The Evolution of Retailing: A Meta Review of the Literature

Abstract

The evolution of retailing has interested academics across a range of disciplines including economics, history, geography, and marketing. Due to its interdisciplinary appeal, the corpus of knowledge on retailing is composed of many disparate variables of analysis - from transaction costs, and entrepreneurs, to environmental factors, and the dispersion of stores. In consequence, the literature that attempts to explain retailing evolution presents as a patchwork, and extant theories remain disconnected because of their narrowness of focus. This literature review applies a macro and systems theory approach to the multi-discipline literature, and links together bodies of work that, until now, have remained conceptually unconnected. This provides a meta typology of six factors that could explain change in retailing: economic efficiencies; cyclical patterns; power inequities; innovative behaviour; environmental influences; and interdependent parts of the system in co-evolution.

Keywords: retailing institutions; literature review; history of marketing thought; marketing history, systems theory.
**Introduction**

In the long input-output chain involving the production and distribution of consumer goods to persons and households, retailers are the terminal members in the value chain. Retail stores provide the arena in which the final economic transaction takes place, and consumption begins. That is, retailing represents the inflection point at which costs are converted into revenues, and the exchange behaviour that underpins marketing occurs, and its development has been of continuing interest to social science researchers. Yet the impact of retailing on marketing has been largely ignored, setting aside the work of a few notable scholars, particularly Stanley Hollander.

The neglect of the role of retailers reflects a wider neglect of the study of marketing channel formats in general (Quinn and Murray 2005; Williams 2009). Classical economic theory dominated business thinking for at least the first half of the twentieth century and its assumption of “costless” transactions meant that “until recently many economists produced institution-free models and were proud of the irrelevance of institutions for their theories…” (Economics) assumed that institutions were a ‘free factor’ or, in other words, that organizing economic activity was not costly” (Landesmann and Pagano 1994, p. 199). The services of retailers are not, of course, cost-free. Retailers bear the risks of forward purchase, later payment, and uncertain demand. As described by the functionalist writers, retailers solve problems of coordination in providing the canonical utilities of time, place, and product assortment which, for consumers (the ultimate purchasers of goods and services), means the ability to buy products when, where and in whatever combination we wish - without recourse to the original producers.

Despite the importance of retailers in marketing, the growth of knowledge on retailing change has been slow - proceeding in discrete bits - with the exceptions of the writings of
Chandler (1977), and Hollander (for example, 1960, 1964, 1967, 1972-1973, 1978, 1980, 1981, 1986a, 1986b), and significant contributions from Dixon (1963, 1994, 2001), Fullerton (1986), Savitt (1989), and Tedlow (1990). The slow pace of knowledge development is explained in large part by the ahistorical nature of most research in marketing. As Dixon (1982) argues, in denying history, and therefore our intellectual heritage, marketing scholars have tended to redevelop - rather than develop further - concepts that already exist. The main concepts that constituted a channel of distribution, for example, had been present in the literature for centuries, but it was not until the nineteenth century that Shaw first listed them in a “slow and painstaking” working out of ideas (Dixon 1982).

Since the arrival of department stores in the mid 1800s, retailing has undergone significant change in five main areas: formats, technologies, forms of ownership, customer types, and geography (Chandler 1977). In Western countries, the most visible change has been the transition of stores through a sequence of different formats. In approximate chronological order of their development, the main retailing formats have been small general stores, department stores, chain stores, supermarkets, discount department stores (such as Walmart), franchises, category killers or superstores, and e-retailing. This sequence is imprecise, however, because precursors for each format type existed prior to their mainstream arrival, and as other researchers argue, the pace of change is a process in which at any one time, “there are likely to be participants at a variety of stages of the cycle” (Dawson 1979, p. 351). Different country settings produced different dates of emergence, and there is a difference between individual and global chronologies.

While several theories have been put forward to explain this progression, the development of a global model has proved “elusive” (Filser and McLaughlin 1989). Each of the
extant approaches has made valuable contributions to our understanding, but all are limited by their narrowness of view (Brown 1987, Levy et al. 2005). Explaining the reasons for the transformation of retailing formats therefore remains as “one of the most intriguing and difficult problems in the marketing discipline” (Markin and Duncan 1981).

The most commonly accepted view on retailing evolution is the “synthesis” by Brown (1987), which distils the literature into three main approaches to understanding the transformative process: cyclical, conflict, and environmental. In other words, changes in retail formats are the outcome of some combination of environmental forces, and a cyclical sequence of institutional conflicts (Anitsal and Anitsal 2011). This three-variables explanation of retailing evolution has endured for decades (e.g., Anitsal and Anitsal (2011), Betancourt and Gautschi (1986), Davies (1998), Evans (2011), Evans, Barnes, and Schlacter (1993), Gist (1968), Levy et al. (2005), Markin and Duncan (1981), Roth and Klein (1993), and

In marketing’s most recent attempt to explain the evolution of retail formats, the “Big Middle”, Levy et al. (2005) combine economic with cyclical factors in arguing that the most successful retailers are - in the long-run - “inexorably drawn” into being volume-driven generalists. Brown et al. (2005) build on this in their recognition of the influence of technology and population migration, yet acknowledge the need for further research on environmental factors. Although the “Big Middle” advances our understanding of retailing change, it is deterministic and largely based on large-scale retailing experience in the USA. While each of the schools of thought on retailing provides partial explanations of its development, the types of independent variables that have been studied make interdisciplinary comparisons difficult. As shown in the summary of literature in Table 1, the range of independent variables in extant knowledge extends across disciplines, from the economic costs of exchange, environmental
influences, individual stores, and their geographic dispersion, to power in relationships in the
distribution channel (Pfeffer and Salancik 1978). Given the difficulty of comparing the
incommensurate variables of the literature, one way of analysing such cross-disciplinary
knowledge is not in terms of final or “ultimate causes” (Schumpeter 1947), but on the basis of
initial governing “first causes”, or fundamental similarities in the assumptions and beliefs about
why change takes place. In other words, researcher’s choices of topics and methods indicate
implicit beliefs about relevance, and therefore also about probable regularities (Hollander 1978).

