that revolves around purchasing behavior. The current findings add to a growing body of literature on the application of NNs to the study of complex cognitive phenomena and the examination of nonlinear data that is subsequently able to provide the predictive ability of the future events with a convincing degree of accuracy. The methods used for this consumer data and product category may be applicable to other data and product categories as well, which would allow for an assessment of the generalizability that the models developed are able to offer. The present study confirms previous findings and contributes additional evidence that suggests that Utilitarian and Informational Reinforcement variables central to the BPM framework are increasingly useful in consumer behavior analysis. The empirical findings in this study provide a new understanding of the application of these variables in predicting and forecasting consumer choice.

Finally, a number of important limitations need to be considered. First, the models need to be tested using out of sample testing and k-fold cross validation to assess the model performance. Second, the current investigation was limited by the nature of the dataset. Even though the data employed here included a very large number of cases and many variables with a high number of individual consumers, it was limited to 52 weeks of purchasing behavior and a single product category. The limited time span prevents certain test designs such as where the data is split chronologically and models are trained on the first weeks (months, years) of the data and are subsequently tested on the last weeks of the data. Tests of such nature provide an obvious benefit of testing the model on the real data taking the experiment even further away from the laboratory into the real market situation. It is also not assumed at any point in the paper that these results should also be applicable to other product categories, which remain to be examined. Third, even though the current research was not specifically designed to evaluate the data with a continuous dependent variable, one source of weakness in this study, which could have affected the measurements of consumer loyalty, was that the probabilistic loyalty value had to be converted into binary. The nature of decision-making is rarely represented in a form of a choice between few alternatives or a binary type of a problem (such as belonging to one of the two groups) as examined here, but rather is a probabilistic measure. As essential predictive information is lost during the transformation of probabilistic value into binary, the tests designed to employ probabilistic variables should offer better results. This then also is a promising area of future research as NNs models could be compared to other traditionally employed methods such as multiple regression.

This research has thrown up a number of questions in need of further investigation, and a number of possible future studies using the same experimental set up are apparent. It is therefore recommended that further research be undertaken in the following areas: (1) The predictive and (2) explanatory capacity of the connectionist framework and NNs models working with the consumer behavior data and the process of consumer decision-making

needs to be further examined by employing the variable contribution algorithms and out of sampling testing to accurately assess model ability working with the new data. (3) The specific contribution to the BPM theoretical framework that connectionism could provide needs to be discussed. Further experimental investigations are needed to estimate the extent of NNs model capacity to predict new previously unseen data. Further research might investigate the best algorithms to maximize model predictive ability relative to its size, and the relative ease of integrating the application of these models in the marketing industry. Considerably more work will need to be done to determine the explanatory capacity of NNs models and their ability to explain the consumer decision-making process. The research here might take a more academic direction and involve significant interdisciplinary collaboration from such research fields as cognitive psychology, behaviorism, connectionism, economics, and marketing. Future research should also concentrate on the investigation into how and to what extent connectionism could be integrated with the established BPM theoretical framework, namely just regarding the computational modeling or regarding the connectionist representation of consumer behavior process on a cognitive level, should the suggested integration transpire.

