UNIVERSITY OF SÃO PAULO
POLYTECHNIC SCHOOL

Charles Lincoln Kenji Yamamura

Dominant design and machine learning in new product forecasting

São Paulo
2022

CHARLES LINCOLN KENJI YAMAMURA

Dominant design and machine learning in new product forecasting

Revised Version
(the original version is available at the post graduate program unit)

Thesis presented to the Polytechnic School of the University of São Paulo (USP) for the degree of Doctor of Science.

Concentration area: Production Engineering

Advisors:
Prof. Dr. Fernando Tobal Berssaneti
Prof. Dr. José Carlos Curvelo Santana

São Paulo
2022

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, ______ de ____________________ de __________

Assinatura do autor: ________________________

Assinatura do orientador: ________________________

Catalogação-na-publicação

Yamamura, Charles Lincoln Kenji

Dominant design and machine learning in new product forecasting / C. L. K. Yamamura -- versão corr. -- São Paulo, 2022.

142 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.Previsão de demanda 2.Design dominante 3.Machine learning 4.Conhecimento de domínio 5.Invariantes I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II.t.

Nome: Charles Lincoln Kenji Yamamura

Título: Dominant design and machine learning in new product forecasting

Tese apresentada à Escola Politécnica da
da Universidade de São Paulo para a obtenção
do título de Doutor em Ciências

Área de Concentração Engenharia de Produção

Aprovado em:

Banca examinadora

Prof. Dr. Fernando Tobal Berssaneti
Instituição: Escola Politécnica da Universidade de São Paulo – Depto Enga. Produção
Julgamento: Aprovado

Prof. Dr. José Alberto Quintanilha
Instituição: Escola Politécnica da Universidade de São Paulo – Depto Enga. Transportes
Julgamento: Aprovado

Prof. Dr. Paulo Carlos Kaminski
Instituição: Escola Politécnica da Universidade de São Paulo – Depto Enga. Mecânica
Julgamento: Aprovado

Prof. Dr. Bruno Sanches Masiero
Instituição: Universidade de Campinas – Fac. de Enga. Elétrica e de Computação
Julgamento: Aprovado

Prof. Dr. David Kohan Marzagão
Instituição: King’s College London / University of Oxford – Dept. of Engineering Science
Julgamento: Aprovado

Miguel and Duda,

may the Future bring you thrilling new adventures.

Hold on to your dreams and never give up.

ACKNOWLEDGEMENTS

This PhD dissertation marks the fulfilment of a long-time dream and it would unlikely to happen without the friends and colleagues who stayed by my side along this exciting but by no means lenient journey.

I owe an immense debt of gratitude to my advisor, Prof. Dr. Fernando Berssaneti, for the continuous guidance, support, and encouragement. His deep knowledge, experience, and wisdom were the beacons in my wading through academic research. I thank Prof. Dr. José Quintanilha for his always present advice, wit, and friendship. Prof. Dr. José Curvello Santana, my co-advisor, for the invaluable guidance and support. Prof. Dr. Eduardo Zancul is the preceptor who originally encouraged me with his sapience and friendship. Prof. Dr. Paulo Kaminski, his guidance, wisdom, and generosity were fundamental to my success in the program. Prof. Dr. Celma Ribeiro has furbished my quantitative skills and her cardinal lessons will not be forgotten. Prof. Dr. Denise Dantas showed me the way of the *human touch* in design. I thank Prof. Dr. Bruno Masiero for the enlightening suggestions. My gratitude to Prof. Dr. Linda Hoo for the advice; Prof. Dr. Leandro Patah, for both trust and advice; Prof. Dr. Cynthia Campos and Prof. Dr. Cira Pitombo, for the technical contributions; Lidia Nogueira, for the continuous assistance; Claudia Machado, for the support and witty humor; Harmi Takiya, for the support. I thank Prof. Jim Euchner and Tammy McCausland for the precious lessons and suggestions.

My gratitude and deep affection to Pri, Marcio, Duda, Miguel, Nana, and Moana, who hosted me as a new family member during the hardships of the pandemic and have always stood by my side during this pilgrimage. I thank Veronica Rothe for the longtime friendship and support. Maria Clara Gonçalves, for the always graceful and enlightened companionship. My gratitude to Samantha Paul, Fabio Boccia, Patricia Carvallo, Emy Yanagizawa, Cristiane Meirelles and David Yamamura, for the encouragement; Francisco Ocejo and Samuel Obara, for both trust and friendship; Raquel Schilis, for the words of wisdom; Sandra Mancini, Larissa Duarte and Debora Correa, for the incentive.

My gratitude to the Production Engineering Department of the Polytechnic School of the University of São Paulo, which has always entrusted and welcomed me. I thank the Coordination for the Development of Higher Education Personnel (CAPES) for the scholarship; the Vanzolini Foundation for financial support; the Electromobility EPUSP Research Group, for the support.

Fernando Ranieri (in memoriam), a lifelong friend and supporter who is sorely missed.

La machine elle-même, plus elle se perfectionne, plus elle s'efface derrière son rôle. Il semble que tout l'effort industriel de l'homme, tous ses calculs, toutes ses nuits de veille sur les épures, n'aboutissent, comme signes visibles, qu'a la seule simplicité, comme s'il fallait l'expérience de plusieurs générations pour dégager peu à peu la courbe d'une colonne, d'une carène, ou d'un fuselage d'avion, jusqu'à leur rendre la pureté élémentaire de la courbe d'un sein ou d'une épaule. Il semble que le travail des ingénieurs, des dessinateurs, des calculateurs du bureau d'études ne soit ainsi en apparence, que de polir et d'effacer, d'alléger ce raccord, d'équilibrer cette aile, jusqu'à ce qu'on la remarque plus, jusqu'à ce qu'il n'y ait plus une aile accrochée à un fuselage, mais une forme parfaitement épanouie, enfin dégagée de sa gangue, une sorte d'ensemble spontané, mystérieusement lié, et de la même qualité que celle du poème. Il semble que la perfection soit atteinte non quand il n'y a plus rien à ajouter, mais quand il n'y a plus rien à retrancher. Au terme de son évolution, la machine se dissimule.

(Saint-Exupéry, 1939)

(*The machine itself, the more it is perfected, the more it effaces behind its role. It seems a man's whole industrial effort, all calculations, all sleepless nights over the blueprints, don't reach, as visible signs, anything but the sole simplicity, as if the experience of several generations were necessary to retrieve, little by little, the curve of a column, a fairing, or an airplane fuselage, just to give it the elemental purity of a breast or a shoulder's curve. It seems the craft of engineers, designers, and research office calculators is nothing more in appearance than to polish and to obliterate, to alleviate this connection, to balance that wing, just as it is no longer noticed, just as there were no longer a wing attached to a fuselage, but a perfectly blossomed form, finally relieved from its waste, a sort of spontaneous device, bonded mysteriously, and with the same quality of a poem. It seems perfection is reached, not when there is nothing more to be added, but when there is nothing more to be suppressed. At the end of its evolution, the machine conceals itself.*)

RESUMO

Yamamura, C. L. K. (2022). *Design dominante e aprendizagem de máquina na previsão de novos produtos* (Tese de Doutorado). Escola Politécnica da Universidade de São Paulo, São Paulo.

Devido à escassez de dados históricos e de ferramentas analíticas adequadas, os gestores normalmente recorrem à intuição e a heurísticas para a tomada de decisões sobre novos produtos. Ainda assim, as decisões não deveriam ser baseadas unicamente em intuição. Técnicas analíticas poderiam trazer consistência e confiabilidade às decisões. A aprendizagem de máquina pode capturar relações não lineares presentes em problemas reais, mas geralmente apresenta a desvantagem de exigir enorme quantidade de dados. O uso de invariantes presentes no conhecimento do especialista pode contornar esse problema, acelerando a convergência de problemas a soluções, sem a necessidade de grandes bancos de dados. Este trabalho propõe um método de previsão e suporte à tomada de decisões estratégicas sobre novos produtos, utilizando o conceito de design dominante e algoritmos de aprendizagem de máquina. O design dominante é um conjunto de atributos principais que definem uma categoria de produtos e é adotado pela maioria dos participantes no mercado. O período de evolução tecnológica anterior à emergência do design dominante é caracterizado por alta incerteza e competição entre diferentes conceitos de produto. Previsão nessa fase refere-se a entender os componentes tecnológicos principais e mapear suas trajetórias. Com a emergência do design dominante, a indústria entre numa fase de relativa estabilidade, com foco no aperfeiçoamento de processos e a inovação de produtos passa a ser incremental. A previsão consiste em determinar uma combinação de atributos vencedora e medir o tamanho da demanda. Invariantes no design dominante, codificadas implicitamente no conhecimento dos profissionais da indústria, são utilizadas para reduzir o espaço de possíveis variáveis. Uma hipótese de produto – conjunto dos atributos mais relevantes – é esboçada a partir do conhecimento dos especialistas do setor. Um banco de dados, a matriz de atributos e valores alvo, modela o mercado. Uma rede neural artificial extrai as relações não lineares no banco de dados, simulando o mercado. O método é demonstrado por estudos de caso da indústria automobilística, produzindo resultados significativos e mostrando que a abordagem pode ser utilizada em estratégia de novos produtos.

Palavras-chave: Previsão de demanda. Design dominante. Aprendizagem de máquina. Conhecimento de domínio. Invariantes.

ABSTRACT

Yamamura, C. L. K. (2022). *Dominant design and machine learning in new product forecasting* (Doctoral dissertation). Polytechnic School of the University of São Paulo, São Paulo.

Due to the lack of both historical data and adequate analytic tools, managers usually rely on intuition and heuristics in new product strategy decision-making. But decisions should not be based solely on intuition. Analysis brings consistency to managerial judgment. Machine learning can capture nonlinear relations in real-life problems, leading to more sound and reliable decisions. However, machine learning usually adds the caveat of requiring significant amounts of data. Using invariants from the expert's domain knowledge can circumvent that hurdle, accelerating the convergence of problems to solutions, without the need of significantly large datasets. This study proposes a framework to predict and to support new product strategy decisions, using the theoretical concept of dominant design and machine learning algorithms. A dominant design is the set of core features that define a product category and is adopted by the majority of players in a market. The period of technology evolution preceding the emergence of a dominant design is characterized by high uncertainty and intense competition among different product concepts. Forecasting in this phase concerns understanding core technology evolution patterns and mapping their trajectories. After a dominant design emerges, the industry shifts to a period of relative stability, with focus on process improvement. Product innovation is incremental henceforth. Forecasting centers on assessing winning combinations of features and predicting demand size. Invariants in the dominant design, implicitly encoded in the industry knowledge, are used to reduce the space of potential feature variables. A product hypothesis – a set of the most relevant features mapped to customer value – is drawn from experts' domain knowledge, research, and analysis. A database is synthesized from the hypothesis, consisting of a matrix of feature variables and fitness indicators (target values) which models the market. An artificial neural network extracts the complex nonlinear relations in the database, simulating the dynamics of the market. The framework is demonstrated with cases from the automobile industry, yielding meaningful results and showing the approach can assist new product strategy making.

Key words: Demand forecasting. Dominant design. Machine learning. Domain knowledge. Invariants.

FIGURES AND TABLES

CONTENTS

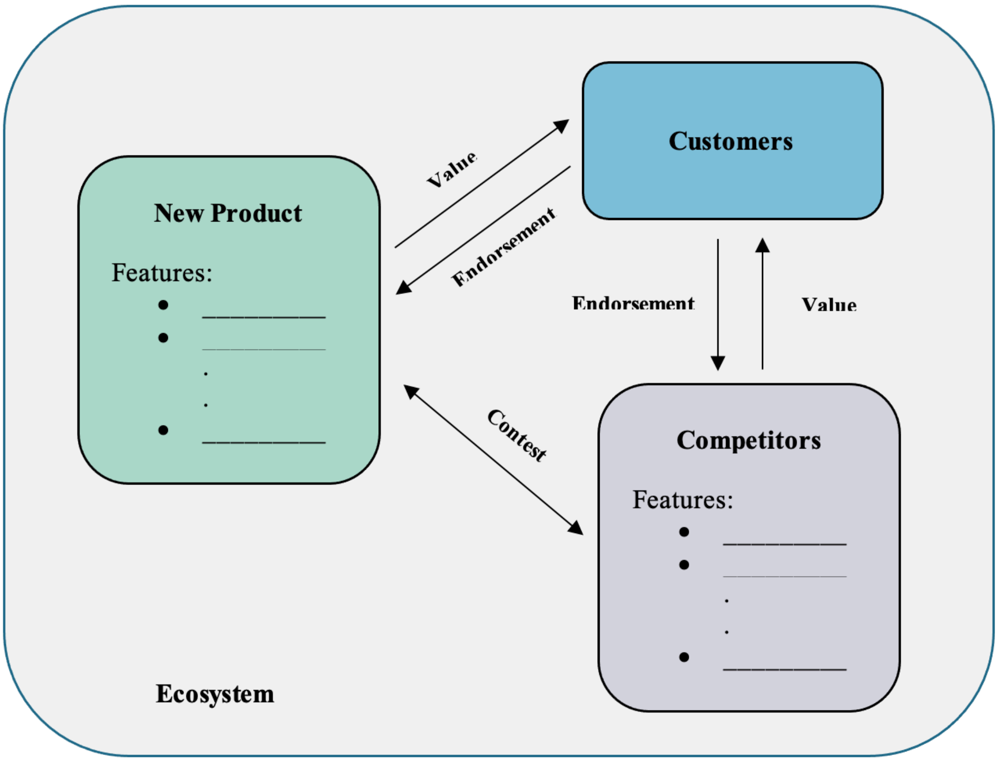
1 INTRODUCTION

Popular culture oftentimes produces witty pearls of wisdom. Countrymen in the state of Minas Gerais, Southeast of Brazil, are known for their diligence, affability, and introspection. A humorous popular reflection, with a regional colloquial accent, “oncotô? proconvô? concovô?” (“where am I? where am I going? how will I get there?”), spontaneously and systematically captures the essence of strategic thinking.

Complex products, like automobiles, take years to be developed (about four or five years), and must remain competitive for another few years (about eight years) after they are launched (Ulrich, Eppinger, & Yang, 2020). They come in a wide range of segments, prices, sizes, usages, customer needs and tastes, geographical conditions, technologies, legal requirements, and cultural flavors. Predicting and deciding a new product strategy is an overwhelming endeavor, enclosed in fizzle information, fleeting customer aspirations, and a protracted time horizons (Weber, 2009).

Concepts are decided today, but products must be sold in a distant tomorrow. It is not uncommon for a recently launched product to be matched or surpassed by new competitors, who raise the bar and leave the former behind in both customer value and market performance. The overall context of new product strategy is encapsulated in figure 1.

Figure 1 – New product strategy context diagram.



New product strategy is a vital activity for a firm's continuity, competitive advantage, and prosperity (Sorli et al., 2010; Ulrich, Eppinger, & Yang, 2020). It is also a notoriously difficult task, as it deals with complex (many variables in nonlinear relations) and dynamic problems (Simon, 1958; Spangler, 1991). To complicate matters further, some customer needs are *latent*, i.e., customers have difficulty to articulate them, and those needs often remain unmet by existing products (Ulrich, Eppinger, & Yang, 2020). Beyond reacting to customer needs, product strategy must predict and anticipate them – usually with little or no available data, as the new product is not in the market yet (Kahn, 2002).

Product features are variables that lead to potential customer value. However, according to Afrin, Nepal, and Monplaisir (2018), the most difficult task in new product strategy is not to identify feature variables, but to estimate their relative weights. Due to the complexity and dynamic nonlinearity among features, it is very hard to mentally assess their relative importance. The human mind has limited processing capacity and the functions (relations) that generate those weights are dauntingly complex and uncertain (Schneider & Leyer, 2019).

Owing to the scarcity of both data and adequate analytic tools, managers usually rely on intuition and heuristics to make decisions. Intuition is driven by the decision-maker's knowledge and perception of the business environment (customers, competitors, and technology), and expected changes over time. A common heuristic is *reasoning by analogy*, searching for similarities between present circumstances and past experiences (Spangler, 1991).

However, even grounded on experience, intuition may lead to wrong decisions. It strives to extract meaning from available information on customers and competitors, but the process is vague and subjective. There is variability in managers' individual knowledge, skills, and experience (Spangler, 1991). Moreover, businesses may struggle to keep track of existing knowledge and to use it effectively. Ideally, judgment in decisions should be supported by analysis, for more reliable outputs (Hair, 2007).