This review therefore pulls back from a detailed examination of individual schools of
thought, to look at shared primary assumptions that can be discerned from the literature. It takes
a systems theory approach to the literature, in its “search for fundamental quantities and a listing
of parts” (Halliday 1984, p. 135). This is the approach recommended by Boulding (1956, p. 203)
“to look over the empirical universe and to pick out certain general phenomena which are found
in many different disciplines, and to seek to build up general theoretical models relevant to these
phenomena”. Indeed, the different schools of study are not incompatible in an epistemological
sense, and there are many points of crossover between bodies of work.

This meta approach to analysing the literature necessarily involves reductive
generalisations in order to derive the constituent elements of different theories (Brown, 2011).
This reductive process provides a means of making sense of disparate phenomena arising from
different disciplines, and links together bodies of work that, up until now, have remained
conceptually unconnected. This process also necessitates an examination of the historical
development of different bodies of thought, since assessing the validity of existing theories is
difficult unless they are considered within the historical context in which they occurred (Jones
and Shaw 2006). It is only by studying the unfolding of intellectual history, through tracing the
development of changing areas of interest captured in “old” research, that context-dependence and the accumulative nature of knowledge become apparent.

This review therefore analyses the wider themes of the cross-discipline literature based on assumptions made about fundamental “first” causes of change. It is not intended as a comprehensive account of the literature, but rather presents a birds-eye view on a landscape of knowledge and its most salient features. This provides a systems theory typology of underlying causes of retailing change. This typology does not in any way invalidate the contributions of the numerous and discrete explanations of exchange behaviour. Rather, it provides a framework in which we can seek to acknowledge - and support or refute - the litany of ideas contained in extant knowledge, together with their underlying causal themes. Table 1 summarises, in approximate chronological order of their emergence and development, six broad themes in the English-language literature on retailing change, where change is assumed, in the first instance, to take place as a function of:

1. Economic efficiencies
2. Patterns operating in nature
3. Power inequities
4. Innovative behaviour
5. Environmental influences, and
6. Interdependent parts of the system in co-evolution

In the rest of this document, we critically discuss these six themes. These themes are arrayed in their approximate chronological order of emergence to highlight how later themes were partially informed by the preceding perspectives.

Insert Table 1 about here
Change as A Function of Economic Efficiencies

Economic approaches dominated management thinking about distribution for at least the first half of the twentieth century, reflecting the push for “scientific” management advocated by Taylor (1947). Early writers included Coase (1937) and Friedman (1953), whose influence reflected and prolonged the hegemony of economic explanations of exchange behaviour, although the role of entrepreneurial activities of the managerial team is also acknowledged in some early economic work (e.g. Nystrom 1930).

Economic approaches focus on both micro and macro structural changes in the distribution industry and their central concerns include costs, productivity and concentration (Bucklin 1981, 1983; Mallen 1973; Takeuchi and Bucklin 1977; Tucker 1978). Early work by Barger (1955) and Cox et al. (1965) provided some insight into how changes in retail structure influenced retail cost and gross margins. Generally, economic approaches are summed up by Mallen (1973), who saw total distribution cost efficiency as the reason intermediaries emerge or disappear - the relevant cut-off points when economies versus diseconomies come into play. Under this logic, “the basic raison d’être of middlemen is providing external economies to producer firms” (Mallen 1973, p. 20).

Related to this are studies on horizontal concentration in the industry or changes in the number or types of intermediaries, such as Tucker (1978), Hall and Knapp’s (1959) survey on the relation between the number of shops and retailing productivity, or Jefferys’ (2011) study of retailing in Britain from 1850-1950. The impetus for studying horizontal concentration stemmed from the belief that there were “too many stores” (Ford 1935), an issue which “has been vexing retailers at least since Adam Smith”. A lack of data, however, meant industry “concern about over-trading tended to result from value judgements - from the shopkeepers’ sense that
competition had passed the limits of fairness” (Monod 1996, p. 26-7).

Transaction cost analysis (TCA), extended by Williamson (1975), focussed on how the role of transaction costs shaped the structure of market relations. Although its original focal unit was exchange dyads, TCA has been extended in more recent work focussing on the choice of using local intermediaries versus the establishment of a company’s own market entry channel, in other words “make or buy” decisions (Anderson and Coughlan 1987).

Another stream of economic explanation is that of the functionalist writers, whose research concerned the specific utilities of intermediaries, such as Alderson’s (1957) time, place, and product assortment utilities, or Bucklin’s (1967) list of channel functions: transit, inventory, search (communication), persuasion, and production. A definitive list of the functions of retailers is yet to emerge (Nyberg 1998), though standard textbooks suggest anticipating customer wants, developing assortments of products, storing products, sharing risk, acquiring market information, and financing and transport. Despite their mechanical nature, functionalist approaches provide a means for conceptualising channel structure in terms of the patterns of relationships among organisations and the way the work is divided, and the cost efficiency of performing each activity (Dixon and Wilkinson 1986).

Economic approaches rely heavily on ceteris paribus assumptions and presume that equilibrium solutions will prevail. Finally, economic approaches do not explain why “uneconomic channels” and cost-inefficient retailers persist over time (Filser and McLaughlin 1989), or why retail and wholesaling formats resist change even when it appears to offer economic benefits.

The unifying characteristic of works in this category is their logical or rational nature. Their governing assumptions, however, concern the justification of intermediaries, which posits
economic efficiency as a first cause of change.

**Patterns Operating in Nature**

The second main explanatory theme in retailing change focuses on patterns operating in nature. Common to these models is a belief in the presence of repeating phenomena, through either time and/or space. There are three different emphases in this group: cyclical models, conflict models and spatial positioning.

Cyclical models propose that retail formats pass through several predictable stages, before an inevitable decline. Cyclical models include McNair’s wheel of retailing (Brown, et al. 2005), the product or institutional life cycle (Davidson, Bates, and Bass 1976), and the observations of a general decline during the third generation of family-run businesses (Dreesmann 1968), prompting the maxim: “from shirtsleeves to shirtsleeves in three generations” (Lazonick 1994). The wheel of retailing - which predicts a pattern of growth, development, maturity, and decline - has widespread acceptance in marketing instruction, and in some standard textbooks, it is the only explanation proffered for retailing change. Although its simplicity is appealing, the wheel of retailing has limited power in explaining why department (and other) stores that should (according to its own logic), have been superseded, continue to trade today. Rather than being a theory of change, the wheel of retailing is perhaps “simply an artifact of a particular industrial evolutionary stage” (Hollander 1981, p. 89) and does not adequately inform why format innovations are adopted or disseminated at a sectoral level. Overall, this theoretical approach presents a questionable explanation of change and has been criticised for lacking universality, particularly in developing world contexts (Brown 1991; Savitt 1982).