References

- Adya, M., & Collopy, F. (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of forecasting*, 17(5-6), 481-495.
- Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*: Sage Publications, Inc.
- Anderson, P. F. (1986). On method in consumer research: a critical relativist perspective. *The Journal of Consumer Research*, 13(2), 155-173.
- Bashford, S. (2009). Bring me sunshine. *Marketing*, 32-33.
- Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental analysis of Behavior*, 22(1), 231-242.
- Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental analysis of Behavior*, 32(2), 269-281.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. New York: Oxford University Press.
- Calder, B. J., & Tybout, A. M. (1987). What Consumer Research Is. *Journal of Consumer Research*, 14(1), 136-140.
- Cornwell, B., Chi Cui, C., Mitchell, V., Schlegelmilch, B., Dzulkiflee, A., & Chan, J. (2005). A cross-cultural study of the role of religion in consumers' ethical positions. *International Marketing Review*, 22(5), 531-546.
- Cunningham, L. F., Young, C. E., Moonkyu, L., & Ulaga, W. (2006). Customer perceptions of service dimensions: cross-cultural analysis and perspective. *International Marketing Review*, 23(2), 192-210.
- Curry, B., & Morgan, P. H. (2003). Neural networks, linear functions and neglected non-linearity. *Computational Management Science*, 1(1), 15-29.
- Curry, B., & Moutinho, L. (1993). Neural networks in marketing: Modelling consumer responses to advertising stimuli. *European Journal of Marketing*, 27(7), 5-20.
- Davies, F. M., Goode, M. M. H., Moutinho, L. A., & Ogbonna, E. (2001). Critical factors in consumer supermarket shopping behaviour: A neural network approach. *Journal of Consumer Behaviour*, 1(1), 35.
- Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Boston: Allyn & Bacon.
- Er, M., Wu, S., Lu, J., & Toh, H. (2002). Face recognition with radial basis function (RBF) neural networks. *Neural Networks, IEEE Transactions on*, 13(3), 697-710.
- Foxall, G. R. (1980). Marketing models of buyer behaviour: A critical view. *European Research*, 8(5), 195.
- Foxall, G. R. (1984). *Corporate Innovation: Marketing and Strategy*. London: Croom HELM Ltd.
- Foxall, G. R. (1990/2004). *Consumer psychology in behavioral perspective*. London and New York: ROutldge. Reprinted 2004, Freerick, MD: Beard Books.
- Foxall, G. R. (1999). The substitutability of brands. *Managerial and Decision Economics*, 20(5), 241-257.
- Foxall, G. R. (2003). The behavior analysis of consumer choice: An introduction to the special issue. *Journal of Economic Psychology*, 24(5), 581-588.
- Foxall, G. R. (2009). *Interpreting consumer choice: The behavioural perspective model*: Routledge.
- Foxall, G. R. (2010). *Interpreting consumer choice*. New York: Routledge.
- Foxall, G. R. (2014). *Corporate Innovation (RLE Marketing): Marketing and Strategy*: Routledge.

- Foxall, G. R., & James, V. K. (2001). The behavioral basis of consumer choice: A preliminary analysis. *European Journal of Behavior Analysis*, 2, 209-220.
- Foxall, G. R., & James, V. K. (2003). The behavioral ecology of brand choice: How and what do consumers maximize? *Psychology and marketing*, 20(9), 811-836.
- Foxall, G. R., Oliveira-Castro, J. M., James, V. K., Yani-de Soriano, M., & Sigurdsson, V. (2006). Consumer behavior analysis and social marketing: the case of environmental conservation. *Behavior and social issues*, 15(1), 101-124.
- Foxall, G. R., Oliveira-Castro, J. M., & Schrezenmaier, T. C. (2004). The behavioral economics of consumer brand choice: Patterns of reinforcement and utility maximization. *Behavioural Processes*, 66(3), 235-260.
- Foxall, G. R., & Schrezenmaier, T. C. (2003). The behavioral economics of consumer brand choice: Establishing a methodology. *Journal of Economic Psychology*, 24(5), 675-695.
- Foxall, G. R., Wells, V. K., Chang, S. W., & Oliveira-Castro, J. M. (2010). Substitutability and independence: matching analyses of brands and products. *Journal of Organizational Behavior Management*, 30(2), 16.
- Foxall, G. R., Yan, J., Oliveira-Castro, J. M., & Wells, V. K. (2011). Brand-related and situational influences on demand elasticity. *Journal of Business Research*.
- Güneren, E., & Öztüren, A. (2008). Influence of Ethnocentric Tendency of Consumers on Their Purchase Intentions in North Cyprus. *Journal of Euromarketing*, 17(3/4), 219-231.
- Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Oxford: Maxwell Macmillan International.
- Hebb, D. O. (1949). *The organisation of behaviour*: New York: Wiley.
- Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental analysis of Behavior*, 4(3), 267.
- Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental analysis of Behavior*, 13, 243-266.
- Herrnstein, R. J. (1997). The matching law. In H. Rachlin & D. Laibson (Eds.), *Papers in Psychology and Economics*. New York: Sage.
- Herrnstein, R. J., Rachlin, H., & Laibson, D. I. (1997). *The matching law: Papers in psychology and economics*. Cambridge, Mass: Harvard University Press.
- Holbrook, M. B. (1987). What Is Consumer Research? *Journal of Consumer Research*, 14(1), 128-132.
- Kagel, J. H., Battalio, R. C., & Green, L. (1995). *Economic choice theory: An experimental analysis of animal behavior*: Cambridge Univ Pr.
- Lawrence, S., Giles, C., Tsoi, A., & Back, A. (2002). Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on*, 8(1), 98-113.
- Lu Hsu, J., & Han-Peng, N. (2008). Who are ethnocentric? Examining consumer ethnocentrism in Chinese societies. *Journal of Consumer Behaviour*, 7(6), 436-447.
- McKee, A. (1984). Social economy and the theory of consumer behavior. *International Journal of Social Economics*, 11(3/4), 45.
- Moore, R. S., Beauchamp, M. B., Barnes, C. Z., & Stammerjohan, C. A. (2007). Artificial neural networks : an introduction and application in a consumer behavior setting. *Marketing Management Journal*, 17(1), 27-38.
- Oliveira-Castro, J. M., Diogo, F., Foxall, G. R., & Schrezenmaier, T. C. (2005). Dynamics of repeat buying for packaged food products. *Journal of Marketing Management*, 21, 1(2), 37-61.