A machine learning (ML) algorithm extracts patterns that are difficult to map mentally or heuristically, making it suitable to handle real-life problems in product strategy. However, it comes with the caveat of usually requiring significantly large amounts of training data. When dealing with large and complex problems, obtaining sufficient data to train algorithms is still a major challenge. According to Ng (2016), data hunger is the Achilles' heel of *supervised learning* – the modality of ML where algorithms are trained with labeled samples.

A way to mitigate the voracity for data in ML is to use *invariants* – properties that do not change when a system suffers a transformation (Aleksandrov, Kolmogorov & Lavrent'ev,

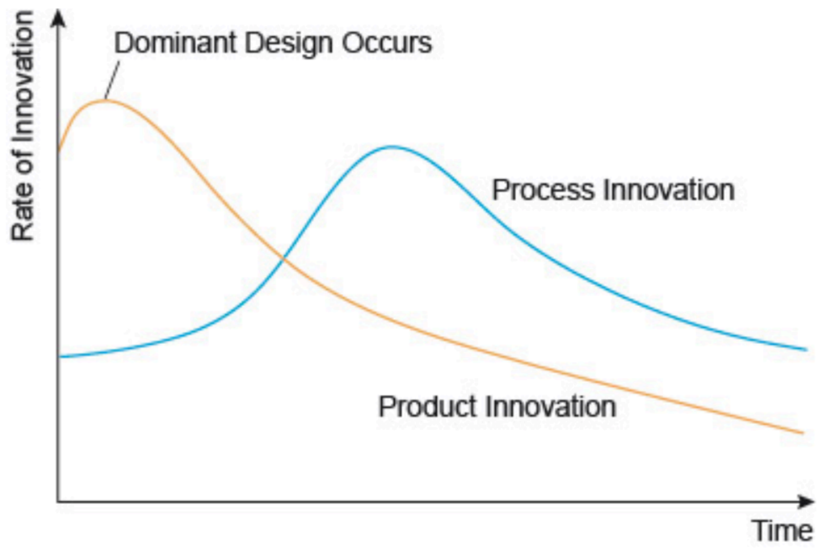
1999) – from the experts' domain knowledge (Vapnik & Izmailov, 2019, 2020). They use knowledge and experience to provide the initial conditions or assumptions – the invariants of the problem – for the ML simulation. Those are *shortcuts* that accelerate the rate of convergence of problems to solutions, instead of relying on significantly large datasets to capture patterns in data. The manager (domain expert) is an active participant in the whole analytic process, not a mere supporter just supplying directives and raw data, as is the case in conventional data analytics.

Summarizing, forecasting a new product's probability of success is one of the major business challenges, due to complex variables in nonlinear relations that are difficult to map mentally or by using statistical linear methods. Machine learning algorithms, like artificial neural networks, have the capacity to capture those nonlinear relations but usually require enormous amounts of data. This research proposes a new product forecasting method that precludes the need for huge datasets. The concept of *dominant design* provides the theoretical foundation for the proposed method.

1.1 Industrial Evolution and Dominant Designs

Utterback and Abernathy (1975, 1978) introduced the concept of industrial evolution as two out-of-phase wave curves of product and process innovations (figure 2). Initially, there are many firms, intense experimentation, variation in products, high growth rates, and high firm mortality. Anderson and Tushman (1990) call this phase the era of *ferment*. Battery electric vehicles (BEVs) are an example of industry in such stage of evolution (Brem & Nylund, 2021a). When the two innovation curves meet, a *dominant design* emerges as the amalgamation of prior competing and independent product innovations. The dominant design is a set of integrated product features (subsystems) that defines a product category and is adopted by the majority of players in the industry (Christensen et al., 1996; Brem & Nylund, 2021b).

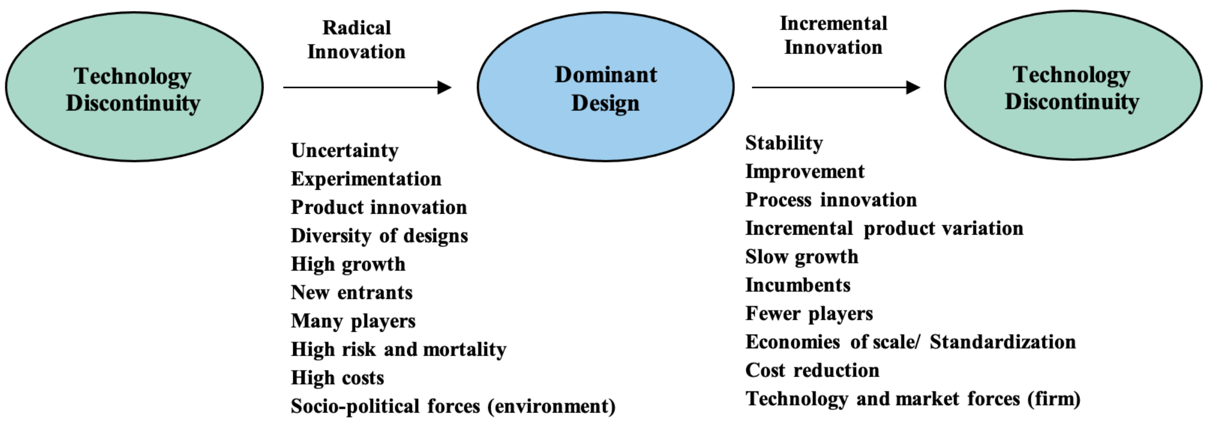
Figure 2 – Model of industrial evolution.



Source: Utterback & Abernathy, 1975; 1978.

The emergence of a dominant design marks the transition of focus from product innovation to process innovation (Utterback & Suarez, 1993; Cecere et al., 2015). Although the radical product innovation phase is over, there is an increase in incremental product variations (figure 3). A key aspect of the emergence of a dominant design is the dramatic reduction in product costs (Anderson & Tushman, 1990). The increase in production volumes accelerates learning, standardization, and modularization of components (Murmann & Frenken, 2006). Prices fall and most potential consumers adopt the new product. There is a parallel between technology evolution and consumer adoption processes, as innovation seeks demand from customers (Noel et al., 2019). A new technology discontinuity may break the process and start a new era of ferment (Anderson & Tushman, 1990).

Figure 3 – Technology evolution.



Source: Adapted from Anderson & Tushman, 1990; Argyres et al., 2015.

Inspired by Kuhn's model of scientific revolutions (1962), Dosi proposed an analogy between science and technology (1982), advancing the notions of technology paradigm and technology trajectory. A paradigm is the definition of relevant problems and solutions, existing knowledge and procedures for problem-solving, to the exclusion of alternatives. The paradigm exhibits momentum and progress occurs by updating its prescriptions from complex technology and market feedbacks, setting its trajectory. The trajectory results from the interaction of both paradigm and immediate experience (Kuhn, 1962), correcting a stream of imbalances in problem solving (Murmann & Frenken, 2006).

In both science and technology, the core components and principles of a paradigm become *invariants* in a stable knowledge base and are not revisited every time a new research or development program is started (Dosi, 1982; Murmann & Frenken, 2006). Continuous change (incremental innovation) occurs along an existing technology paradigm, while discontinuous change (radical innovation) is related to the emergence of a new paradigm. Unsolved problems can signal the need for a paradigm shift.

In isolation, neither market nor technology can explain a paradigm. Each technology paradigm has a set of economic and technology tradeoffs, related in a complex feedback process. Beyond technical and market factors, the selection process is strongly influenced by political and social dynamics, making the prediction of a dominant design difficult, if not impossible (Dosi, 1982; Anderson & Tushman, 1990). Being a compromise between sociopolitical and market forces, a dominant design is usually below the optimum technical performance frontier. A technological discontinuity or radical innovation happens when a new technology significantly pushes an existing technology frontier (Anderson & Tushman, 1990; Midler & Baume, 2010; Brem et al., 2016).

Technological discontinuities can be competence enhancing or competence destroying (Anderson & Tushman, 1990). Battery electric vehicles are competence destroying by making the knowledge of incumbent firms obsolete. New entrants usually initiate the obsolescence of existing technologies, but the incumbents' experience also contributes to the creation of a new technological order (MacDuffie, 2018). For instance, the combination of new digital technologies and traditional manufacturing is necessary to bring BEVs to fruition (Perkins & Murmann, 2018; Sovacool et al., 2019).

Currently, governments of leading economies are strongly influencing the technological discontinuity in the automobile industry. However, if there is too much government intervention or not enough scale, there is a risk of dominant designs not emerging, or to emerge prematurely and the result to be idiosyncratic – from both technical and market points-of-view

(Tushman & Murmann, 1998). European railway tracks and the space shuttle are cited as examples.

A systems view provides a better understanding of technology innovation and dominant designs (Murmann & Frenken, 2006). Dominant designs do not initially emerge at the product system level but first in components or subsystems (Brem & Nylund, 2021a). When dominant designs in the set of central or core subsystems consolidate, a dominant design emerges at system level. Core components are those that affect the largest number of product characteristics or features, i.e., they have many connections (Murmann & Frenken, 2006). Peripheral components, affect few characteristics and thus have fewer connections. The larger the number of connections in a product, the higher is its complexity (multiplicity of variables).

Core components have the largest impact in overall product performance. Changes in core components have a lower probability of success than changes in peripheral components, due to their complexity (Murmann & Frenken, 2006; Brem & Nylund, 2021a). For BEVs, central technologies are energy batteries, electric motors, and power control systems (MacDuffie, 2018; Enge et al., 2021; Yoshimoto & Hanyu, 2021; Xiong et al., 2022). In the era of ferment, companies tend to invest heavily in proprietary core subsystems. Each player fights fiercely in an effort to turn their solutions into industry dominant standards – instead of cooperating to create common standards – delaying the adoption of global dominant designs.

An operational principle is the knowledge and definition of how a set of components is integrated and works to accomplish its purpose (Murmann & Frenken, 2006). An operational principle settles the key dimensions of a system's design space, making possible to distinguish between variation within a dominant design and changes that create a new design. The set of all possible variations along the same operational principle defines its design space.

Architecture is the way components of a system are connected and organized (Henderson & Clark, 1990). A dominant design is a family of designs with common and stable core subsystems and architecture (Tushman & Murmann, 1998). However, dominant designs are unlikely to be present in all components (subsystems). Once the core components of a design are settled, development shifts to peripheral components. Core components become invariants which are not revisited in a new design (Henderson & Clark, 1990), reducing the design space, and restricting variations to peripheral features (Murmann & Frenken, 2006). A replacement of core components implies a change in the dominant design and a paradigm shift. Core components and features play an important role in product forecasting, first as key variables (in the radical innovation phase) and then as invariants (in incremental stage).

1.2 New Product Forecasting and Strategy

Decisions are preceded by understanding a problem and predicting their outcomes (Kahn, 2002). Effective decision-making implies good forecasting (Weber, 2009). New product forecasting is predicting the most likely outcome of a product concept, given a set of assumptions (Kahn, 2006). Forecasting *existing products* is performed using statistical linear techniques, like times series and regression methods. But *new product* forecasting is predominantly judgmental and seeks meaningfulness as the performance criterion – instead of statistical accuracy – as there is little or no historical data (Kahn, 2014). Even if analytic tools were used, the activity is still highly dependent on managerial judgment – for the initial assumptions, to fine tune the process, and to make the final decision to act (Bhattacharya, 2018).

New product forecasting and decisions are based on surveys and managerial judgment, sometimes on extrapolations made by assessing similarities with competitor products (Sharma, 2020). A meaningful forecast is one that can be used for decision-making, providing understanding of the problem (Kahn, 2002). The aim is to yield realistic and operationally useful approximations in predictions, instead of numerical precision.

Formulating strategies is a notoriously difficult task, as the problems they tackle are complex, unstructured, and fuzzy (Spangler, 1991). More than reacting to customer needs, product strategy needs to predict and to anticipate them (Day, 1994). Strategy is driven by experience and data, but also by the decision-makers' perceptions of the environment – competition, customers, and technology – and their expectations about the future. Competitors may be gauged either by listening and observing their explicit and implicit manifestations, or through modeling and simulation of their behavior (Spangler, 1991).

In predicting new product strategies, accuracy is not as important as approximate ranges that allow managers to understand the impacts and risks of their choices. The process starts with (usually little) available data and qualitative knowledge – from experience, judgment, and research. Human judgment can provide context, which machines would struggle to discern (The Economist, 2020). Assumptions are made, mental models are constructed (Nonaka & Takeuchi, 1995), alternative scenarios are designed and evaluated, considering contingencies. Decisions are taken assuming predictions will be numerically wrong, but within a meaningful ballpark (Kahn, 2014).

Kahn (2014) identifies a spectrum of seven possible new product categories: cost reduction, improvement (replacing an existing product), extension (addition to an existing product line), new usage, new (geographical) market, "new-to-the-company", and "new-to-the-world" products. Products that are new to the world are radical products, as they disrupt an existing industry and customer perceptions of the market (Kahn, 2018). The other six types are incremental products. Radical products are very challenging and risky (Ching-Chin, 2010), and companies must balance a mix of incremental and radical new products for a healthy business (Kahn, 2018).

Forecasting becomes harder and less accurate as products move along the incremental to radical scale, and applied techniques tend to drift from statistical to intuitive (Kahn, 2014). From a survey of 168 industrial companies (Kahn, 2002, 2014), average prediction accuracies are 72% (cost reduction), 65% (improvement), 63% (line extension), 54% (market extension), 47% (new-to-the-company), and 40% (new-to-the-world). The average accuracy was 58% (Kahn, 2002, 2010; Armstrong, 2002), but an acceptable level should be around 76% (Kahn, 2010; Ching-Chin, 2010) (table 1).

Table 1 – Prediction accuracy by new product category.

Category	**Demand forecast accuracy**
Cost reduction	72%
Improvement	65%
Line extension	63%
New geographical market	54%
New to the company	47%
New to the world	40%
Acceptable	**76%**

Sources: Kahn, 2002, 2014; Ching & Chin, 2010.

1.3 Judgment and Analysis

Product strategy problems are complex, ambiguous, unstructured, and not easy to be formulated explicitly and quantitatively (Kahn, 2002; Spangler, 1991). Hence managers usually rely on intuition and heuristics to make decisions (Krabuanrati & Phelps, 1998). Intuition is driven by the decision-maker's knowledge and perception of the business environment (customers, competitors, technology) and expected changes over time (Spangler, 1991). Decision makers usually attribute their confidence on intuitions to experience. But relying solely on intuition may be unsatisfactory, as it may lead to wrong decisions (Spangler, 1991; Krabuanrati & Phelps, 1998). Intuition in fact does not (or should not) work independently from analysis (figure 4a), the two being complementary for effective decision-making (Schneider & Leyer, 2019; Simon, 1987).

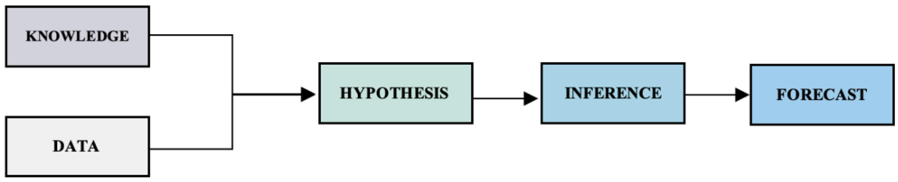
Conventional analytic techniques for prediction are data-driven, i.e., they start from the identification of patterns in data. Then hypotheses are formulated and tested. Theory may be part of the process or not, reason why data analytics is sometimes criticized as not being sufficiently rigorous and scientific (Hair, 2007). Data scientists use their technical knowledge and expertise to select and organize data, and to engineer features in the model; but the focus is still on data itself (figure 4b). In most cases, system and domain experts are different persons. Domain experts are often the decision makers. Usually, the latter are involved in the initial and final stages of the data analytics process, not in intermediate data analysis (Amershi et al., 2014; Gil et al., 2019).

Dealing with large and complex problems usually puts the caveat of requiring substantial amounts of data. The volume of required data is proportional to the complexity of the model (number of independent variables) and the random variation in data itself (Hyndman & Kostenko, 2007). At a minimum, it is necessary to have more observations than variables. But usually a much larger quantity of observations is needed because of data randomness.