Other cyclical models predict a pendulum-like movement between two ends of a spectrum (Doody 1963; Savitt 1989). They include the retail accordion, or the general-specific-
general approach (Brand 1967). This model describes the evolution of distribution formats in a polar pattern of the number of lines of merchandise handled by the predominant format. An illustration is the general store of the nineteenth century, which “gave way to the specialist, single-line businesses of the twentieth, which were themselves superseded by the mass merchandisers of the post-war era” (Brown 1987, p. 9). The mass merchandiser was followed in the 1950s, by the advent of business format franchising, when highly specific outlets again developed. By the end of the twentieth century, the giant superstore had re-emerged, for example the U.K. 40,000 square feet store offering dry-cleaning, photo-processing, health checks, a restaurant, crèche, pharmacy, and cut-price petrol (Jones 1998). The polarity of the retailing sector in terms of small and large entities has been an ongoing theme in the literature (Palamountain 1955), and more recently, work has been undertaken on the polarisation of retailing on such variables as service and size (Brown 1987; Monod 1996). Although cyclical in their rhythmic nature, these polar models can be differentiated from cyclical models such as the wheel of retailing because they lack the terminality of eventual decline. Instead, they allow for ongoing oscillation in a given plane of interest, and as such, they move towards a more general theory of change.

A more analytical class of works based on polar patterns includes conflict models or dialectic theory. Dialectic theories hold that innovation is brought about when newcomers challenge, and force change in traditional formats. Change takes place through a process of

\[
\text{Thesis} \rightarrow \text{Antithesis} \rightarrow \text{Synthesis} \rightarrow \text{Thesis} \rightarrow
\]

where the established or maturing class of format forms the thesis, and new, innovative formats the antithesis (Gist 1968). Both types adapt to produce a synthesis or composite institutional form, which becomes, in turn, the new thesis, and the pattern continues. While this seesaw of
change and counter change exhibits polarity, dialectic theory diverges from other cyclical models in its perception of the source of change, the innovator (the firm or its founder), and also because the nature of change is viewed as a two-way process. Dialectic or conflict models have a good general fit with historical evidence, and can be applied to a variety of cases of change. Examples of the phenomenon of synthesis are readily observable. For example, department stores were challenged by the advent of large discount houses that began to appear in the 1950s (with some exceptions, e.g., Five and Dime stores were introduced prior to World War 1 in some countries). By the 1970s, in most Western nations, a new synthesised format, discount department stores, such as Walmart, had appeared on the landscape, which Lichtenstein (2009) attributed largely to emerging technologies (e.g., barcodes) and changes in government legislation. Similarly, the growth of voluntary group trading in the grocery sectors in the 1960s and 1970s throughout the USA, UK and continental Europe was largely in response to the emergence of supermarket multiples and direct manufacturer-retailer linkages (Quinn and Leavy 2005). Early conflict or dialectic models grew out of the work of Nieschlag (Fullerton 1986), and Schumpeter (Bliss 1960). Schumpeter argued that discontinuous change (“change that matters”) happens when innovators force the existing body of retailers to adapt and imitate. Nieschlag, on the other hand, believed that the innovators themselves are modified in the process of diffusion of the innovation (Fullerton 1986). The unifying characteristic of all the literatures above is their teleological nature, in their shared belief that “there is a grand perpetuity” (Hakansson and Lundgren 1997, p. 19). This gives patterns operating in nature as a fundamental or “first” cause of change. However, in isolation, such theories provide only limited insight into retail evolution, and conflict models have been criticised for not fully explaining retail format adaptation. While the emergence of discount stores in the United States forced some retailers to reposition towards up-
market stores, other retailers responded by cutting prices or buying out competitors to maintain their market position (Brown 1987).

Less metaphysical but also concerned with a search for patterns, are the studies on retail location, which has been explored by researchers from diverse disciplines, although geographers have made the main contributions (Clarke, Bennison, and Pal 1997; Cohen 1996; Graff 1998). The central focus of these works is on the patterns of spatial distribution of shops, the associated patterns of shoppers’ behaviour and store choice (Craig, Ghosh, and McLafferty 1984), and the implications of these patterns for managerial or urban planners. Ingene and Lusch (1980), for example, examined geographic market selection for department stores using data from more than 200 United States metropolitan areas in an attempt to predict the potential profitability of specific markets. More recent scholars such as Coles (1996), in his geo-historical study of German retailing, found “distinct patterns and processes of retail evolution existed”, while Sparks (1990) concluded that “spatial and structural change in retailing…clearly… operate in harness”.

Allaway et al. (1994) found that suburban diffusions of innovations “emerge at both predictable and random locations” which can be influenced by the “purposeful efforts” of the innovator. In their application in site selection and normative purpose, the works concerning geographical patterns differ from others in this category. More recent research adopts an ecological perspective in understanding the spatial fragmentation in retail centres (Rotem-Mindali 2012). In common with other pattern research, however, is their understanding of evolution as a fundamentally unstable process of continual flux. This contrasts with the assumptions of economic theory of “ceteris paribus” and equilibrium in markets.

**Power Inequities**

The literatures in this category assume that, contrary to economic assumptions, not all behaviour
is optimal or logical. Rather humans typically act - in the first instance - in their own best interests, which may or may not coincide with rational expectations (“bounded rationality”). The unifying assumptions of this group of works are that “the *basic* variables are social variables” (Stern and Brown 1969, p. 13, italics in original), and that “economic relations are embedded in social ones” (Nelson 1995, p. 76).