- Oliveira-Castro, J. M., Foxall, G., & James, V. (2008). Individual differences in price responsiveness within and across food brands. *The Service Industries Journal*, 28(6), 733-753.
- Oliveira-Castro, J. M., Foxall, G. R., & Schrezenmaier, T. C. (2005). Patterns of consumer response to retail price differentials. *Service industries journal*, 25(3), 309-335.
- Oliveira-Castro, J. M., Foxall, G. R., & Wells, V. K. (2010). Consumer brand choice: money allocation as a function of brand reinforcing attributes. *Journal of Organizational Behavior Management*, 30(2), 15.
- Oliveira-Castro, J. M., Foxall, G. R., Yan, J., & Wells, V. K. (2011). A behavioral-economic analysis of the essential value of brands. *Behavioural Processes*.
- Oliver, R. L. (1999). Whence consumer loyalty? *The Journal of Marketing*, 33-44.
- Pachauri, M. (2002). Consumer behavior a literature review. *Marketing Review*, 2(3), 319.
- Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, 101(4), 587.
- Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior*, 30(2), 73.
- R-Development-Core-Team. (2010). R: A language and environment for statistical computing. Vienna, Austria.
- Rachlin, H., Kagel, J. H., & Battalio, R. C. (1980). Substitutability in time allocation. *Psychological Review*, 87(4), 355.
- Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge ; New York: Cambridge University Press.
- Rowley, H., Baluja, S., & Kanade, T. (2002). Neural network-based face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(1), 23-38.
- Rumelhart, D. E., & McClelland, J. L. (1987). Parallel distributed processing, exploitation in the microstructure of cognition-Vol. 1: Foundations. *Computational Models of Cognition and Perception, Cambridge: MIT Press, 1987, 1*.
- Sent, E. M. (2004). Behavioral economics: how psychology made its (limited) way back into economics. *History of Political Economy*, 36(4), 735.
- Simon, H. A. (1982). Models of bounded rationality.
- Simon, H. A. (1987). Behavioural economics. *The new Palgrave: A dictionary of economics*, 1, 221-225.
- Smolensky, P. (1995). On the proper treatment of connectionism. *Connectionism: Debates on psychological explanation*, 2, 28-89.
- SPSS-Inc. (2007). *SPSS Statistics Base 17.0 User's Guide*. Chicago, IL.
- van Kenhove, P., Vermeir, I., & Verniers, S. (2001). An Empirical Investigation of the Relationship between Ethical Beliefs, Ethical Ideology, Political Preference and Need for Closure. *Journal of Business Ethics*, 32(4), 347-361.
- Van Wezel, M. C., & Baets, W. R. J. (1995). Predicting market responses with a neural network: the case of fast moving consumer goods. *Marketing Intelligence & Planning*, 13(7), 23-30.
- Watson, J. J., & Wright, K. (2000). Consumer ethnocentrism and attitudes toward domestic and foreign products. *European Journal of Marketing*, 34(9/10), 1149.
- Wells, V. K., Chang, S. W., Oliveira-Castro, J., & Pallister, J. (2010). Market segmentation from a behavioral perspective. *Journal of Organizational Behavior Management*, 30(2), 176-198.
- Wells, V. K., & Foxall, G. R. (2011). Special issue: Consumer behaviour analysis and services. *The Service Industries Journal*, 31(15), 2507-2513.

- West, P. M., Brockett, P. L., & Golden, L. L. (1997). A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, 16(4), 370.
- Yan, J., Foxall, G. R., & Doyle, J. R. (2012). Patterns of reinforcement and the essential values of brands: i. incorporation of utilitarian and informational reinforcement into the estimation of demand. *The Psychological Record*, 62(3), 361.