On the other hand, big data tends to increase the complexity of strategic decision-making (Intezari & Gressel, 2017). It is the challenge of *information overload* (Merendino et al., 2018) – to manage and make sense of overwhelming amounts of data. Quantity of data does not lead perforce to quality of insight (Merendino et al., 2018). It does not preclude the need for rigorous inferential construction (Intezari & Gressel, 2017).

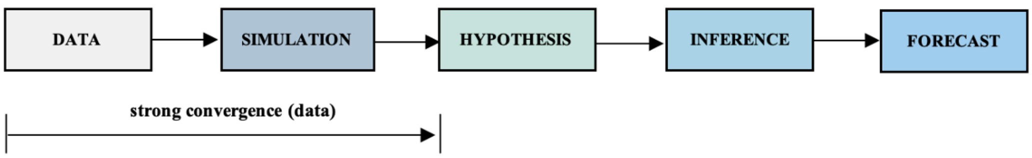
Figure 4 – Product forecasting approaches.

a)



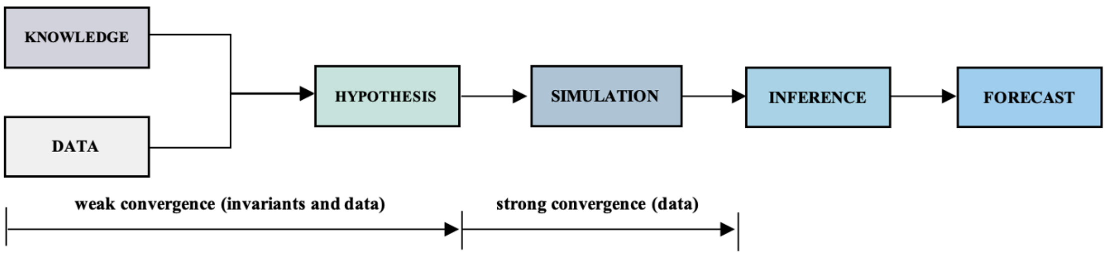
Purely inductive managerial approach (data is often limited or not available).

b)



Conventional algorithmic induction (requires significantly large datasets).

c)



Proposed method, using domain knowledge and machine learning simulation (domain knowledge reduces the need of very large datasets).

Multiple variables and ambiguous relations are unlikely to fit in simple models such as linear, logarithmic, quadratic, and exponential (Feng & Shanthikumar, 2018). Thus, fitting a model usually requires more complex models and huge data sets. On the other hand, in most business tasks – like in new product strategy – obtaining sufficient data to train algorithms is still a major hurdle. It is the Achilles' heel of machine learning (Ng, 2016).

1.4. Machine Learning

A machine learning algorithm is a computer program that improves its performance with data inputs, without being explicitly programmed (Goodfellow, Bengio, and Courville, 2016). It distinguishes itself from conventional algorithms by its capacity to learn or to improve with new data. Most current achievements in machine learning are of the *supervised learning* type, where algorithms improve by processing *labeled* data (Goodfellow, Bengio and Courville, 2016).

An artificial neural network is a machine learning algorithm that maps functions of complex feature variables to target values in nonlinear hierarchical relations (Goodfellow, Bengio and Courville, 2016; Aggarwal, 2018). Machine learning methods with multiple levels of representation are also called *deep learning* methods (LeCun, Bengio and Hinton, 2015). Data are processed by inputting feature data and adjusting the weights in the connections through layers of hidden nodes. The artificial neural network calculates errors by successive forward and backward iterations, with the goal of minimizing the derivatives of errors using stochastic gradient descent. As data move into higher layers of representation, the algorithm processes more abstract relations through the network (Aggarwal, 2018).

A categorical principal component analysis (CATPCA) reduces dimensions in the space of variables into a set of uncorrelated summary variables (principal components), assigning metric values to non-metric variables in a dataset with different levels of measures (nominal, ordinal, or numeric) and nonlinear relations. The algorithm seeks a small number of linear combinations that explain as much of variance in data as possible (Linting & Van Der Kooij, 2012; Linting et al., 2007).

1.5 Predicates and Invariants

Regularities or *invariants* in the laws of nature are what make science possible (Wigner, 1949). Invariants are conditions that do not change (they are constants) while an object suffers a transformation (Aleksandrov, Kolmogorov & Lavrent'ev, 1999). Invariants are also referred to as initial conditions, assumptions, and constraints (Wigner, 1949; 1960). They are dependent on the circumstances, context, and focus of a problem. Isolating invariants is not a trivial task. It is a difficult art and results from the scientist's knowledge, skill, and ingenuity (Wigner, 1960).

Vapnik and Izmailov (2019) proposed the theory of *learning using statistical invariants* (LUSI). Human learning requires far fewer examples than machine learning due to the use of intelligence, in contraposition to brute force. Learning has two modes – strong and weak convergence. Strong convergence uses observations from data to select a function (from a set of admissible functions) that minimizes the probability of error. Besides data, in the form of observations, weak convergence uses *invariants* (unchanging patterns in data) to accelerate the rate of convergence to a solution function (Vapnik & Izmailov, 2020).

Experts provide specialized information in the form of explanations, comments, and metaphors. Those are invariants, ubiquitous and found in almost any problem, according to the authors (Vapnik & Izmailov, 2019). In a technology paradigm, *invariants* are the core components, features, and attributes embodied in the dominant design. Incremental innovation sets a technology trajectory along the paradigm, generating feature *variables* in peripheral components and attributes (Dosi, 1982; Murmann & Frenken, 2006; Henderson & Clark, 1990).

In learning theory, a *predicate* is a logical statement or abstract function (Vapnik, 2020), which manifests itself either as a (feature) *variable* or an *invariant*. The difference between invariants and feature variables is increasing invariants leads to more accurate predictions, while the opposite is true for an increase in features, requiring more training data. On the limit, if all predicates were invariants, the dependent variable itself would be invariant and there would be only one function: the solution to the problem (Vapnik & Izmailov, 2020; Ling, Jones & Templeton, 2016).

Strong convergence occurs in a potentially infinite space of functions and is a "brute force" approach. Weak convergence uses "intelligence" for problem solving, in a more restricted set of functions (Vapnik & Izmailov, 2019) – the design space in the case of technology innovation (Murmann & Frenken, 2006). Many potential feature variables in a problem are irrelevant to the task at hand. Based on their knowledge and experience, experts construct invariants latent in the problem, instead of initializing from a blank sheet of paper.

Vapnik and Izmailov (2019) exemplify (jocosely) with the *duck test*: "if it looks like a duck, swims like a duck, and quacks like a duck, it probably is a duck" – just three attributes to qualify a duck, instead of many possible variables to qualify birds in general. Learning theory handles methods for converging the (potentially infinite) set of admissible functions to a much smaller set of desired functions. As there are only two modes of convergence (strong and weak), Vapnik and Izmailov (2020) claim their framework is the *complete theory of learning*. The new product forecasting method proposed in this study uses invariants from experts' domain knowledge – imbued in the canon established by the dominant design – reducing the dataset needed to capture relations and patterns in market simulation and forecasting (figure 4c).

The ultimate learning question is how to design a compact set of predicates encompassing the problem. *Compact* means it is *closed* (finite number of elements) and bounded (within a delimited range) (Aleksandrov, Kolmogorov & Lavrent'ev, 1999). Vapnik and Izmailov (2020) offer the analogy of Vladimir Propp's "Morphology of the Folktale" (1968), which summarizes not only Russian fairy tales, but also the narratives of novels, movies, theater, television, and games, in just thirty-one structural elements. The work attempts to ambitiously condense human life in a compact set of elements.

Vapnik and Izmailov (2020) also mention “the Art of War”, from Sun Tzu (1963), who prescribes thirty-three rules for effective warfare strategy, and is a widely adopted framework in both the military and business. The construction of a set of predicates that captures the essence of a problem can only be accomplished through human expertise and ingenuity. Predicate selection requires sound domain knowledge and is the foundation of both learning and problem-solving (Vapnik & Izmailov, 2020).

In our model, experts use their domain knowledge and experience to create a hypothesis (microtheory or cognitive model), to be tested and validated by a machine learning algorithm. It results from exclusion of dominant design core component invariants from the model, reducing the design variable space. The product hypothesis is further streamlined by handpicking peripheral features and components deemed to be the essential variables from the customer point-of-view, in accordance to the experts’ industry knowledge and judgment.

The hypothesis refer to key customer variables in a product category or segment. From a particular manufacturer’s point-of-view, even some of those variables can be invariants at certain moments of time, if it is unable to change its customer effects. E.g. brand image can be a given in the short term. Fuel economy and performance may be limited by past investments in powertrain technology.

1.6 Key Definitions

Product	Complex set of feature attributes that embody customer value and is built in nested hierarchical technology levels (system, subsystem, component) that evolve cyclically (Woodruff, 1997; Murmann & Frenken, 2006).
Customer value	Expression of the benefits perceived by customers; the capacity to satisfy their needs and aspirations (Woodruff, 1997).
Product concept	Unique set of feature variables mapped to customer value (Simon, 1996; Woodruff, 1997). Product concepts are embedded in product micro-theories (Marcus, 2020).
Micro theory	Theory or cognitive model about a small-scale phenomenon or a limited area of a wider subject (Marcus, 2020).
Cognitive model	Knowledge or representation about a particular phenomenon and its properties (Marcus, 2020).
Product hypothesis (cognitive model, micro theory)	Function that maps the product concept (set of feature variables) to fitness indicators (target variables) (Ng, 2022)
Feature variable	Dependent variable expressing a product characteristic or specification (Ng, 2022; Weber, 2006)
Fitness indicator	Independent target variable that expresses the adequacy of a product concept (Ng, 2022; Vincent & Brown, 2005)
Technology	Human made system of integrated components with the objective of satisfying customer needs and aspirations (Murmann & Frenken, 2006).
Innovation	An invention – new product component or architecture – accepted by the market (Nylund et al., 2021).
Dominant design	A set of features that defines the architecture of a product category and holds 50% of market share or more for at least four years (Anderson & Tushman, 1990; Christensen et al., 1996; Srinivasan et al., 2006).
Innovation shock / discontinuity	A firm introduces a new product that unexpectedly surges in demand and disrupts the existing technology paradigm (Argyres et al., 2015).
Complexity	Multiple integrated components and connections (Nylund et al., 2021).
Architecture	The way components in a system are connected, organized, and integrated (Brem & Nylund, 2021a).
Core subsystem	Subsystem with a lot of components and connections to other systems, strong impact in the overall system, and high strategic importance (Tushman & Murmann, 1998).

Peripheral subsystem	System with less components, connections, and lower impact and strategic significance than a core subsystem (Tushman & Murmann, 1998).
Ecosystem	Community of multiple actors and activities aligned to create and deliver value to customers. It includes producers, suppliers, competitors, distributors, customers, and other stakeholders, and adds *complementors* (Adner, 2006; Adner & Kapoor, 2010).
Domain knowledge	Knowledge pertaining to a particular specialized field, as opposed to general knowledge (Yu, 2007). In this study, domain knowledge is the sum of knowledge, experience and perceptions about customers, technologies, and competitors related to specific markets and products.
Machine learning algorithm	Computer program that improves its performance with new data inputs, without being explicitly programmed (Mitchell, 1997; Goodfellow, Bengio & Courville, 2016).
Artificial neural network	Machine learning algorithm that maps functions of complex feature variables to target values in nonlinear hierarchical relations (Goodfellow, Bengio & Courville, 2016; Aggarwal, 2018).
Predicate	Logical statement or abstract function of a relation in a phenomenon (Vapnik & Izmailov, 2020; Aleksandrov, Kolmogorov & Laurent'ev, 1999).
Variable	A predicate that can assume distinct (continuous or discreet) values as the phenomenon itself changes (Vapnik & Izmailov, 2020; Aleksandrov, Kolmogorov & Laurent'ev, 1999).
Invariant	A predicate that colapses into a single value, or constant. In other words, it is a condition that does not change when a phenomenon suffers a transformation (Vapnik & Izmailov, 2020; Aleksandrov, Kolmogorov & Laurent'ev, 1999).
Game	Model of the interaction among players using strategies, in pursuit of payoffs (Sharif & Heydari, 2013).
Players	Individuals with the same strategies and payoffs (Vincent & Brown, 2005; Vincent et al., 2011).
Minority game	Game where the winners are the players who stay in the minority, i.e., they play an advantageous strategy that differs from the strategy played by the majority (Arthur, 1994, 1997).
Prediction	Construction of hypotheses and analysis of their sensitivity to theory and data (Simon, 1996).

1.7 Research Aims and Objectives

This study aims to improve new product forecasting and strategy. It provides an effective set of principles and tools to add discipline, understanding, and replicability in the detection of complex patterns in forecasting problems that are difficult to map mentally. The objectives of this research are:

- To identify and assess the role of dominant designs in new product forecasting.
- To assess and propose a method using dominant design (recorded implicitly in the experts' domain knowledge and experience) and machine learning to forecast complex mass products, precluding the need for enormous datasets.

Thus, the hypotheses to be tested in this investigation can be stated as follows:

1. ***The concept of dominant design is a theoretical connecting thread that explains new product forecasting.***
2. ***Invariants from the dominant design, in combination with a machine learning algorithm, enhance the performance of new product forecasting.***

1.8 Original Contribution

The original contribution of this research lies in using **the concept of dominant design as a theoretical foundation to explain and guide new product forecasting**. A novel method for incremental new product forecasting is proposed as a direct consequence of the theoretical construction, in combination with machine learning algorithms (figure 5).

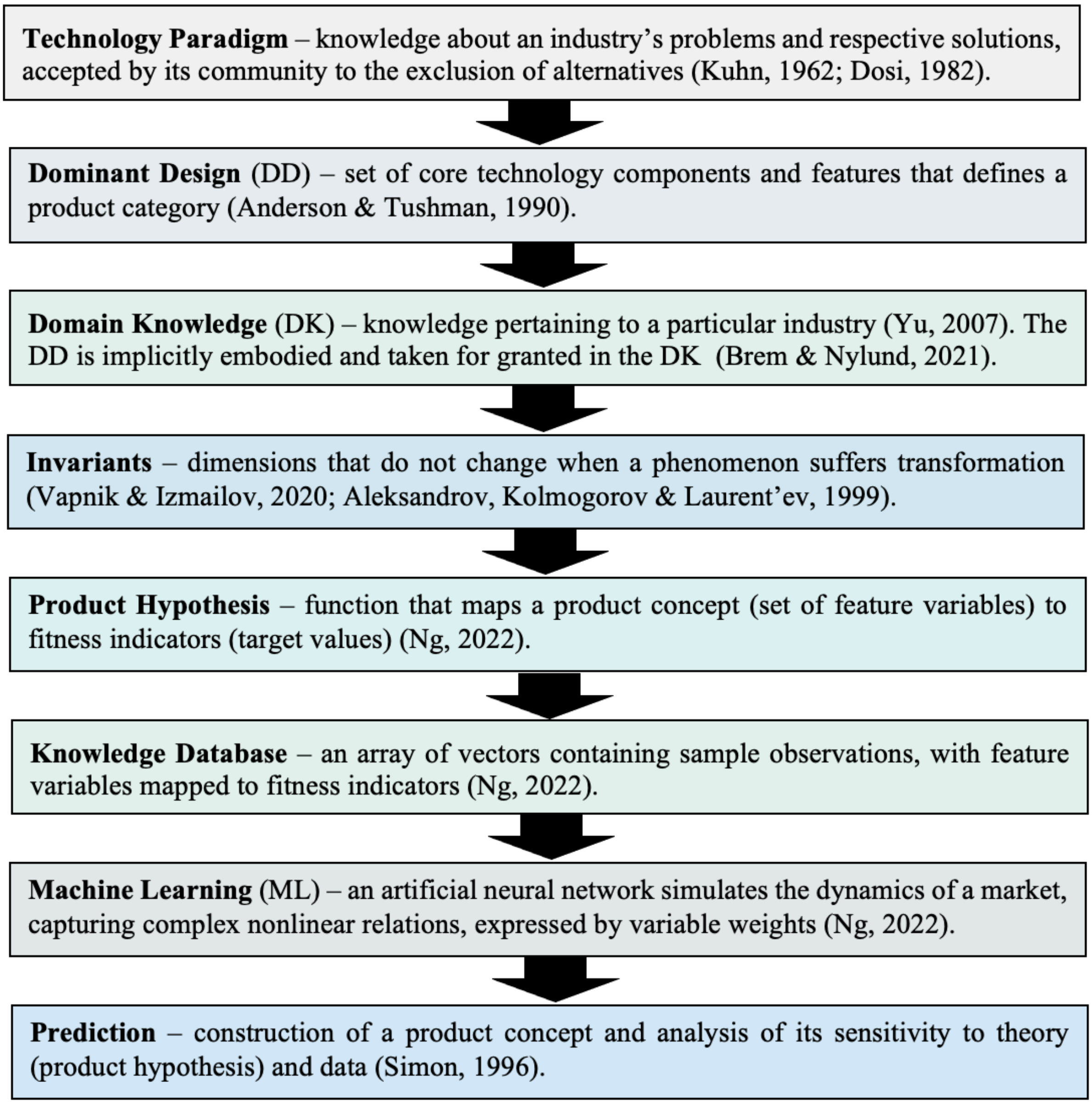
Prior to the emergence of a dominant design (in the radical innovation stage) core technology analysis is used to explain the technology evolution process and to map its trajectory. After the emergence (in the incremental innovation phase) a dominant design provides the canon, in the form of invariants present in experts' domain knowledge and experience, to reduce the design space and streamline the construction of a *product hypothesis* – a *compact* set of most relevant product features mapped to customer value.