The growth of interest in behavioural analysis corresponded with a rising discontent with the dominant economic paradigm (Nyberg 1998). The recognition that manufacturers had lost in prominence because of the rise of large-scale retailers, was being discussed in academic journals in the early 1940s (Wingate 1941), and retailers were coming to be seen “as whole individuals not simply economic agents” (Dixon 2002, p. 741). By the 1950s, the idea of the channel as a mechanical conduit was giving way to the channel as a set of relationships, or a “behaviour system” that changes over time (Alderson 1957).

Chandler (1977), Chernatony (1989), and Mallen (1967) described how the role of “channel captain” shifted over the past two centuries: from the wholesaler in the last half of the nineteenth century, to the manufacturer because of the Industrial Revolution, and then finally, to the retailer in the twentieth century. Central to this recognition of the rise of retailers was the construct of power. The research on power has been “truly inter-disciplinary” stemming from political theory and philosophy to become “one of the central concepts of the social and human sciences *per se*” (Clegg 1989, p. xvii). In marketing, the power literature deals with questions of evolution only indirectly. That is, it does not seek to explain evolution, but its substance highlights how distribution channels are shaped by power relationships in the channel, and has led to research into such behavioural issues as collaborative efforts within the channel and, more recently, the development of trust (for example, Morgan and Hunt 1994, Palmatier, Dant and
Grewal 2007, Young and Albaum 2003). Closely related to the research on power was the construct of conflict. Palamountain (1955) and Mallen (1967) identified the relationships of retailers with other retailers (“intertype competition”) as the major source of distributive conflict in channel relationships, and every innovation in retailing has met with resistance, from the anti-department store, and anti-chain store movements (Hollander 1972-73, p. 13), and resistance to group trading in numerous wholesale sectors (Quinn and Leavy 2005), to such incremental changes as the introduction of chromium turnstiles in stores (Kingston 1994).

Although there are difficulties in measuring the constructs of power and conflict, the inclusion of power in an analysis is of particular importance. Klein (1980) claims that economists have swept considerations of power under the “ceteris paribus rug” because it does not fit their paradigm. Instead, they pay heed only to the power differential between firms, and the typical unit of analysis is usually confined to the manufacturer:retailer dyad (for example, Chernatony 1989; Krishnan and Soni 1997; Segal-Horn and McGee 1989; Tynan and Cunningham 1995).

The theory of countervailing power directly addresses the issue of change in its more general proposition that, the aggregation of power in any part of a system often results in a corresponding but opposing force in another part of the system (Emerson 1962; Galbraith 1967; Gaski and Nevin 1985; Polanyi 1975). Galbraith (1967, p. 120) wrote: “power on one side of a market creates both the need for, and the prospect of reward for the exercise of countervailing power from the other side. This means that, as a common rule, we can rely on countervailing power to appear as a curb on economic power”. He predicted in this same 1967 work, the rise of retailers’ power as a countervailing force against the dominance of manufacturers at that time. History thus supports Galbraith’s thesis of countervailing power, since retailers have indeed
taken the dominant position in the channel, driven in large part by the imposition on manufacturers of trade allowances and “slotting fees” to rent shelf space (Lariviere and Padmanabhan 1997).

Countervailing power makes explicit the idea of imbalance in relationships as a primary driver of change. As Segal-Horn and McGee (1989, p. 26) argue: “much of the shift in balance of power towards retailers has itself been a response to the previous market power of suppliers”. This brings into focus the idea of relative power and improves on the shortcomings of economic approaches with their reliance on horizontal or industry-wide competition as the chief regulating mechanism. In countervailing theory, competition extends both horizontally and vertically or, as Galbraith predicted (1967, p. 119) “big sellers (manufacturers) cause the rise of big buyers” (retailers).

If perceived power inequities are a force of change in retailing, then we should expect to see a backlash against retailers’ current domination of the distribution channel, which has been evident since at least the 1980s. Since then, however, there have been very few manifestations of countervailing power against retailers, unless we consider the proposition that the rise of electronic retailing is being driven, at least in part, by the need to curb the power of store-based retailers. For manufacturers and suppliers, the Internet presents a means of bypassing retailers that is, for the first time in the history of direct channels, a viable alternative offering relative and absolute cost-efficiencies, plus access to a market that spans the globe. Similarly, consumers boost their relative power whether consciously or not, through the adoption of electronic commerce, because they multiply the number of alternative suppliers available, as the cost of “comparison shopping” across the globe approaches zero (Bakos 1991). One could conceivably argue from the lens of big middle concept - in which successful retailers are “inexorably drawn”
into being volume-driven generalists - that electronic retailing is the emergent big middle of the early twenty-first century, as evidenced in such online giants as Amazon and Alibaba.

Countervailing power, in common with the polar models discussed in the last section, describes a pendulum-like pattern, which views change as a two-step process of action and reaction. Its underpinning, however, is overtly behavioural in that countervailing power is seen as a solvent to human tensions, and in some way necessary as a corrective balance or “curing” power (Boulding 1989). The theory of countervailing power can be applied to a variety of cases of change, and generally fits well with historical evidence. Wilkinson (1990) points out, however, that its predictive power is limited in scope. It does not explain the reasons behind the success or failure of specific retailing formats. Related to this conceptual difficulty is the question of whether countervailing power is the end product, or merely a stage in a continual process of change (Hollander 1981, p. 89). That is, if we accept the morphing to form that is described by dialectic theories, then “original” power is more likely a countervailing response to some even earlier accumulation of power. In essence, countervailing power is concerned with relative power in a two-step process of change, but dialectic theory allows for ongoing change, and so moves towards a metatheory of change.

The contributions of power literature include its recognition of the role of humans as active agents in shaping exchange behaviour and, therefore, in effecting change over time. The theory of countervailing power acknowledges change more directly, and makes explicit the concept of change as a two-way process. The significance of both bodies of work is in their acknowledgement that power exists in the channel of distribution, and in the shared primary assumption that change is human-centred. However, as noted earlier it does not explain the reasons underlying success and failure of specific retailing formats.
**Innovative Behaviour**

This literature is the least developed of the six causal streams, even though the role of the entrepreneur was recognised at least as early as Adam Smith. The focus of innovation literature, however, centres not so much on conflict and self-interest, which prescribe the power literature, but on the particular role of innovators, or innovative behaviour in general.

Works in this category include the hagiographies of great stores and their founders. Americans celebrated department store founders John Wanamaker (Sobel 1974), Marshall Field (Wendt and Kogan 1952), and Roland H. Macy (Hower 1943), while the British lauded Gordon Selfridge (Honeycombe 1984), and the founders of Harrods (Callery 1991) and Liberty’s (Adburgham 1975). These celebratory works provide rich descriptions of the “personalities” behind department store retailing (Converse 1959), men who were by definition excellent promoters. Often referred to as “merchant princes”, these department store founders prompted the remark by economist Alfred Marshall that “in the hands of an exceptional genius (department stores) sometimes rise with meteor-like rapidity and splendour” (Perkins and Meredith 1996, p. 7). A parallel case can be made for Sam Walton, the founder of Wal-Mart.

The more analytical class of literature on the role of innovators stems from the eighteenth century writings of Adam Smith. Early works adopted a functionalist approach to entrepreneurs, similar to the functionalist works written on retailing, which recognised the value of “entrepreneur-coordinators” in adding “marginal product”, especially in their function as business forecasters (Coase 1937). The importance of innovators was elevated by Schumpeter (1947, p. 150), who saw their role as pivotal to change, and synonymous with progress itself. Schumpeter argued that the “creative response” of innovators (as opposed to the “adaptive” responses of others) was an “essential element in the historical process… (which) shapes the
whole course of subsequent events and their ‘long-run’ outcome”. In a similar vein, Aldrich and Zimmer (1986), and Anitsal and Anitsal (2011) argue that new organisational forms emerge because of the entrepreneurial activities of new entrants. Some more contemporary examples would be Ray Kroc (who founded Mc Donald’s in 1955) and James McLamore and David Edgerton (who launched Burger King in 1954), and triggered the franchising revolution which catapulted the franchising format as the fastest growing form of retailing in the world (Dant and Grunhagen 2014; Dant, Grunhagen and Windsperger 2011). Many writers regard entrepreneurship as a collective activity, which nowadays is subsumed as a part of the corporate mission. Cole’s “multiple entrepreneur” (Silk and Stern 1963) concerned entrepreneurship as a corporatised collective function, which Chandler (1961) endorsed and explained as a corollary to the development of “multi-function enterprise”. The consequence, as Schumpeter pointed out (1950, p. 132), is that “innovation is itself being reduced to a routine”.

The common emphasis of this category of works is on the importance of the role of innovators and innovative behaviour in effecting change. This moves beyond the behavioural locus of “self-interest” that underlies the power literature, to a broader consideration of human behaviour, in which self-interest may well play a part. The two schools are not incompatible in an epistemological sense. Recognising the role of innovative behaviour addresses the issue of change directly. It posits change as an endogenous process, in contrast with the next category of literature, which looks to external causes.

**Environmental Influences**

Compared with the literature so far discussed, works in this category are unified by their tendency to look at change as context-driven, or “at retailing from the outside” (Samson 1981, p. 32). Most works of an historical nature implicitly accept the environment as a first cause of
change, but have drawn criticism from relying upon anecdotal and not analytical evidence (Brown 1987). Indeed, one of the characteristics that separates historical research from other approaches is its broad domain, and attention to the macro factors surrounding the object of study, for example, Miller’s (1981) history of the world’s first department store, the *Bon Marché*. In general, historical works provide rich descriptive narratives of retailing evolution, which pay heed to the time and space context of change, and its antecedents. While such historical studies promise to extend our understanding of the environment of retail format evolution, Brown (1987) found they have limited value in predicting retail trends. The assumption upheld by historical research is that change cannot be understood outside of its context. While historical works provide the pedigree for marketing’s long history (for example, McArthur 2013), most extant historical work does not fully explain the reasons underlying retailing change and evolution. For example, while *Sears* and *Montgomery Ward* mail order trading was successfully adapted to environmental shifts in the 1920s (e.g., investment in warehousing facilities), many American department stores were slow to recognise the value in growing suburban locations (Brown 1987; Emmet and Jeuck 1950). So, too, retailers in the United Kingdom have been criticised for somewhat ‘lack-lustre’ responses to environmental changes in the 1970s and 1980s (Brown 1991b).

Another more recently developed body of works also gives primacy to historical context, but its arguments rest on the epistemological necessity of considering history. There are two main types of argument central to the claim that history matters. The first concern path-dependence arguments (David 1994; Hakansson and Lundgren 1997) about the succession or *flow* of events, and sunk-cost decisions, such as “lock-in” and the irreversible investments involved in, say, the choice of competing technologies (Cohen and Stewart 1994; David 1993) which predicted future events. Path dependence contributes to our understanding of cumulative
causation, with the observation that “extraneous features of the initial conditions, the historical context in which organizations are formed, can become enduring constraints” (David 1994, p. 214). Indeed, in his study of US department stores from 1910-1965, Wood (2011) found “lock-in” to the traditional downtown locations “posed significant obstacles” to adaptation and progress, and contributed to their ultimate decline.

The second type of arguments supporting consideration of history concerns behavioural expectations. These contend that history is important because it contains precedents for behaviour, which give rise to conventions and expectations of role conformity in relationships (Hodgson 1998). Habits and expectations are important because deviance from role expectations can be a significant source of conflict (Rosenberg and Stern 1970), which itself provides impetus for change.

Another group of works in this category of environmental influences identifies specific causal variables in the environment. Examples are Gist’s (1968) nomination of demographic, social, economic, cultural, technological, and consumer character variables, or the PESTLE framework of political, economic, socio-cultural, technological, legal and environmental variables. Extensions on PESTLE include Clarke’s (2005) application to international markets, which adds the variables of history, language, financial, rules, demographic, infrastructure and culture.

The major significance of these works is their explicit acknowledgment of multiple influences in the environment, especially consumers as a force of change - a “variable” that until this point in the development of thought on retailing evolution had been largely dealt with only indirectly (Buckley 2011). A spattering of literature appeared on “the consumer movement” following the rise of activist groups in the 1930s, but its focus was largely about consumers as a
new organised political entity, or largely as a party to be controlled (Blattberg and Glazer 1994, Craig-Lees 1991, Dameron 1941). Reynolds et al. (2007) pay heed to the role of consumers in driving change in retail formats, but fleetingly. More recent work by Bailey et al. (2010) examines consumer reactions to self-service supermarket formats in England (c. 1947-75) utilising a ‘life course’ perspective. Such research underscores the importance of recognising that consumers constitute a fundamental half of the exchange relationship and further research from a consumer perspective is warranted to address what is a significant structural gap in our understanding of retailing evolution (Cohen 1990).