The use of invariants accelerates the convergence of a set of admissible functions (potential solutions) to the desired function, precluding the need for a vast amount of training data, as this is usually available in limited supply. A product hypothesis is designed as a cognitive model of the market, according to the experts' knowledge and experience. In the proposed method, it is a function that maps product features to fitness indicators (customer value, expressed by market share in the experiments).

A database is synthesized using the features in the product concept and available sales data on similar products. Each observation in the sample (a vector in the data matrix) contains evaluated features from the concept (predictor variables) and market share (target value) calculated from sales figures. The *weights* of feature variables are estimated by a machine learning algorithm. Although the variables themselves are selected using domain knowledge, getting adequate weights from expert judgment can be difficult, due to the nonlinear relations in variables (Afrin, Nepal & Monplaisir, 2018).

The combination of database and the trained machine learning algorithm is a model of the market and plays the role of an integrated knowledge repository and continuous learning tool. New product concepts can be simulated to estimate their market fitness. They are updated as decisions are taken and new information is received, being applicable to future new projects.

Figure 5 – Proposed new product forecasting method using invariants from the dominant design and machine learning.



1.9 Method

This study is grounded on literature review, construction of theoretical frameworks, machine learning simulation, and case studies. Artificial neural networks (ANNs) were used in the experiments to simulate and validate new product concepts. Literature research included academic documents, technical reports from government agencies, think-tanks, and consultancies, specialized press articles, and books in a wide range of topics, including technology evolution, product innovation, forecasting, strategy, decision-making, game theory, predictive analytics, machine learning, energy transition, automobile industry, and clean energy vehicles. Key constructs were defined, and conceptual frameworks were built from those sources.

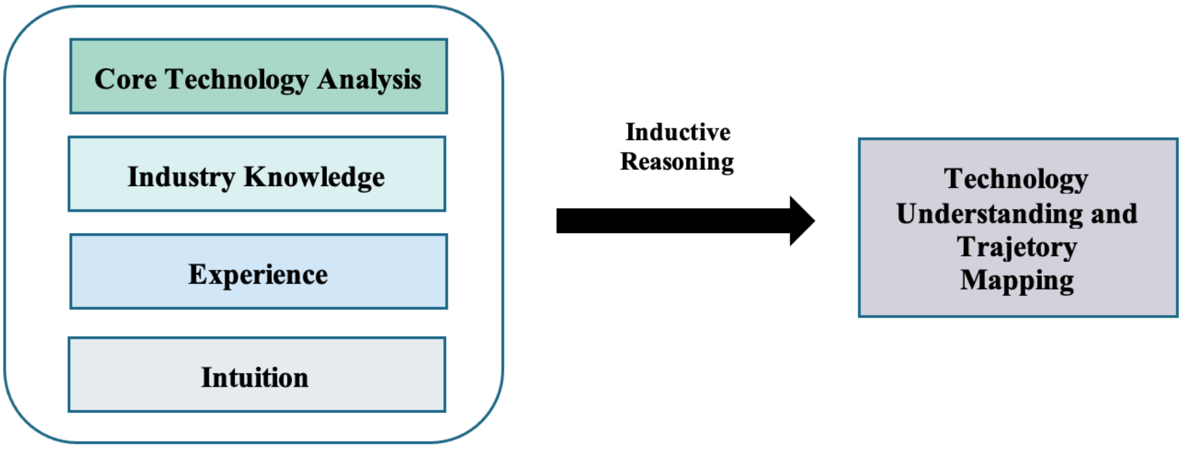
The inductive case study is the foundation of the research methodology, seeking explanation of complex phenomena under the logic of theory (Eisenhardt et al., 2016). Case studies in this investigation refer to electric cars in Brazil, for technology forecasting and strategy prior to the emergence of a dominant design, and small SUVs, in the incremental stage of product innovation. The cases exhibit both typical and extreme elements of the studied phenomena, namely the role of dominant designs in the technology evolution process and their implications in new product forecasting and strategy. Typically, a case study allows deep understanding and prediction of a particular phenomenon (Eisenhardt, 1989). The study of an extreme case provides a clear vision of the problem and can lead to new insights (Eisenhardt et al., 2016).

We rely on the concept of dominant design as the guidepost to orient the forecasting process. The initial step is to understand if there is a dominant design in the industry and, if not, whether it will emerge – it does not have to necessarily arise (Srinivasan et al., 2006; Cecere et al., 2015; Brem et al., 2016; Chen et al., 2017). A dominant design is present in an industry if there has been a standard set of core technology subsystems holding fifty percent or more of a market for over three years (Murmann & Frenken, 2006; Srinivasan et al., 2006; Anderson & Tushman, 1990).

Prior to the emergence of a dominant design, an industry is in radical innovation or "ferment" stage (Anderson & Tushman, 1990; Chen et al., 2017), and strategists want to know the likelihood of a dominant design to emerge, identify its core components, to estimate the timing of emergence, and how the emergence process will unfold. To assess the emergence of a dominant design, it is necessary to go down in the hierarchical analysis, to the subsystems and components level. Core subsystems need to be identified. Those are the subsystems or components with the highest level of variables and connections to other subsystems, and the largest impact in the overall product system performance.

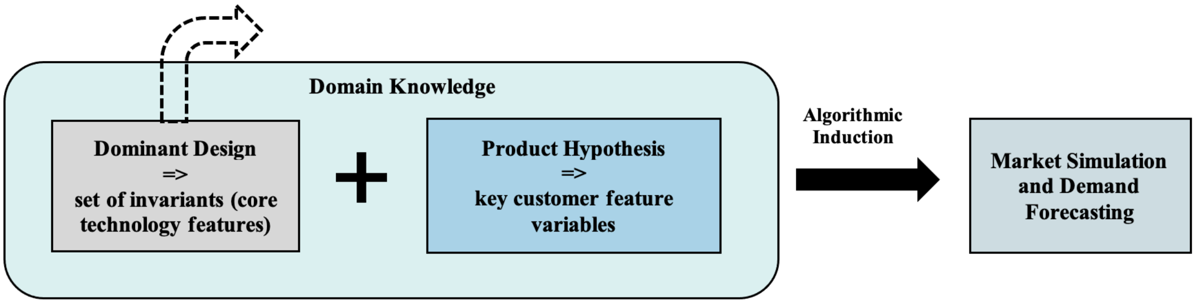
Once core subsystems are identified, strategists need to study the evolution of those subsystems. Prediction at this stage consists in understanding behavior patterns and trends in the technology, not measuring precise demand (figure 6). The emergence of dominant designs is determined not only by technical and market factors, but also by socio-political variables, implying the winner will not always be the "best" technology (Chen et al., 2017).

Figure 6 – Product forecasting in the radical innovation stage, prior to the emergence of a Dominant Design.



The emergence of a dominant design marks the transition from radical to incremental innovation. For this stage, the method proposes using domain knowledge and machine learning to new product forecasting and decision making (figure 7). The dominant design contains the standard features that must remain constant for a product to be accepted and be successful in the market. Among the potentially infinite features in a design space, many are held constant and incorporated into the industry knowledge, and taken for granted by experts and professionals. Invariants in the dominant design significantly reduce the set of features (design space) that defines a new product concept.

Figure 7 – Incremental product forecasting, using invariants from the dominant design and a product hypothesis extracted from the expert's domain knowledge.



Invariants are "negative" elements that are *left out*, *not built-in*, to significantly reduce the design space. Even after setting invariants in a product concept, there is still a wide range of possible variables. It is the experts' task to further streamline the variables and determine a compact set of the most relevant features from the customer's point-of-view. It is their cognitive model or hypothesis, grounded on knowledge, experience, research, and intuition.

A knowledge database is synthesized from that hypothesis and sales data on similar products (there is no direct historical data from a product that was not launched yet). Evaluated features and target values (market share calculated from sales) in the observations are vectors in a database matrix. The database is processed by a machine learning algorithm (an artificial neural network), which captures patterns and estimates the weights of variables. Weights are the parameters that set the relative importance of multiple features variables in nonlinear hierarchical relations.

In conventional data analytics, professionals start to tackle problems handling complex variables empirically without understanding them properly. They simply try to fit a curve on data (Marcus, 2018). The approach is long on data but short on knowledge. In the domain knowledge-based approach, experts start from understanding and building a cognitive model (also called product hypothesis, micro-theory, or customer narrative) to represent (model) the problem. The convergence of problems to solutions using data only is strong convergence; learning using both data and invariants from knowledge is weak convergence. The combination of both database and machine learning algorithm works as a knowledge repository and learning tool to predict the fitness of alternative product concepts.

Different professionals may produce different product concepts, based on their interpretations of competitive product and customer value. Despite their relativity and subjectivity, those concepts are operationally useful and essential ideas that can be improved through learning (Marcus, 2018, 2020). A machine learning algorithm measures the correlation between product concepts (sets of feature variables) and fitness indicators (target variables), giving objectivity to the initially subjective hypotheses. A cognitive model (product hypothesis) suggests causality between the set of feature variables and target values. The better a product concept, the higher is its probability of success (Marcus & Davis, 2019).

We demonstrate the incremental innovation framework with an experiment on small SUVs (sport utility vehicles) in the Brazilian market. The experiment shows a machine learning algorithm is a useful tool to support new product strategy decisions. The trained artificial neural network (ANN) captured the complex nonlinear relations among the variables, producing meaningful results: a prediction accuracy of 91% using the full dataset (82% with partial datasets coming from different probability distributions) – compared to the standard accuracy of 58% achieved by most firms (Kahn, 2002, 2010).

1.10 Research Structure

This PhD thesis adopts the set of research papers model and contains four articles, three of them published and one in the first review process. The thesis also comprises four major sections. This introduction gave context to the work, presenting the research problem and general literature, research aims and objectives, hypotheses, methodology, key definitions, and an explanation of its original contribution.

The next section summarizes the contents of the four research papers, followed by comments and explanation of their integration as a unified and consistent whole. The conclusion section wraps up the work discussing results and presenting implications, limitations, and suggestions for future research. The last part of the study contains full texts of the research papers.

2 SUMMARY OF RESEARCH PAPERS AND INTEGRATION

2.1 Paper #1

The first article, “electric cars in Brazil: an analysis of core green technologies and the transition process” (appendix A), explains the role of dominant designs in the evolution of industries and how core technology analysis can lead to understanding and mapping technology trajectories. The emergence of a dominant design is the landmark of the transition from product innovation to process innovation and it is essential for achieving economies of scale in production (Anderson & Tushman, 1990). In the case study, the rise of a dominant design is understood as a requirement for the mass dissemination of electric vehicles in emerging countries.

The emergence of a dominant design happens at the technology core subsystem level (Brem et al., 2016). Once dominant designs are established in core subsystems, they integrate into a technology system. To understand the process of dominant design emergence, it is essential to analyze and understand the process of core technology evolution. Once a dominant design is established, its features and components turn into standards that are implicitly encoded in the industry’s common knowledge and practice. They are taken for granted among professionals in the field and are no longer revisited or questioned when new products are designed and developed. In other words, they become *invariants* in new product development (Brem & Nylund, 2021a).

Manufacturing companies can focus on the improvement of production techniques because they just follow the prevailing standards and established designs, and do not have to worry about disruptions in core technologies. A technology paradigm is relatively stable, and firms also gather attention and resources on the improvement of product details (Nylund et al., 2021). Traditional automobile manufacturers master incremental product innovation and excel in manufacturing technology. On the other hand, relatively new entrants (like Tesla) are agents of discontinuity and bring radical product innovation in the form of viable electric cars to the automobile industry (MacDuffie, 2018; Perkins & Murmann, 2018).

The theory and process of dominant design emergence are illustrated with the case study of electric cars in Brazil. Core technologies for electric cars are energy batteries, electric motors, and power control systems (Enge et al., 2021). Electric vehicle technology is currently in radical innovation or “ferment” stage, with intense competition among multiple players, including several new entrants (Brem et al., 2016). There are no established core technology standards, and most players invest heavily to achieve dominance with their own technologies.

Preference for in-house development may delay the emergence of dominant designs. Under such circumstances, it is difficult to achieve cost reductions from economies of scale, which are necessary to make electric cars affordable to automobile buyers in developing economies (Brem & Nylund, 2021a). For that reason, hybrid technologies will be needed to bridge the transition from combustion engine vehicles to pure battery electric vehicles in emerging countries. Hybrid vehicles combine internal combustion engines with electric motors and batteries.

Since batteries are the most expensive components in pure battery electric vehicles (typically between 30% and 50% of total vehicle cost), hybrid vehicles are less expensive because their batteries are smaller (Jetin et al., 2020; Yu et al., 2020). In terms of evolution of technologies and their ecosystems, we have a case of *robust resilience* (Adner, 2006; Adner & Kapoor, 2010). The incumbent technology (internal combustion engines) still exhibits vitality while the new technology (battery vehicles) has many obstacles to overcome – cost, range, charging infrastructure, vehicle weight, etc. – before coming to fruition (Teixeira et al., 2015).

Pure battery vehicles capable of covering the full-scale usage range of current internal combustion engine vehicles – including three long distance vacation trips a year – can cost almost three times as much as equivalent combustion engine counterparts (Arora et al., 2021). Without the intermediate development of hybrid electric vehicles, there is the risk of exacerbating a social divide, making new cars a privilege of just the most affluent consumers. Mainstream consumers would keep their old internal combustion vehicles for longer periods, with detrimental impact on global emissions.

Brazil could focus on developing and manufacturing small biofuel electric hybrid vehicles (derived from new global platforms) targeting medium class consumers in emerging markets. But clear and ambitious – albeit realistic – industrial strategy and government policies are in tall order. The paper provided a case study demonstrating the analysis of core technologies in electric vehicles and their ecosystem can lead to a better understanding of the technology evolution process and support industrial forecasting and strategy.

2.2 Paper #2

The second paper, "forecasting new product demand using domain knowledge and machine learning" (appendix B), is the backbone of the paper set, as it elaborates on the incremental product forecasting method using invariants and machine learning. Invariants allow a machine learning algorithm to capture complex patterns with a much smaller dataset than conventional data analytics. A reduction of two orders of magnitude (100x) in the number of observations is possible (Vapnik & Izmailov, 2020). The way to achieve predictive power using small data is to use invariants established by the dominant design and encoded in experts' domain knowledge and experience, reducing the space of potential product feature variables (Brem & Nylund, 2021a).

Invariants significantly reduce the design space – the potential set of feature variables – which is further reduced to its essentials, still using experts' domain knowledge. They are expressed in a set of product variables mapped to a fitness indicator or target variable. In the studied case, it is an expression of customer value quantified as market share. Together, the set of features and indicator makes the cognitive model or hypothesis of a product category. But the model is still incomplete. The weights of variables in complex nonlinear hierarchical relations are missing. They are estimated by an artificial neural network processing the small dataset. Once the model is trained (it exhibits a stable set of parameter weights), it can be used to predict the fitness of different product concepts.