Another school of study that places the environment at the centre of change is ecological theory (Etgar 1984). This approach views change in Darwinian terms of “natural selection whereby only the fittest retail species survive” (Brown 1987, p. 8), or an evolutionary process in which firms that do not adapt to their environment struggle and die out (Davies 1998; Levy et al. 2005; Williams 2009). While this idea may be intuitively appealing, ecological theory treats retailing as “a dependent variable in a world of environmental forces” (Hollander 1981, p. 90), and so ignores the process of how retail formats - and the inter-relations between retailers and the environment - are affecting change (Deutsch 2010). To some degree then, ecological theory is limited by its unidirectional, deterministic and mechanical thinking.

Characteristic of these works is the acknowledgement of the environment, and the “process” of change over time. While ecological theory represents the most promising of these attempts at explaining change, the assumption that the environment determines what succeeds, dismisses the essential principle of co-evolution (i.e., the joint and interactive outcome of managerial intentionality, organisational efforts and the environmental influences) that underlies works in the next category.
Interdependent Parts of the System in Co-Evolution

This is the most recently developed of the six groups of literature, and like the historical/environmental stream discussed above, these works concern the “urge to look at change” (Dosi 1990). Their difference, however, is in their epistemological position, since this body of work recognises interdependencies between components and two-way processes of co-evolution. Several theories are linked by this focus on the “wholes” rather than the “parts”, and the consequent refutation of single-cause arguments because of the implausibility of providing an “explanation of complex phenomena in terms of isolable elements” (Bertalanffy 1975, p. 151). These theories are variously known as evolutionary theory, co-evolutionary theory, and more broadly, systems theory.

Evolutionary theory according to Nelson and Winter (1993, p. 194) is “structurally” different from orthodox theory, presenting a “way of theorizing rather than an explicit formal theory… (though) the lines between these are somewhat blurry”. Nelson and Winter assert that existing “orthodox” theory about change must be “regarded as not merely oversimplified but a misleading structurally inadequate explanation… (because) important mechanisms are disregarded or dealt with shallowly”. The authors argue the value of re-specifying extant knowledge, which is characterised by “little theories aimed at different phenomena”, as “special cases of a broader ‘master theory’” (1993, p. 194-195). While Nelson and Winter acknowledge the purposeful behaviour of firms, the emphasis placed on aspects of “selection” in their evolutionary model on the behaviour of firms and industry response, implies that environments select organisations (Lewin and Volberda 1999).

Lewin and Volberda (1999) argue for a “co-evolutionary theory”, which “requires that sets of co-acting organizations and their environments be the object of study”. The authors
recommend four criteria for the co-evolutionary approach: studying organisations over a long period, examining organisations within a historical context of the firm and its environment, considering multidirectional causalities between micro and macro evolution, and the study of the interactions between multiple nested levels of co-evolution. These co-evolutionary criteria are well attended by systems theory.

Systems theory contends that any object of study cannot be fully understood outside the context of the co-evolution of the economic, social and political systems of which it is a part (Wilkinson 1990). In this approach, the object of study (the system), is viewed as embedded in a much larger system (the suprasystem), that is itself subject to complex change. In turn, the system comprises smaller subsystems from which it is aggregated, and which influence and proscribe the limits of its operations. This three-tiered set of relationships, in turn, must be understood in the context of the environment, and its physical and cultural parameters.

Systems theory recognises that retailing evolution must be both context and path dependent. History is important, because the “dynamics of a system are not only governed by where it is, but also by where it is coming from” (Hakansson and Lundgren 1997, p, 128). Local context is also important, because the terrain on which development takes place can be very uneven (Boulding 1981). Systems theory therefore provides for heterogeneity in both outcomes and initial states, which is absent from the more deterministic and monocultural literature thus far discussed. In its epistemological aim to provide a “way of seeing things which were previously overlooked or bypassed” (Bertalanffy 1975), systems theory also provides a structure that is “openly interdisciplinary in recognising insights” from a range of sciences (Hodgson 1998, p. 173).

As yet, this theoretical approach has not been applied to the topic of retailing evolution,
except by Evans, Barnes, and Schlacter (1993) whose work considers only the cyclical, conflict
and environmental literatures discussed earlier, but ignores the contributions from other
disciplines. Layton’s 2011 work examines the characteristics of marketing in general, which
allow it to be viewed as a system, and suggests, among other propositions, that “persistent
discrepancy in marketing systems is a driver of innovation” (Layton 2011, p. 263).

The systems theory approach addresses the main problem of conventional theory in
which “the broader conceptualization tends to be implicit and ad hoc, rather than explicit and
systematic, (and in which) a number of important issues tend to be neglected” (Nelson and
Winter 1993, p. 197). That is, compared with the adoption of a particular approach in the pursuit
of further knowledge, systems theory controls for “error” in knowledge development, by making
explicit the ways in which single-cause arguments are invalid.

While their emphases may differ, evolutionary, co-evolutionary and systems theory agree
on the basic or holistic system viewpoint, and reach similar conclusions. Their differences then
are largely semantic. Indeed, Bertalanffy introduced the term ‘general system theory’
“deliberately in a catholic sense… So the name, ‘general system theory’ may be used broadly,
similar to our speaking of the ‘theory of evolution’ which comprises about everything between
fossil digging, anatomy and the mathematical theory of selection… It is the introduction of a new
paradigm that matters” (1975, p. 157). The significance of these works is their recognition of the
interdependent and two-way process of evolution.

The main value of systems theory lies in its broad scope. This has particular merit in the
discipline of marketing, since “too much research in marketing concentrates on a major variable
without regard to the grand flow of events… The major variables are never placed in the context
in which they arise nor are they considered as to whether they are part of one or more ongoing
processes” (Savitt 1988). Similarly, Dixon and Wilkinson (1985, p. 62) point out that standard texts deal “solely with marketing management, or… the management of a small system”.