The model is demonstrated by the case of small SUVs in the Brazilian market. A database was designed using sixteen independent variables, identified as the most relevant product features expressing customer value. Industry knowledge, information from specialized press, and textbooks were used as proxies for industry domain knowledge to select those feature variables. Sales figures from 2003 to 2020 were converted to market share values and used as fitness indicators – target dependent variables in the model. The dataset covers the history of small SUVs in Brazil, with slightly over 1,500 observations (a smaller subset proved equally effective to capture data patterns, as can be seen in paper #4), and act as a knowledge repository and continuous learning tool. The algorithm captures customer value and competitive relations, assessing the relative importance of features, and simulating market dynamics.

Initially, the full dataset is shuffled. Training, validation, and test sets are drawn stochastically, and separated from the beginning at 70:15:15 ratios. The ML algorithm produced predictions with a correlation factor R of 0.96 to the real training data. To test for overfitting, i.e., to measure its capacity to capture patterns in new data (not present in the original training dataset), six real products were used to test the method. Neural networks were trained with portions of the dataset from periods prior to the product launch dates (hence there is no data on those products in the training datasets). An average accuracy of 82% (18% error) was observed, showing the model has useful predictive power.

The experiment proved it is possible to use invariants – defined by the dominant design and extracted from the expert's domain knowledge – to significantly reduce the size of data samples to train an algorithmic model and to predict demand for incremental new products.

The requirements for the application of the proposed forecasting method are:

- Complex products (multiple feature variables) produced in relatively high volumes.
- Competitive market with a minimum of about seven distinctive players.
- Presence of a dominant design in the market for at least three years.
- Domain expertise to extract invariants and to design a product hypothesis (set of most relevant product features mapped to fitness indicators).
- Minimum amount of fitness indicator data, like sales, market share, revenues, profits, customer satisfaction etc. – a few hundred observations are needed.
- Machine learning algorithm (e.g., artificial neural network) to capture the nonlinear hierarchical relations in the model.

2.3 Paper #3

The next article, "the minority game in product strategy" (appendix C), explains and explores the implications of the dominant design and incremental technology evolution in competitive market strategy. The minority game happens in two stages or dimensions: compliance to the core technologies (invariants) and differentiation in peripheral features. Following the canon established by the dominant design is a requirement to be accepted and stay competitive in a relatively stable market (Brem & Nylund, 2021a). But once those demands (core technologies, product features, and attributes) from the dominant design are met, competitive advantage is obtained by adroit differentiation.

The winners of the minority game are the players who adopt advantageous product strategies, not followed by the majority of the market (Arthur, 1994, 1999). In other words, products must contain invariant core components to be part of the game, but they need to differentiate with unique sets of peripheral features that are particularly attractive to customers and are hard to copy. The market is dynamic, competitive strategies evolve continuously, players strive to match and outperform their rivals. The minority game tends to equilibrium overtime, but it is dynamically unstable. If the majority adopts a supposedly winning strategy, it turns into a *losing strategy*, as they defeat the minority principle by following the same strategy. Players are forced to differ and the game is in continuous flow (Casti, 1996; Challet et al., 1997, 1998, 2001, 2004, 2005).

The method using a knowledge database and an artificial neural network is demonstrated simulating the small SUV market in Brazil. Different strategies can be tested and their fitness can be estimated using the model. Just sixteen feature variables in a five score Likert scale can potentially yield 5^{16} (approximately 150 billion) product strategy combinations. Firm and environment constraints significantly reduce the strategy space, but the likelihood of two firms adopting the exact same strategy (all sixteen features with the same scores) is still very small, meaning competitors are likely to differ.

In our complex product (small SUV) experiment, fitness is measured by the relative presence of a strategy, expressed by share in the total automobile market. If all firms in a hypothetical market had the exact same resources, played the same product strategies (e.g., they were commodities with no room for differentiation), under the same conditions, their individual market share would simply be the market size divided by the number of players. But winners in complex markets are the competitors who seize larger than average market share. They are part of the winning minority. The study identified eight historical winners in the small SUV market, with consistent above average market share.

A categorical principal component analysis – CATPCA – method (Campos et al., 2020; Kuroda et al., 2013; Linting & Van Der Kooij, 2012; Linting et al., 2007) was used to understand the relations among the sixteen variables, clustering them, and yielding three main groups or components. The first component clusters comfort, convenience, equipment, infotainment, and finish. They are attributes related to technology and directly associated with price. Expensive vehicles tend to incorporate more technology and convenience features. A second axis identified cabin and trunk space inversely associated to agility and economy, all related to vehicle size. Large vehicles tend to be more spacious both in the cabin and trunk. Smaller vehicles are usually more agile and energy efficient. A third axis related brand and robustness. Attributes like durability and reliability tend to be incorporated into a brand's reputation.

The analysis of winning minority strategies allowed the identification of distinctive traits. Winners usually have strong brand reputations and are not necessarily the most affordable products. There are winning propositions in affordable, economical, and agile products, but also in more expensive, comfortable, and technology rich alternatives. Affordable but spacious and reliable is also an attractive formula. Those scenarios were used to illustrate the method applied to competitive strategy analysis, combining the model with invariants from domain knowledge, machine learning, and the minority game framework.

2.4 Paper #4

The last article in the series (appendix D) explores how the forecasting method using experts' knowledge and machine learning can be applied to the so-called *front end* of new product development – FEPD (Koen et al., 2001). Also known as *pre-development*, it comprises the exploratory new product activities prior to the official approval and budgeting of a project (Cooper, 1988). Potential alternative concepts are assessed and evaluated on a preliminary basis. Each product concept – set of distinctive product features – is understood as a potential product strategy to be pre-evaluated before choosing a project (or a small set of projects) for in-depth valuation and then development.

Front-end activities in product development usually start with the collection of information on existing products, technology, customers, and competitors. But decision-making is usually fuzzy, vague, and subjective, based on experience, intuition, and heuristics (Reinertsen & Smith, 1991). The article proposes the use of micro theory (product hypothesis or cognitive model) building and statistical learning (in the article, the expression *statistical learning* was used refering to *statistical machine learning*) simulation methods to improve the front-end of product development.

An earlier version of the forecasting method using domain knowledge and machine learning was used in the article case. The database contained information on small SUV sales in Brazil from 2013 to 2018. The sample contained 715 observations (less than half the sample size used in the case study of article #2) but the method was the same and illustrated its application in the front-end of product development. It shows how machine learning can contribute to add rigor and a systems approach to product strategy and development. The use of invariants in the experts' domain knowledge reduces the space of potential feature variables and increases the precision of predictive analytics, significantly reducing the size of data samples at the same time.

Selecting product concepts at the FEPD means scanning a large number of ideas in a short period of time with little or no budget. The simulation of alternative product concepts using machine learning increases the performance of predictions with a tool that is analytical, have low cost, and is relatively easy to use. While article #3 demonstrates the application of the proposed method in competitive strategy, article #4 shows how it can be used by firms developing their own products.

2.5 Unification of Papers

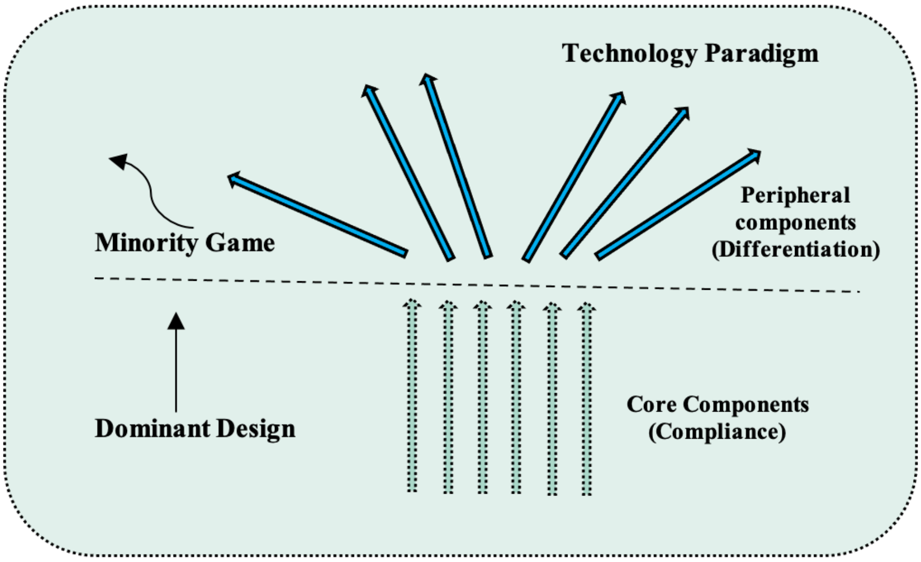
An innovation discontinuity or shock starts a period of intense experimentation and uncertainty, with multiple product variations and new competitors (Argyres et al., 2015). In this period of "ferment", the strategist's task is to understand how the technology is going to evolve and try to determine its trajectory (Brem et al., 2016). Analyzing the technology into its core components allows strategists to identify evolution trends and map their patterns. Although it is not possible to determine which exact solutions will prevail in complex technologies (amalgamating into dominant designs), due to sociopolitical variables that act on top of technical and market factors, it is possible to estimate the likelihood of dominant core components to emerge, their approximate timing, and how the processes will unfold (Anderson & Tushman, 1990).

The study on electric cars in article #1 illustrates how the study of core technology components can be used to understand and map a technology trajectory. The objective is not to determine the precise shape or moment of emergence of a dominant design, but to monitor and extrapolate its evolution path, providing elements for on-going decisions in a period of high uncertainty prior to the emergence of dominant designs. In the radical innovation period, firms are not worried about exact sales volumes. Their priority is to understand where the industry is heading, which technologies to invest in (maybe to bet on), and what to do.

Once a dominant design is established, the industry moves to the incremental innovation phase, with relative stability in core product features, customer profiles, and competitive behavior. The core features of the dominant design are implicitly encoded in the knowledge and practice of the industry and its professionals (Henderson & Clark, 1990). A technology paradigm bounds the trajectories that incremental innovation may take and evolve in. It is possible to perform more accurate product forecasting, strategy, and planning. The market exhibits stable patterns, although they may be too complex to be mapped mentally (Schneider & Leyer, 2019). Machine learning algorithms can be used to capture the complex nonlinear market behavior. Article #2 demonstrates how invariants from the dominant design and a machine learning algorithm can be used to model a market in incremental innovation stage.

Article #3 shows how the methods and tools presented in the two previous articles can be interpreted and used in competitive product strategy, with addition of the *minority game* framework (Arthur, 1994, 1999). A minority game is combination of both compliance and differentiation in distinct dimensions or stages. Overall, a new product needs to comply to dominant design prescriptions to be part of the game; its core concepts are sine qua non requirements for competitiveness and survival in the market (Argyres et al., 2015). But once the dominant design canon is satisfied, there is still considerable leeway in terms of possible peripheral product features – incremental innovation – to appeal to customers and distinguish a product from its competitors (figure 8).

Figure 8 – The Minority Game.



The objective is to differentiate by incremental product innovation, achieving competitive advantage by playing the minority game. For instance, a firm may choose to offer distinctive product features at a premium or to be competitive in cost, reducing product content. But relations among product variables are complex and hard to figure out mentally. Playing the minority game demands both creativity and knowledge. Since good forecasting is a condition for good strategy making (Wilson, 1999), the article explores how the proposed forecasting method can be applied to successfully play the minority game. The article relies on the Brazilian small SUV case to demonstrate the forecasting method, illustrating its application in competitive product strategy. Besides the artificial neural network used to model competitive strategies and extract relations among the feature variables, Categorical Principal Components Analysis (CATPCA) is used to understand the relations in clusters of variables.

Article #4 describes how the proposed method can be applied to the initial stage of product development. In the "front-end" of product development (FEPD), business ideas are identified, product concepts are generated and screened for project approval. As the project is not budgeted yet, this early and often underrated phase of product development needs an efficient and succinct method to analyze a wide range of potential product ideas.

Instead of using heuristics and intuitive guessing, the method adds rigor and systemic view to decisions in the FEPD by employing the proposed new product forecasting approach to initially screen alternative product concepts. A full-scale technical and economic due diligence to evaluate the range of potential product concepts in not feasible at this stage, because of both financial and time constraints. There are little resources to study a project that is not officially approved yet. Predictive accuracy, speed, low cost, and ease of use are some of the requirements addressed by the proposed method using machine learning.

3 CONCLUSION

New product strategy is a challenging task due to complex, nonlinear, and dynamic relations among its variables and lack of historical data. Product strategy making encompasses understanding a problem, forecasting outcomes, and making a decision. Managers usually rely on judgment derived from intuition and experience, and seasoned experts can handle those decisions with dexterity. But intuition and experience could still be wrong, and current practices may not be adequate. There is too much variability and loss of experience and knowledge. There is also a lack of systems-view and integration.

On the other hand, there used to be a mix of reluctance and perplexity to use machine learning to support strategic decision-making. But events are changing, and time may be ripe for a better integration between human domain knowledge and machine learning algorithms in new product forecasting and strategy.

In Greek mythology, a centaur was a creature with head, torso and arms of humans, body and legs of horses. Nowadays, it refers also to the human-machine combination playing *freestyle* chess against humans and machines alone or other "centaurs". In 2005, a pair of amateur players with three ordinary computers and a well-planned strategy won an international freestyle tournament, beating supercomputers and also grandmasters with powerful computers, but not so well-structured methods. Kasparov (2017) summarized: "*weak human + machine + better process* was superior to a strong computer alone and, more remarkably, superior to "*strong human + machine + inferior process*".

In the same token, a combination of experts' domain knowledge and machine learning can lead to enhanced new product strategy decisions. A usual caveat of machine learning applications is the requirement of large amounts of data to train algorithms. A way to accelerate the convergence of a problem to its solution, without the need of large datasets, is to use predicates and invariants from the experts' knowledge and experience.

Big Data uses hundreds of thousands (sometimes millions) of observations. By selecting appropriate invariants and variables to reduce the design space, it is possible to shrink the training dataset by up to two orders of magnitude. In our experiment, the full dataset has about 1,500 observations and sixteen variables, a ratio of about a hundred observations per variable. Partial simulations showed about three hundred observations (about twenty observations per variable) are sufficient to provide a meaningful prediction.

The proposed approach – integrating domain knowledge and machine learning – effectively supports new product strategy making. Initial conditions, assumptions, and constraints are extracted from the technology paradigm, encoded in the dominant design and present in the industry experts' domain knowledge. The use of invariants accelerates algorithmic learning without significantly large datasets. A machine learning algorithm is the implementation tool.

In the proposed method, a database and a neural network model the market, mapping product features to customer value. Multiple alternative concepts covering the strategy space can be tested simultaneously, early in the development process, and at low cost. As decisions are taken, new knowledge is integrated into the model, which is used to support new project decisions, and may be shared among different organization areas for continuous learning. The added layer of consistency and discipline provided by the approach reduces both uncertainty about customer value and the risk of wrong and unused information.

The combination of theoretical construction and experiments proved the **concept of dominant design as a theoretical foundation to explain new product forecasting**, the original contribution of this investigation to the advancement of academic knowledge, yielding meaningful results that support new product strategy decision making. A **novel method is proposed from the theoretical framework**, presenting the following distinctive features:

- It uses the theoretical concept of dominant design as the guidepost to both explain and improve new product forecasting.
- Employs the invariants from the dominant design – embodied in the experts' domain knowledge and experience – to reduce the design space, accelerating the learning rate and avoiding large datasets.
- Uses domain knowledge to construct a product hypothesis, selecting a closed set of feature variables mapped to target variables.
- Synthesizes a database from the hypothesis and a limited amount of sales data.
- Estimates weights directly from data – instead of judgment – using machine learning.
- Acts as a knowledge repository and continuous learning tool.

The automobile industry has applied artificial intelligence mainly in fields such as advanced manufacturing, fleet management, predictive maintenance, and self-driving cars. Most of those applications have in common a well-defined and stable environment, a fixed set of rules and clear objectives. Strategic decision making, however, is inherently ill-defined, complex, and adaptative, and it poses significant challenges. Human intelligence is much better on dealing with ambiguity, making analogies, and grasping the overall picture. A symbiosis between humans and machines can produce superior results. The better the expert, the better the tool. By continuously learning, the converse is also true: the better the tool, the better the expert. This investigation has shown human and machine intelligence interface to simulate, to understand, and to predict new product markets is a sound proposition.

3.1 Limitations and Future Research

Currently, the world is experiencing a period of high uncertainty and instability, with disruptions in both markets and supply chains caused by the Covid-19 pandemic, wars, inflation, and shortage of materials. The panorama of markets after those events is not clear yet. Although the theoretical foundations of this research should remain, the tools and operational details of the proposed method (like the dataset used to train the models) may need to be re-assessed when global markets settle down into possible new circumstances (value chains, price levels, consumer behavior, social demands, legislation etc.). It is difficult to appraise if the current market dataset will be effective to predict new product concepts under the new scenario. The model may need to be retrained with new data from the post crisis era, even if the principles stay the same. A review of the method under the new circumstances can bring additional and useful insights.

This PhD thesis was based on studies about the automobile industry. Conducting research on a single industry has the advantage of providing depth of analysis and identification of unique patterns and insights that contribute to expand the knowledge in a field of study. The logic of the research leads to infer both method and insights from this study are equally applicable to other complex mass products and businesses (e.g., digital technologies, entertainment, food, apparel, and publishing industries). But extensive investigation of other industries, sectors, and contexts will certainly contribute to solidify and deepen the assessments of the present work.

An artificial neural network was used as the default machine learning algorithm in this study. Artificial neural networks are among the most popular and capable machine learning tools used in a wide range of applications, from financial credit assessment and recommendation systems to self-driving cars and robotics. However, there is a wide range of other machine learning tools both currently used and under scrutiny, like decision trees, support vector machines, and extreme learning machines. Investigations using a variety of algorithmic tools may further enrich research in the field of machine learning applied to new product forecasting and strategy.

REFERENCES[1]

Abernathy, W. J., & Utterback, J. M. (1978). Patterns of industrial innovation. *Technology Review*, 80(7), 40-47.

Adner, R. (2006). Match your innovation strategy to your innovation ecosystem. *Harvard Business Review*, 84(4), 98.

Adner, R. & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, 31(3), 306-333.

Afrin, K., Nepal, B. & Monplaisir, L. (2018). A data-driven framework to new product demand prediction: Integrating product differentiation and transfer learning approach. *Expert Systems with Applications*, 108, 246-257.

Aggarwal, C.C. (2018). *Neural Networks and Deep Learning: a Textbook.* New York: Springer.

Aleksandrov, A.D., Kolmogorov, A.N. & Laurent'ev, M.A. (1999) *Mathematics: its Content, Methods, and Meaning* Gould, S.H. (transl.). Mineola, NY: Dover Publications.

Amershi, S., Cakmak, M., Know, W.B. & Kulesza, T. (2014). Power to the people: the role of humans in interactive machine learning. *AI Magazine*, 35(4), 105-120.

Anderson, P., Tushman, M.L. (1990). Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative Science Quarterly*, 35(4), 604-633.

Armstrong, J.S. (2002). Benchmarks for new product forecast errors. *Journal of Product Innovation Management*, 19, 133-143.

Argyres, N., Bigelow, L., Nickerson, J. A. (2015). Dominant designs, innovation shocks, and the follower's dilemma. *Strategic Management Journal*, 36(2), 216-234.

Arora, A., Niese, N., Dreyer, E., Waas, A. & Xie, A. (2021). *Why Electric Cars Can't Come Fast Enough*. Boston Consulting Group.

Arthur, W.B. (1994). Inductive reasoning and bounded rationality. *American Economics Review,* 84(2), 406-411.

Arthur, W.B. (1999). Complexity and the economy. *Science,* 284(5411),107-109.

Bhattacharya, P. (2018). Artificial intelligence in the boardroom: enabling 'machines' to 'learn' to make strategic business decisions. *IEEE Fifth International Conference on Information Technology Trends*, 170-174.

Brem, A. & Nylund, P.A. (2021a). Home bias in international innovation systems: The contingent role of central technologies in the emergence of dominant designs in the electric vehicle industry. *Journal of Cleaner Production*, 321, 128964.

[1] In accordance with the APA (*American Psychological Association*) style.

Brem, A. & Nylund, P. A. (2021b). Maneuvering the bumps in the new Silk Road: open innovation, technological complexity, dominant design, and the international impact of Chinese innovation. *R&D Management,* 51(3), 293-308.

Brem, A., Nylund, P.A. & Schuster, G. (2016). Innovation and de facto standardization: The influence of dominant design on innovative performance, radical innovation, and process innovation. *Technovation*, 50, 79–88.

Campos, C.I., Pitombo, C.S., Delhomme, P. & Quintanilha, J.A. (2020). Comparative analysis of data reduction techniques for questionnaire validation using self-reported driver behaviors. *Journal of Safety Research*, 73, 133-142.

Casti, J.L. (1996) Seeing the light at El Farol: a look at the most important problem in complex systems theory. *Complexity,* 1(5), 7-10.

Cecere, G., Corrocher, N. & Battaglia, R. D. (2015). Innovation and competition in the smartphone industry: Is there a dominant design? *Telecommunications Policy*, 39(3-4), 162-175.

Challet, D. & Zhang, Y.C. (1997). Emergence of cooperation and organization in an evolutionary game. *Physica A: Statistical Mechanics and its Applications*, 246(3-4), 407-418.

Challet, D. & Zhang, Y.C. (1998). On the minority game: Analytical and numerical studies. *Physica A: Statistical Mechanics and its Applications*, 256(3-4), 514-532.

Challet, D., Chessa, A., Marsili, M. & Zhang, Y.C. (2001). From minority games to real markets. *Quantitative Finance*, 1(1), 168-176.

Challet, D., Marsili, M. & Ottino, G. (2004). Shedding light on el farol. *Physica A: Statistical Mechanics and its Applications*, 332, 469-482.

Challet, D., Marsili, M. & Zhang, Y.C. (2005). *Minority games: interacting agents in financial markets*. Oxford, England: Oxford University Press.

Chen, T., Qian, L. & Narayanan, V. (2017). Battle on the wrong field? Entrant type, dominant designs, and technology exit. *Strategic Management Journal*, 38(13), 2579-2598.

Ching-chin, C., Ieng, A.I.K., Ling-Ling, W. & Ling-chieh, K. (2010). Designing a decision-support system for new product sales forecasting. *Expert Systems with Applications*, 37(2), 1654-1665.

Christensen, C. M., Suarez, F. F. & Utterback, J. M. (1998). Strategies for survival in fast-changing industries. *Management Science*, 44(12-part-2), S207-S220.

Cooper, R.G. (1988). Predevelopment activities determine new product success. *Industrial Marketing Management*, 17(1988) 237-247.

Day, G.S. (1994). Continuous learning about markets. *California Management Review*, 36(4), 9-31.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3), 147-162.

Eisenhardt, K.M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.

Eisenhardt, K.M., Graebner, M.E. & Sonenshein, S. (2016). Grand challenges and inductive methods: rigor without rigor mortis. *Academy of Management Journal*, 59(4), 1113-1123.

Enge, P., Enge N. & Zoepf, S. (2021). *Electric Vehicle Engineering*. New York, NY: McGraw Hill.

Feng, Q. & Shanthikumar, J.G. (2018). How research in production and operations management may evolve in the era of big data. *Production and Operations Management*, 27(9), 1670-1684.

Gil, Y., Honaker, J., Gupta, S., Ma Y., D'Orazio, V., Garijo, D. & Jahanshad, N. (2019). Towards human-guided machine learning. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 614-624.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. Boston, MA: MIT Press.

Hair, J.F. (2007). Knowledge creation in marketing: the role of predictive analytics. *European Business Review,* 19(4), 303-315.

Henderson, R.M. & Clark, K.B. (1990). Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 35(1), 9-30.

Hyndman, R.J. & Kostenko, A.V. (2007). Minimum sample size requirements for seasonal forecasting models. *Foresight*, 6, 12-15.

Intezari I, A. & Gressel, S. (2017). Information and reformation in KM systems: big data and strategic decision-making. *Journal of Knowledge Management*, 21(1), 71-91.

Jetin, B. (2020). Who will control the electric vehicle market? *International Journal of Automotive Technology,* 20, 156–177.

Kahn, K.B. (2002). An exploratory investigation of new product forecasting practices. *Journal of Product Innovation Management*, 19(2), 133-143.

Kahn, K.B. (2006). *New Product Forecasting: an Applied Approach*. Armonk, NY: ME Sharpe.

Kahn, K.B. (2010). New product forecasting. (2010). *Wiley International Encyclopedia of Marketing*.

Kahn, K.B. (2014). Solving the problems of new product forecasting. *Business Horizons*, 57(5), 607-615.

Kahn, K.B. (2018). Understanding innovation. *Business Horizons*, 61(3), 453-460.

Kasparov, G. (2017). *Deep Thinking*. New York, NY: Public Affairs.

Koen, P.A. et al. (2001). Providing clarity and a common language to the "fuzzy front end, *Research-Technology Management*, 2, 46-55.

Krabuanrat, K. & Phelps, R. (1998). Heuristics and rationality in strategic decision making: an exploratory study. *Journal of Business Research*, 41, 83-93.

Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago, IL: The University of Chicago Press.

Kuroda, M., Mori, Y., Masaya, I. & Sakakihara, M. (2013). Alternating least squares in nonlinear principal components. *Computational Statistics*, 5(6), 456–464.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444.

Ling, J., Jones, R. & Templeton, J. (2016). Machine learning strategies for systems with invariance properties. *Journal of Computational Physics*, 318, 22-35.

Linting, M., Meulman, J.J., Groenen, P.J. & Van Der Kooij, A.J. (2007). Nonlinear principal components analysis: Introduction and application. *Psychological Methods*, 12(3), 336.

Linting, M. & Van Der Kooij, A. (2012). Nonlinear principal components analysis with CATPCA: A tutorial. *Journal of Personality Assessment*, 94(1), 12–25.

MacDuffie, J.P. (2018). Response to Perkins and Murmann: Pay attention to what is and isn't unique about Tesla. *Management and Organization Review* , 14(3), 481-489.

Marcus, G. (2018). *Deep learning: a critical appraisal*. arXiv preprint arXiv:1801.00631.

Marcus, G. (2020). *The next decade in AI: four steps towards robust artificial intelligence*. arXiv preprint arXiv, 2002.06177.

Marcus, G. & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York, NY: Pantheon.

Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L. & Canhoto, A. (2018). Big data, big decisions: the impact of big data on board level decision-making. *Journal of Business Research*, 93, 67-78.

Midler, C. & Beaume, R. (2010). Project-based learning patterns for dominant design renewal: The case of Electric Vehicle. *International Journal of Project Management*, 28(2), 142-150.

Mitchell, T.M. (1997). *Machine learning*. New York, NY: McGraw-Hill.

Murmann, J.P. & Frenken, K. (2006). Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. *Research Policy*, 35(7), 925-952.

Ng, A. (2016). What artificial intelligence can and can't do right now. *Harvard Business Review*. https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now

Ng, A. (2022). *Supervised Machine Learning.* Machine Learning Specialization. Coursera. https://www.coursera.org.

Nonaka, I. & Takeuchi, H. (1995). *The knowledge-creating company: how Japanese companies create the dynamics of innovation*. Oxford, England: Oxford University Press.

Nylund, P. A., Brem, A. & Agarwal, N. (2021). Enabling technologies mitigating climate change: the role of dominant designs in environmental innovation ecosystems. *Technovation*, 102271.

Perkins, G. & Murmann, J. P. (2018). What does the success of Tesla mean for the future dynamics in the global automobile sector? *Management and Organization Review*, 14(3), 471-480.

Propp, V. (1968). *Morphology of the Folktale* (Scott, L. transl.). Austin, TX: University of Texas Press.

Reinertsen, D. & Smith, P.G. (1991). The strategist's role in shortening product development, *Journal of Business Strategy*, 4, 18-22.

Saint-Exupéry, A. (1939). *Terre des Hommes*. Paris, France: Gallimard.

Schneider, S. & Leyer, M. (2019). Me or information technology? adoption of artificial intelligence in the delegation of personal strategic decisions. *Management and Decision Economics*, 40(3), 223-231.

Sharif, P. & Heydari, H. (2013) An introduction to multi-player, multi-choice quantum games: Quantum minority games & kolkata restaurant problems. In: Aberjel F et al. (ed) *Econophysics of Systemic Risk and Network Dynamic*s (pp. 217-236). Milano, Italy: Springer.

Sharma, A.K., Goel, N., Rajput, J. & Bilal, M. (2020). An Intelligent model for predicting the sales of a product. *10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 341-345.

Simon, H.A. (1987). Making management decisions: the role of intuition and emotion. *Academy of Management Perspectives*, 1(1), 57-64.

Simon, H.A. (1996). *The Sciences of the Artificial*. Boston, MA: MIT Press.

Sorli, M., Sopelana, A., Taisch, M., Al-Shaab, A., Keat, J., Flores, M. & Martinez, L. (2010). Applying lean thinking concepts to new product development. *APMS International Conference Advances in Production Management Systems Book of Abstracts*, 50.

Sovacool, B. K., Rogge, J. C., Saleta, C. & Masterson-Cox, E. (2019). Transformative versus conservative automotive innovation styles: Contrasting the electric vehicle manufacturing strategies for the BMW i3 and Fiat 500e. *Environmental innovation and societal transitions*, 33, 45-60.

Spangler, W.E. (1991). The role of artificial intelligence in understanding the strategic decision-making process. *IEEE Transactions on Knowledge and Data Engineering*, 3(2), 149-159.

Srinivasan, R., Lilien, G.L. & Rangaswamy, A. (2006). The emergence of dominant designs. *Journal of Marketing*, 70(2), 1-17.

Sun Tzu. (1963). *The Art of War* (Griffith, S.B. transl.). Oxford, England: Oxford University Press.

Teixeira, A.C.R., da Silva, D.L., Neto, L.D.V.B.M., Diniz, A.S.A.C. & Sodré, J.R. (2015). A review on electric vehicles and their interaction with smart grids: the case of Brazil. *Clean Technologies and Environmental Policy*, 17(4), 841-857.

The Economist. (2020). Automobiles: roadblock. *The Economist*, Technology Quarterly, 13.

Tushman, M. L. & Murmann, J.P. (1998). Dominant designs, technology cycles, and organization outcomes. *Academy of Management Proceedings*, 1998 (1), A1-A33.

Ulrich, K. T., Eppinger, S. D. & Yang, M.C. (2020). *Product Design and Development*. New York, NY: Mc Graw-Hill.

Utterback, J. M., & Abernathy, W. J. (1975). A dynamic model of process and product innovation. *Omega*, 3(6), 639-656.

Utterback, J. M., & Suárez, F. F. (1993). Innovation, competition, and industry structure. *Research policy*, 22(1), 1-21.

Vapnik, V. (2019). Complete statistical theory of learning. *Automation and Remote Control*, 80(11), 1949-1975.

Vapnik, V. & Izmailov, R. (2019). Rethinking statistical learning theory: learning using statistical invariants. *Machine Learning*, 108, 381-423.

Vapnik, V. & Izmailov, R. (2020). Complete statistical theory of learning (learning using statistical invariants). *Proceedings of Machine Learning Research*, 128, 1-37.

Vincent, T.L. & Brown, J.S. (2005). *Evolutionary game theory, natural selection, and Darwinian dynamics*. Cambridge, England: Cambridge University Press.

Vincent, T.L., Vincent, T.L.S. & Cohen Y (2011). Darwinian dynamics and evolutionary game theory. *Journal of Biological Dynamics*, 5(3), 215-226.

Weber, J. (2009). *Automotive development processes: processes for successful customer-oriented vehicle development*. Berlin, Germany: Springer Science & Business Media.

Wigner, E.P. (1949). Invariance in physical theory. *Proceedings of the American Philosophical Society*, 93(7), 521-526.

Wigner, E.P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics*, 13, 1-14.

Wilson, D.F. (1999). Competitive marketing strategy in a volatile environment: theory, practice and research priorities. *Journal of Strategic Marketing*, 7(1), 19-40.

Woodruff, R.B. (1997). Customer value: the next source for competitive advantage. *Journal of the Academy of Marketing Science*, 25(2), 139-153.

Xiong, J., Zhao, S., Meng, Y., Xu, L. & Kim, S. Y. (2022). How latecomers catch up to build an energy-saving industry: The case of the Chinese electric vehicle industry 1995–2018. *Energy Policy*, 161, 112725.

Yoshimoto, K. & Hanyu, T. (2021). Nissan e-Power: 100% electric drive and its powertrain control. *IEEJ Journal of Industry Applications*, 20013024.