A meta perspective, therefore, provides an “alternative paradigm” for current-day readers in understanding and planning for the future - a view of change in a broader context, and as part of an ever-changing environment. This extends and anchors the microeconomic nature of knowledge in marketing, and provides a more useful approach for managers. As Layton (2011, p. 274) writes: “developing a theory of marketing systems may be an essential step to widening and deepening the scope of marketing thought, offering new horizons for our discipline”.

**Discussion and Conclusions**

When we look at the development of knowledge in its approximate chronological order, it can be observed that our understanding of retailing evolution has grown slowly, and is accumulative and both context and path-dependent. That is, to the extent that theory embodies empirical evidence, schools of thought develop in their own cultural and historical contexts and are influenced by local circumstances, habits, and attitudes (Coats 1976). The different “world views” (Bertalanffy 1975) that have evolved to explain exchange behaviour were products of their time and place, and the body of ideas that purports to explain retail evolution is in itself path dependent.

In the early twentieth century, the capacity and efficiency of manufacturing industries for handling an unprecedented flow of goods dominated contemporary thought. This was reflected in the growth of academic interest in the mechanical and rational science of “Taylorism”, with little thought being given to the “social” tasks of exchange that are performed by retailers. In the post-war 1920s, when supplies of goods around the Western world became plentiful, a more humanistic conception of the distribution channel began to take hold. Attention shifted towards the demand end of the channel - as captured in the emergence of the “customer is king”
philosophy. The comparatively sophisticated shopper of the 1920s “was no longer ‘a searcher but a choosser’, more prone to exercise independent judgment than simply to accept what stores featured… The balance of power between retailer and customer had shifted in a most disturbing way” (Benson 1986, p. 102). The corollary of this was the idea that retailers - and not manufacturers - occupied the cardinal position in the channel, because of their proximity to the consumer, and their concomitant role in influencing the consumer’s purchase decision.

By at least the 1950s, the prominence and power of manufacturers had fallen, in the face of the rise of large volume retailing intermediaries (Bucklin 1973), and a body of academic research on the significance of power relations had appeared. The concept of power inequities continues to be a central focus of research in the discipline of marketing, although its role in shaping change has been addressed only indirectly.

These bodies of knowledge that evolved over the twentieth century were typified by a neglect to explain the historical and cultural contexts under which their findings were made (Savitt 1988). This lack of attention to the influence of context may explain why the Western literature on retailing evolution falls short in its application to other nations (Davis and North 1971). Extant literature fails to explain why department stores did not take off in Eastern European countries (Pasdermadjian 1954, p. 7), or in much of Latin America (Rassuli and Hollander 1986, p. 9). In addition, most theories were developed largely without reference to other competing theories. This review has revealed significant overlap and thematic similarities across different disciplines.

Evolution in retailing entailed a compound of macro and micro influences and a process of co-evolution. The “causal variables” are multiple, and evolving, and the links between them are complex. If we look at the fundamental assumptions and beliefs that underpin the literature,
however, we can combine extant knowledge, despite its seemingly incompatible focal units.

This review provides a systems analysis of the underlying “causes” of retailing evolution, as proposed by leading scholars from multiple disciplines. While showing some fit with neo-institutional economic approaches to explaining retail evolution (North 1990), it does not attempt to develop a theory to explain retail evolution, but rather to recombine extant bodies of work that previously had remained conceptually unconnected. It presents a systematic typology of the cross-discipline corpus of literature according - not to final or “ultimate” causes but - to “first cause” arguments, or broad governing assumptions about change in retailing. By taking this macro perspective, we are better able to appreciate the many parts and purposes of extant knowledge, at least in English-language literature, and rather than assuming a default methodology, to not take for granted the supremacy of a particular theoretic point of view. This article argues that evolution in retailing cannot be properly understood outside of the historical, physical, and social environments in which it takes place, and without considering the interrelated systems of which it is a part. The methodology of a systems approach, and the content it yields, thus provides a broader and deeper understanding of retail evolution, in its ever-changing environment.

Six broad causal themes were identified in the extant literature on retailing evolution, wherein change is assumed to take place primarily as a function of economic efficiencies, patterns operating in nature, power inequities, innovative behaviour, environmental influences, and interdependent parts of the system in co-evolution. First, regarding economic approaches to explaining change, this review acknowledged their usefulness in providing a priori frameworks that explain why intermediaries appear in a channel, and in predicting the number of levels to be expected in a channel. Given the static assumptions and shortcomings of economic explanations
in explaining “uneconomic” behaviour, however, this review concurs with Stern and Brown’s assertion (1969, p. 13) that “the basic variables are social variables”. The economic tasks imposed on retailers by suppliers are just one subset of the “demand function”, and the behaviour and goals of actors are governed by more than merely functional considerations of costs, prices, and demand. Economic efficiency plays a role as a driver of change, but it leaves a great deal of change in retailing unexplained.

The second causal theme covers works bound by a shared faith, explicit or otherwise, in the existence of generalisable patterns in nature. Such patterns were discussed as cyclic or circular, polar or pendulum-like, and locational or geographic. Circumstantial evidence points to the existence of repeating patterns in a wide range of spheres (for example, Kondratieff 1950; Zarnowitz 1992), but inconclusively so. Schumpeter (1939) drew from history several synchronous waves that he attempted to crystallise in a scientifically predictive theory, although why it was possible to count off these cycles was “indeed difficult to see”. The general pattern of rise and fall, however, is repeated in the rhythmic tendencies to equilibrium in economic theory, and also in the dialectics of conflict models and countervailing power, in which seesaw patterns of tradeoff in the relative power of actors are observed. The persistence of claims for patterns in nature is notable, but seeking proof of their existence is outside the scope of this review. Ultimately, such patterns may be more useful in a predictive, rather than explanatory way.