Yu, T. (2007). *Incorporating prior domain knowledge into inductive machine learning* (PhD dissertation, University of Technology Sidney). Recovered from: https://www.researchgate.net/publication/228723728_Incorporating_Prior_Domain_Knowledge_Into_Inductive_Machine_Learning

Yu, X., Sandhu, N.S., Yang, Z. & Zheng, M. (2020). Suitability of energy sources for automotive application—A review. *Applied Energy,* 271, 115169.

APPENDIX A: Electric cars in Brazil: an analysis of core green technologies and the transition process

Journal	Sustainability – Scopus CiteScore 2021 = 5.0 (86% percentile)
Authors	Charles Lincoln Kenji Yamamura; Harmi Takiya; Cláudia Aparecida Soares Machado; José Carlos Curvelo Santana; José Alberto Quintanilha; Fernando Tobal Berssaneti
Reference	Yamamura, C. L. K., Takiya, H., Machado, C. A. S., Santana, J. C. C., Quintanilha, J. A., & Berssaneti, F. T. (2022). Electric Cars in Brazil: An Analysis of Core Green Technologies and the Transition Process. *Sustainability*, *14*(10), 1-19.
DOI	https://doi.org/10.3390/su14106064

sustainability MDPI

Article

Electric Cars in Brazil: An Analysis of Core Green Technologies and the Transition Process

Charles Lincoln Kenji Yamamura [1,*], Harmi Takiya [1], Cláudia Aparecida Soares Machado [2], José Carlos Curvelo Santana [3], José Alberto Quintanilha [4] and Fernando Tobal Berssaneti [1]

[1] Department of Production Engineering, Polytechnic School, University of São Paulo, São Paulo 05508-010, SP, Brazil; harmi.harmi@gmail.com (H.T.); fernando.berssaneti@usp.br (F.T.B.)
[2] Department of Transportation Engineering, Polytechnic School, University of São Paulo, São Paulo 05508-070, SP, Brazil; claudia.machado@usp.br
[3] Department of Management Engineering, Federal University of ABC, Alameda das Universidades, Anchieta, São Bernardo do Campo 09606-045, SP, Brazil; jose.curvelo@ufabc.edu.br
[4] Institute of Energy and Environment, University of São Paulo, São Paulo 05508-010, SP, Brazil; jaquinta@usp.br
* Correspondence: charles.yamamura@usp.br

Abstract: This paper explores the transition to electric cars in Brazil. The country has been successful to reduce its carbon footprint using biofuels, but it is facing a dilemma in vehicle electrification. It cannot shift abruptly to battery electric vehicles, as current consumers are unable to afford them and investment in recharging infrastructure is uncertain. However, it has a significant manufacturing base, and it cannot isolate itself from global industrial trends. This study relies on the inductive case study method, identifying the core green technologies in vehicle electrification and extrapolating their trends, to explain how the transition process is feasible. The emergence of a dominant design (set of core technologies defining a product category and adopted by the majority of players in the market) in small and affordable segments is essential for the diffusion of electric cars in developing countries. Biofuel hybrid technologies may support the transition. The Brazilian industry can engage in electric vehicle development by designing small cars based on global architectures, targeting consumers in emerging markets. The article contributes by using a dominant design core technologies framework to explain and map the transition to electric vehicles in developing countries, supporting academic research, government, and industry planning.

Keywords: electric car; technology transition; dominant design; vehicle electrification; clean energy; materials usage; vehicle battery; hybrid car; developing countries; Brazil

check for updates

Citation: Yamamura, C.L.K.; Takiya, H.; Machado, C.A.S.; Santana, J.C.C.; Quintanilha, J.A.; Berssaneti, F.T. Electric Cars in Brazil: An Analysis of Core Green Technologies and the Transition Process. *Sustainability* **2022**, *14*, 6064. https://doi.org/10.3390/su14106064

Academic Editor: Domenico Mazzeo

Received: 15 April 2022
Accepted: 13 May 2022
Published: 17 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Climate change is one of the great challenges of our time. The Paris Agreement—an international treaty signed in 2015 by 196 states and the European Union—aims to limit global warming to below 2 and preferably 1.5 degrees Celsius compared to pre-industrial levels, to mitigate the harmful effects of climate change [1]. Human-generated greenhouse gas (GHG) emissions must be reduced to close to zero to achieve that objective.

As the transport sector is one of the major global greenhouse gas emitters (21%) [2], a group of national governments, municipalities and regional governments, automobile manufacturers, and many other institutions and businesses signed a declaration at COP26—the 2021 United Nations Climate Change Conference—to transition to zero tailpipe emission vehicles by 2035 in leading countries and no later than 2040 in emerging countries [3]. Among the signatories were the UK, Sweden, Canada, Mexico, Chile, Ford, General Motors, Mercedes-Benz, Volvo, and Jaguar Land Rover. However, there were abstentions from the United States, China, Germany, Brazil, Volkswagen, Toyota, and BMW, among several others. Completely abandoning internal combustion engines is still a significant

Electric Cars in Brazil: An Analysis of Core Green Technologies and the Transition Process

Abstract: This paper explores the transition to electric cars in Brazil. The country has been successful to reduce its carbon footprint using biofuels, but it is facing a dilemma in vehicle electrification. It cannot shift abruptly to battery electric vehicles, as current consumers are unable to afford them and investment in recharging infrastructure is uncertain. But it has a significant manufacturing base, and it cannot isolate itself from global industrial trends. This study relies on the inductive case study method, identifying the core green technologies in vehicle electrification and extrapolating their trends, to explain how the transition process is feasible. The emergence of a dominant design (set of core technologies defining a product category and adopted by the majority of players in the market) in small and affordable segments is essential for the diffusion of electric cars in developing countries. Biofuel hybrid technologies may support the transition. The Brazilian industry can engage in electric vehicle development by designing small cars based on global architectures, targeting consumers in emerging markets. The article contributes by using a dominant design core technologies framework to explain and to map the transition to electric vehicles in developing countries, supporting academic research, government, and industry planning.

Keywords: electric car; technology transition; dominant design; vehicle electrification; clean energy; materials usage; vehicle battery; hybrid car; developing countries; Brazil

1. Introduction

Climate change is one of the great challenges of our time. The Paris Agreement – an international treaty signed in 2015 by 196 states and the European Union – aims to limit global warming to below 2 and preferably 1.5 degrees Celsius compared to pre-industrial levels, to mitigate the harmful effects of climate change [1]. Human generated greenhouse gas (GHG) emissions must be reduced to close to zero to achieve that objective.

As the transport sector is one of the major global greenhouse gas emitters (21%) [2], a group of national governments, municipalities and regional governments, automobile manufacturers and a many other institutions and businesses signed a declaration at COP26 – the 2021 United Nations Climate Change Conference – to transition to zero tailpipe emission vehicles by 2035 in leading countries and no later than 2040 in emerging countries [3]. Among the signatories were the U.K., Sweden, Canada, Mexico, Chile, Ford, General Motors, Mercedes-Benz, Volvo, and Jaguar Land Rover. But there were abstentions from the United States, China, Germany, Brazil, Volkswagen, Toyota, and BMW, among several others. Completely abandoning internal combustion engines is still a significant challenge in many parts of the world, considering technical, economic, social, political, and geographic factors.

1.1. Decarbonization of light vehicles in Brazil

Brazil has a successful history of using biofuels in transport. The 1973 Oil Crisis motivated the creation of the National Alcohol Program (Proalcool), establishing ethanol as an alternative to oil derived fuels. The program contributed to reduce air pollution by replacing lead as an anti-knocking agent, reducing carbon monoxide and hydrocarbon emissions [4]. Biofuels are considered carbon neutral on exhaust emissions as the emitted CO_2 has been previously captured from the atmosphere by photosynthesis. However, there are greenhouse gas emissions in fuel production and transportation. The first Brazilian ethanol powered automobile was launched in 1979.

Automotive gasoline in Brazil is a blend of 27.5% anhydrous ethanol (E27). Ethanol for vehicles (E100) contains up to 4.5% of water. Flex fuel engines can run on any mixture between E27 and E100. There are about 44 million active light vehicles in the domestic fleet and 74% of them are flex fuel [5]. It is estimated 70% of flex fuel vehicles run on gasoline. Ethanol consumption is influenced mainly by the ethanol to gasoline price ratio. Ethanol prices are affected by weather, government policies, international sugar price, crude oil price, and transport costs [6].

Out of 1.98 million new light vehicles registered in 2021 (79% passenger cars and 21% light commercial vehicles), 84% were fitted with flex fuel engines, 3% with gasoline engines, and 13% were diesel powered [5]. Only 2,851 electric cars were sold in 2021, just 0.14% of light vehicle sales [7]. In 2020, 85% of electric power in Brazil was generated from renewable sources (figure 1), led by hydropower 63.8%, followed by wind generation 9.2%, biomass 9.0%, and solar energy 1.7% [8].

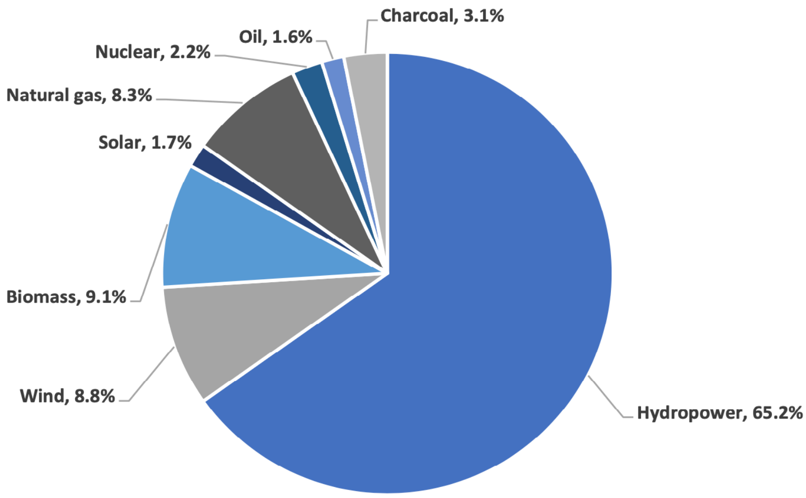


Figure 1. Brazilian sources of electricity in 2020. Source: [8].

The Vehicle Emissions Control Program (PROCONVE) started in 1986, and it progressively reduces new vehicle emission targets [9]. Although it sets targets for carbon monoxide, nitrogen oxides, hydrocarbons, soot, aldehydes, and sulfur oxides, it does not establish CO_2 limits directly. New phases in 2022 and 2025 will introduce progressively more stringent limits on non-methane organic gas and nitrogen oxides, both ground level ozone forming substances [4, 10]. Although there is no current legislation explicitly mentioning vehicle electrification as a route to energy efficiency and decarbonization in Brazil [9], the PROCONVE requirements will demand improvements in current engine technology and possibly the increase in the share of electrified vehicles [4].

Vehicle electrification is the transition from pure internal combustion engine vehicles (ICEVs) to full battery electric vehicles (BEVs), often with intermediate stages of electric hybridization – the combination of combustion engines and electric motors [11, 12, 13]. Electrification is inevitable for carbon neutrality in transports [4, 11, 12, 13, 14]. Besides climate damage, sticking to carbon fuels would isolate a country from the global industry, which would seriously affect its competitiveness and access to technology [11, 12, 15, 16]. But

Brazil is not ready to shift abruptly to pure battery electric vehicles, as most of its population would be unable to afford them, and the massive investment in infrastructure is beyond its current capacity [4, 11, 12, 17, 18].

1.2. Technology evolution and dominant design

Green innovations are technologies and practices that improve the quality of human life and reduce the impact on the environment. They minimize the usage of energy and materials, as well as reducing pollutant emissions and waste [19]. Incremental innovation (continuous change) occurs along an existing technology path, while radical innovation (discontinuous change) is related to the emergence of a new technology [20]. A new technology path usually starts with an *innovation shock*, a rupture from the existing technology [21].

The emergence of a dominant design is a landmark in the transition of technology from the stage to radical innovation to incremental innovation [22]. A dominant design is a set of product features that defines a product category and is widely adopted by the industry as a *de facto* standard competitors must adhere to [19, 20, 23]. The phase prior to the dominant design is called the era of *ferment* and characterized by discontinuous innovation, many competitors, intense experimentation, and high growth rates [22, 24, 25]. Electric cars are in the ferment stage of industrial evolution [20].

A dominant design marks the transition from focus on product innovation to process innovation [20]. Although the radical product innovation phase is over, there is an increase in incremental product variations. A key aspect of the emergence of a dominant design is the dramatic reduction in product costs [22, 24]. The increase in production volumes accelerates learning, standardization, and modularization of components [25, 26]. Prices fall and most potential consumers adopt the new product.

A systems view provides a better understanding of technology innovation and dominant designs [19, 26]. Dominant designs emerge not at the product system level but first in components or subsystems [27]. When dominant designs in the set of central or core subsystems consolidate, a system dominant design emerges. Core components are those that affect the largest number of product characteristics or features, i.e., they have many connections [19, 26]. Peripheral components, affect few characteristics and thus have fewer connections. The larger the number of connections in a product, the higher is its complexity (i.e., it has many variables) [19].

Architecture is the way components of a system are connected and organized [27]. A dominant design is a family of designs with common and stable core subsystems and architecture [28]. However, dominant designs are unlikely to be present in all components (subsystems). Once the core components of a design are settled, development shifts to peripheral components. Core components become invariants which are not revisited in a new design [27, 29], reducing the design space, and restricting variations to peripheral features [26]. A replacement of core components implies a change in the dominant design and a new technology.

Adner and Kapoor [30, 31] expanded the notion of technology evolution by adding the *ecosystems* dimension. Most technologies depend on complementary technologies to come to fruition, delivering value. An ecosystem is a community of multiple actors and activities aligned to create and to deliver value to customers. It includes producers, suppliers, competitors, distributors, customers, and other stakeholders. And adds *complementors*, like BEV recharging infrastructure, energy utilities, battery reutilization and recycling firms [32].

Beyond products, competition happens among ecosystems. Substitution depends on the capacity of a new technology to overcome its challenges, and on the existing technology to keep improving [33, 34, 35]. In *creative destruction*, the new technology overcomes its challenges quickly and the old technology is unable to catch up, being rapidly superseded. The *illusion of resilience* happens when the existing technology is unable to evolve, but it lives a bit longer because the new technology struggles to solve its challenges. However, it is a matter of time before the old technology is disrupted.

When a new technology faces significant entry barriers and the incumbent technology still has room for significant improvement, substitution tends to be slow, with *robust resilience*. Battery electric vehicles in emerging countries are such a case. There are considerable barriers for the dissemination of battery vehicles, and internal combustion engine vehicles can still be improved. However, if the new technology surmounts its difficulties quickly but the incumbent also improves vigorously, replacement is gradual, in a period of *robust coexistence.* The relation between pure internal combustion engine vehicles and hybrid vehicles is akin to that situation. The need to create a new ecosystem for battery vehicles can generate considerable tension and resistance. Hybrids may bridge the gap using existing manufacturing and fuel infrastructure (figure 2).

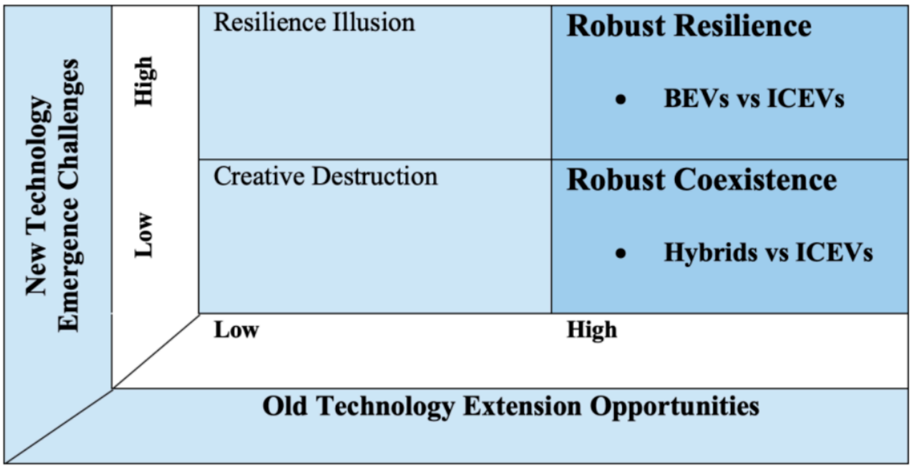


Figure 2. Technology substitution. Source: [34].