Setting aside the possibility of a grand expository pattern as the entire explanation for retail evolution, we are left with causalities that are essentially environmental, or behavioural, or which recognise the interdependency of multiple factors that are co-evolving. There is much evidence to support environmental explanations of the evolution of retailing. Indeed, the rapid changes that took place in retailing in Western markets in the late nineteenth and early twentieth century could
be described almost entirely in terms of environmental influences. The expanding markets of the Industrial Revolution created the need for a distribution system much more efficient than that offered by small stores, fairs, and markets. Department stores back then had better environmental “fitness” than their competitors, although today they seem unfit and corpulent, and heading towards demise. Chain stores also were better adapted to the environment of the late 1920s, because they offered “low” prices at a time of steep inflation after World War 1, and located outside the rising rents of downtown city sites (where department stores congregated), moving out to the suburbs where land was cheap, and the population was growing.

Overall, historical evidence suggests that different kinds of environments yield different outcomes, and rates of change, and other factors must be taken into account. In Australia, for example, chain stores had emerged in Melbourne by 1914, but it was not until 1928 that they became viable in its largest city, Sydney, because of the dominance of Sydney department stores, which were among the most powerful firms in the country (McArthur 2005). Environmental theory is thus limited in its global applicability and, related to this, in its omission of the influence of retail formats on their environments, yet historians such as Leach (1993) and others illuminate the indelible impact of retail stores on their environments and indeed, on our culture of consumption.

This review reveals that most research assumes that phenomena can be studied as if they were disembodied from their environment. With its concentration on single variables, or small groups of variables at best, marketing atomises social phenomena with the result that the discipline presents as a “string of studies” (Savitt 1988, p. 127). In essence, the ahistorical assumptions and “closed system” thinking of single-cause theories are epistemologically indefensible. Instead, we need a meta and open system approach that takes in the panorama of
knowledge over time. For managers and for scholars, this means taking a holistic approach that rejects deterministic truths, recognises ambiguity, and accepts that history and context matter. Finally, if we accept a systems philosophy, we need to allow for causes that have not yet been identified, or accepted, in mainstream literature. The role of consumers in influencing the path of change, and the influence of retail stores on their environments, are two probable causes of retail evolution that deserve further study.

Moreover, the application of this framework in developing countries could prove beneficial in extending prior evolutionary perspectives in these localities (e.g. Goldman 2000, 2001; Goldman and Hino 2005). Precisely because the contextual realities and the experiential socioeconomic history of developing economies differ dramatically from the foregoing Western-centric historical evolutionary account of retailing, we need to embrace the systems theory perspective. Unique among the six perspectives presented in the above narrative, only the systems theory clearly embraces and endorses the premise that multivariate causation must be kept in central focus when developing any encompassing evolutionary account of retailing, and as such would be fully accommodating and accepting of novel contingencies that would no doubt be revealed when investigating the retailing trajectories of developing economies from an emic perspective.
References


Table 1: Summary of Primary “Causes” and Assumptions about Retailing Evolution in the Literature

<table>
<thead>
<tr>
<th>Change occurs primarily as a function of</th>
<th>Bodies of literature</th>
<th>Typical focal units</th>
<th>Contributions</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECONOMIC EFFICIENCIES</strong></td>
<td>• Microeconomic theories</td>
<td>Intermediaries</td>
<td>Institutions emerge (disappear) because of economic (in)efficiencies.</td>
<td>Static ceteris paribus assumptions; cannot explain “cost-inefficient” institutions, nor suboptimal behaviour.</td>
</tr>
<tr>
<td></td>
<td>• Transaction Cost Analysis</td>
<td>Transaction costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Functionalists</td>
<td>Utilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PATTERNS OPERATING IN NATURE</strong></td>
<td>• Wheel of retailing</td>
<td>Retail formats</td>
<td>Evolution is a fundamentally unstable - but rhythmic – process.</td>
<td>Determinism implied by patterns. Inability to explain the continued existence of formats that should, according to their own logic, have been superseded.</td>
</tr>
<tr>
<td></td>
<td>• Accordion of retailing</td>
<td>Individual institutions</td>
<td></td>
<td>Mainly normative research with implications for urban and other managerial planning.</td>
</tr>
<tr>
<td></td>
<td>• Retail Life Cycle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Conflict models and dialectic theory</td>
<td>Opposing parties</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>POWER INEQUITIES</strong></td>
<td>• Geographic theory</td>
<td>Geographic dispersion of stores</td>
<td>Understanding of spatial distribution of shops in geographic context.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Power literature</td>
<td>Dyads (manufacturer: retailer)</td>
<td>Change is human-centred and primarily behavioural.</td>
<td>Reification of power despite “enormous difficulties” in its measurement. Dyadic scope deals with evolution only indirectly.</td>
</tr>
<tr>
<td></td>
<td>• Countervailing power</td>
<td>Opposing parties</td>
<td>Change is a two-step process prompted by perceived power inequities.</td>
<td>Limited predictive power, and ignores variables other than power.</td>
</tr>
<tr>
<td>INNOVATIVE BEHAVIOUR</td>
<td>• Individual store histories</td>
<td>Individuals institutions and their founders</td>
<td>Rich descriptions and visual evidence of stores.</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Entrepreneurial literature</td>
<td>Innovators or innovative behaviour</td>
<td>Addresses the issue of change directly; offers a broader analysis of how human behaviour affects change beyond the motive of power.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hagiographies; single successful institutions and/or their founders.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ENVIRONMENTAL INFLUENCES</th>
<th>• Historical works</th>
<th>External context</th>
<th>Change cannot be understood outside of its context; cumulative causation must be considered; marketing has a long pedigree.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Ecological theory</td>
<td>General environmental characteristics</td>
<td>Exogenous forces affect evolution, and change is a “process”. There are multiple environmental influences, including consumers.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specific variables</td>
<td>Overlooks the principle of co-evolution and how institutions, and the inter-relations between institutions and the environment, are effecting change.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTERDEPENDENT PARTS IN CO-EVOLUTION</th>
<th>• Systems theory</th>
<th>Sets of relationships in their environmental context</th>
<th>Allows cross-discipline viewpoints, heterogeneity in outcomes and initial conditions, and two-way processes of co-evolution.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Evolutionary theory</td>
<td></td>
<td>Assumes all components are knowable. Inability to extract “universal” truths because of context dependence.</td>
</tr>
<tr>
<td></td>
<td>• Co-evolutionary theory</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