Electric vehicle technology depends on the development of batteries with enough energy storage and power delivery, vehicle design (electric motors, control systems, architecture), on recharging infrastructure, power supply from the grid, battery reutilization and recycling. The competition is not only between technologies but between the ecosystems supporting internal combustion vehicles (manufacturers, oil industry, biofuels industry, suppliers, dealers) and electric vehicles (manufacturers, battery makers, charging firms, power suppliers, battery reuse and recycling firms).

This paper posits large scale diffusion of electric cars in developing countries is unlikely to happen before the emergence of a dominant design for small passenger vehicles. Those vehicles should be affordable but practical, comfortable, and safe for small family usage, including enough driving range for holiday trips. The significant cost reduction necessary to make battery electric vehicles accessible to current automobile buyers can only be achieved with the economies of scale that follow the emergence of a dominant design [27]. The transition to electric cars promises more efficient use of energy and materials in the automobile industry. As the industry moves from fuel intensive to materials and energy intensive [36, 37], battery reutilization and recycling – still in early stage and with no established standards and procedures – will also be a key element in the transition.

This article originally contributes by using the dominant design core technology framework to analyze, explain, and map a feasible transition process to electric cars, applied to the specific case of a developing country. The study provides knowledge to support academic research, policy and decision making in both industry and government. The existing literature on vehicle electrification in Brazil and the theory of dominant designs was explored in this section. Materials and methods are discussed next. The core technologies for vehicle electrification and the convergence to dominant designs are explained after. From the analysis, relevant electrification technologies are framed, mapped, and discussed. The study is concluded and suggestions for future research are offered.

2. Materials and Methods

We rely on the inductive case study method, seeking relevant factors and the explanation of a complex phenomenon under the logic of theory [38]. The case of transition to electric vehicles in Brazil combines both extreme and typical elements of the vehicle electrification general case. Following Eisenhardt et al. [38], extreme cases provide a broad and clear perspective of a problem and can facilitate new insights. This study aims to understand the emergence of a dominant design in electric vehicles, their dissemination in developing countries, and a potential bridging process using hybrid technologies, to support strategy and policy making.

Our research draws on the academic literature on industrial evolution, product innovation, technology transition, and dominant designs. To investigate the Brazilian light vehicle electrification case, we relied on secondary data on energy policy, powertrain technologies, alternative fuels, vehicle electrification, electric cars, biofuels, clean energy, materials reuse, recycling, and value chains, from government agencies, research institutes, and think tanks, technical reports and articles from consulting firms and specialized press. From the analysis of those documents, we extracted the major trends in the core green technologies relevant to the emergence of electric cars – batteries, electric motors, control systems, architectures, charging infrastructure, battery reutilization and recycling. Next, we inferred the main implications for developing countries and for Brazil in particular, considering the biofuels experience. Then we draw a map of the technology transition, with ensuing explanations (Figure 3).

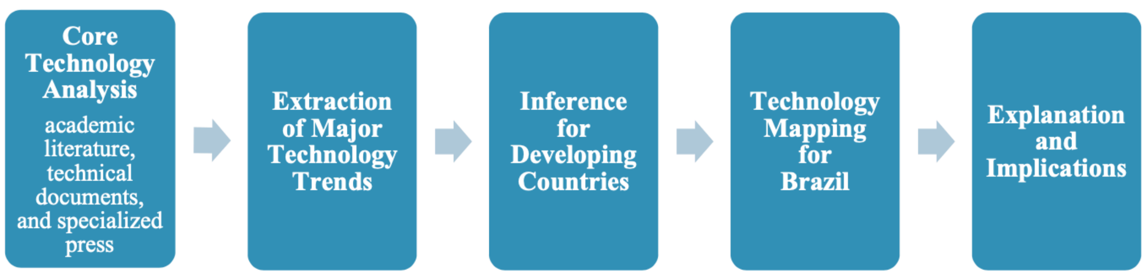


Figure 3. Summary of method. Source: the authors.

The most difficult task is not to understand what will happen but when and how it will happen. For instance, will the transition process be gradual or abrupt? Beyond the study of competing technologies, it is necessary to explore ecosystems – institutions, customers, manufacturers, suppliers, complementors, and infrastructure – in complex interactions, and their unfolding dynamics. Since this is an investigation on the emergence of dominant designs, following Murmann and Frenkel [26], we make explicit both the level of granularity and the unit of time of the study. We focus on the subsystem level (batteries, electric motors, control systems) to understand the integration into the overall vehicle architecture, at system level. This is an inquiry on the automobile evolution from a developing country's point-of-view, neither from an isolated firm's nor from the global industry perspective. The time scale of reference is in years.

This research refers to light-duty four-wheeled land vehicles, to the exclusion of other terrestrial transports such as heavy trucks, buses, trains, tractors, and motorcycles. The Brazilian National Environment Council (CONAMA) [39] defines two categories of light vehicles: light passenger vehicles and light commercial vehicles (LCVs). In Brazil, Diesel engines are used in heavy vehicles (trucks and buses) and light commercial vehicles but are not allowed in passenger cars. Light commercial vehicles are those: a) with payload over 1,000 kg; b) capable of carrying eight passengers or more, plus the driver; or c) with off-road characteristics. LCVs fall below heavy trucks and buses in capacity, which are also legally defined. Some four-wheel drive SUVs (sport utility vehicles) clear the CONAMA off-road vehicle criteria and are Diesel powered – as they are classified as LCVs – although they are typically used as private vehicles [10]. Although there is no legal definition or global consensus on the concept of SUV, we refer to closed bodywork passenger vehicles that are taller than traditional passenger cars, with higher ground clearance and elevated ride height [29].

3. Results

3.1. The emergence of a dominant design in electric cars

Electric cars are propelled by electric motors fed with energy from batteries, which are charged either from the power grid or by brake energy recovery [40]. Electric cars are praised for energy efficiency (about 90% compared to 25%~50% for combustion engines), low emissions, low operating costs (energy and maintenance), instant torque, smooth linear acceleration, and silent operation [41]. Among their weaknesses are high purchasing prices, limited real-world driving range, lack of charging infrastructure, long recharging times, and not having a charger at home [40, 41]. In emerging countries like Brazil, consumers need affordable vehicles that cover their full usage spectrum, not just daily commuting, including three or more annual holiday trips [16, 40].

The core subsystems in electric vehicles are power batteries, electric motors, and power control systems [26, 27, 42, 43, 44, 45]. BEVs are currently in full ferment mode of technology evolution and core subsystems are undergoing intense development, in wait for a dominant design to emerge (figure 4).

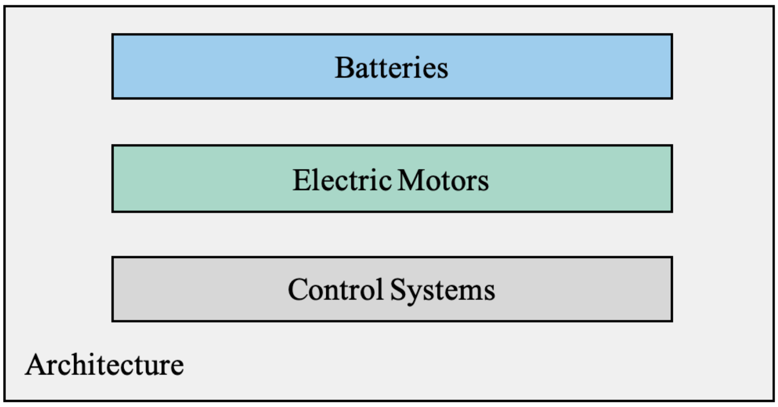


Figure 4. Dominant designs in core subsystems – integrated in the vehicle architecture – define a system dominant design. Sources: [26, 27, 42, 43, 44, 45].

3.1.1. Batteries

Lithium-ion batteries, with their high energy density, durability, and low self-discharge, were originally developed in the consumer electronics industry and made electric vehicles feasible [46]. The most common lithium-ion batteries are the NMC622-graphite type, with nickel, manganese, and cobalt present at 6:2:2 molar ratios in the cathode, and the anode made of graphite. Lithium ions flow between the electrodes through a liquid electrolyte and electrons run in the outer circuit, generating electricity. Nickel, manganese, and cobalt contents determine a battery's capacity, safety, and charge/discharge rate, respectively. NMC111-graphite batteries are no longer used. By 2030, NMC811-graphite is expected to be the common battery chemistry [46, 47].

NMC batteries held 71% of the EV battery market in 2020, nickel-manganese-aluminum (NMA) 25%, and lithium-iron-phosphate (LFP), less than 4% [2]. But LFP batteries are increasing their market share, due to their long life, thermal stability, performance, lower cost, and no use of cobalt – which is often sourced from strategically sensitive countries [48, 49, 50]. LFP batteries are about 25% less energy dense at cell level than NMC batteries. Solid-state lithium batteries – using solid instead of liquid electrolytes – promise significant improvements in cost, energy density, safety, packaging, and weight, but manufacturing at commercial scale is a challenge, unlikely to be surmounted before 2030 [46].

There is a lot of variation in battery chemistry, geometry, and thermal management systems [51]. There are battery cells in cylindrical, prismatic, and pouch geometric forms. Usually, battery cells are assembled in modules and then in battery packages. Battery packages need strong casings to protect them in collisions and to accommodate cooling systems and connectors [52]. Some recent LFP batteries skip modules and assemble cells directly in packages, reducing their weight and hence the disadvantage in specific energy compared to NMC batteries [49]. Structural architectures that eliminate package casings are also being developed [53]. Since battery technology is in ferment stage, the industry will need to find the right balance between standardization – to achieve economies of scale – and flexibility, making room for the fast pace of change and competition [52, 54].

Batteries can cost between thirty and fifty percent of an electric vehicle, depending on battery and vehicle size [41, 55]. In 2020, the average lithium-ion battery pack cost was $137 per kWh [2]. The MIT projects battery packs will still cost $124 by 2030 [46], as it foresees an increase in the cost of the minerals used to produce batteries. The current energy density of lithium-ion batteries is about 220~250 Wh/kg. Solid-state lithium batteries are expected to deliver 400 Wh/kg by 2030 [2].

Battery manufacturing is a logistics conundrum. Batteries are difficult and costly to transport in containers because of their bulk and weight. From the logistic point-of-view, battery production is better located close to vehicle manufacturing facilities [7, 56]. Local battery production – at least at package assembly level – is deemed essential for establishing a BEV industry [55].

3.1.2. Electric motors

Among the electric motors currently applied to vehicle propulsion are the permanent magnet synchronous motor (PMSM), AC asynchronous induction motor, and reluctance motor [43, 57]. Permanent magnet motors currently hold 90% of the EV market [9], due to their energy efficiency and power density. Besides the common radial flux type (magnetic flow perpendicular to the output shaft), the axial (parallel to the output shaft) flux permanent magnet motor is receiving attention recently, due to its higher torque density and compact packaging [58]. However, permanent magnet motors use rare earth metals in their magnets, raising concerns about both toxic extraction wastes and sourcing concentrated in few countries [9, 59]. Induction motors do not use permanent magnets (hence no need for rare earth metals), are less costly, but are also less efficient than permanent magnet synchronous motors, due to energy losses in their copper windings [9]. A reluctance motor uses imbalances produced in the magnetic fields of multiple mismatched poles between the stationary (stator) and rotating (rotor) parts of the motor to produce torque. It provides a good power density to cost ratio. But they can produce torque ripples and noise, and some applications use small permanent magnets in the rotor to mitigate those effects [43, 57].

Although some vehicle manufacturers currently outsource electric motors, there is a trend to design and to build motors in-house, as a source of efficiency, performance, and competitive advantage [59]. However, in-house development can delay the emergence of dominant designs [27].

3.1.3. Control systems

Control systems are devices and software that manage battery energy, to provide powerful, smooth, silent, and efficient operation of electric motors. Controllers achieve those tasks by varying the electricity voltage, shifting from direct current to alternating current, and changing the frequency of the alternating current [43]. Control systems are the brains of electric vehicles. Superior levels of system refinement and efficiency are achieved only with top notch expertise and painstaking development and testing. For instance, it took Nissan ten years to perfect the control of motors and batteries [25].

Masiero et al. [15] recommend Brazilian suppliers engage in the development of control systems, which manage the power from the battery to modulate the electric motor operation. Control systems are a core BEV technology, and they tend to be centralized in company R&D headquarters. Although Brazil may have the competence, government policy and incentives would be needed to encourage global manufacturers to engage their subsidiaries, sharing the development of control systems.

3.1.4. Architectures

System architecture is the way subsystems are connected and integrated [27]. Battery electric vehicle architecture is intimately linked to battery packaging as this is the bulkiest and heaviest component in the vehicle. Incumbent vehicle manufacturers usually start the incursion in electric cars by adapting existing combustion engine vehicle architectures, to reduce capital expenditures, lower volume risks, and time to market. Those manufacturers also tend to adopt off-the-shelf modular components and sometimes build electric cars in assembly lines shared with conventional vehicles [60].

Dedicated architectures are expected to replace internal combustion engine vehicle derived platforms in the future, to improve vehicle packaging, weight, efficiency, and costs [60, 61]. Since electric motors are smaller than internal combustion powertrains (engine and transmission), wheels can be moved to the corners of the vehicle in native architectures, and spaces for transmission tunnels and fuel tank are eliminated, freeing the central floor area for batteries and cabin space. Dedicated architectures also improve vehicle dynamics, by locating the batteries in positions of low center of gravity.

3.1.5. Charging infrastructure and energy management

Despite 80%~85% of EV charging being done at home [46], an adequate public charging network is essential to meet user requirements [62]. In Brazil, there are only about 750 public EV charging stations [63] – most of them level 2 chargers with 7.4 kW and 22 kW of power. There are very few 100-kW level 3 fast chargers, and no connection standards [64]. Building the charging infrastructure is challenging in countries with large territories, like Brazil and India [16, 40, 65]. The cost of installing an EV supply equipment is in the \$30,000~\$80,000 range [62]. According to the Brazilian Automobile Industry Association (ANFAVEA), Brazil will need approximately 150,000 charging points by 2035 [18], requiring between \$ 4.5 and \$12 billions in investment. This will require a joint effort among vehicle manufacturers, companies interested in providing charging infrastructure, and the government. Considering the history of limited investment capacity from the government, the infrastructure is going to take considerably longer than in leading countries [17].

A smart grid incorporates sensing and monitoring technologies to the power network, allowing the bidirectional flow of both energy and information [40]. It is an important element in the integration of electric vehicles, as it provides energy supply and demand management and can help alleviate the demand for grid expansion [40, 64]. The increase in EV power demand generates the risk of system overload – power fluctuations, service degradation, and even blackouts [6, 66, 67]. Studies indicate there is an overlap of EV charging and residence peak loads between two and six o'clock PM. Machine learning methods are being developed to improve charging network management, optimizing the vehicle energy demand side [66, 67].

3.1.6. Battery echelon utilization and recycling

Batteries no longer fit for use in electric vehicles can be reused in less demanding applications like powering residences and commercial buildings, as they retain approximately 80% of the original energy density. The reuse of retired batteries from electric vehicles in other applications is known as *battery echelon utilization* [68]. Among the benefits of echelon utilization are extended battery service life, energy efficiency, economic rents, and reduction of environmental impacts.

Since echelon utilization is an intermediate solution, those batteries will still need to be eventually disposed and recycled [46, 69, 70]. About 80%~85% of weight content in an internal combustion vehicle is currently recycled [71]. Since 30% to 50% of an electric vehicle weight is in batteries, adequate battery recycling and disposal are deciding factors to achieve similar indices in battery vehicles. It is an arduous task, as there is a diversity of battery chemistries and formats, and there is no established recycling procedure – most current batteries are not even designed with recycling in mind [46, 69, 72].

A lithium-ion battery is composed of several recyclable materials (table 1) [73], and the most valuable ones – like nickel and cobalt – are in the cathode. 100% of lithium, nickel, manganese, and cobalt, and 90% of aluminum, copper and plastics in a battery can be recycled. It is estimated the world will need 250,000~450,000 t of lithium, 1.3~2.4 million ton of nickel, and 250,000~420,000 t of cobalt to produce batteries in 2030. Although known reserves are sufficient to meet demand for metals in batteries, temporary shortages and prices increases are expected, caused by fluctuations in demand, and exporting issues, as extraction of minerals like lithium and cobalt is highly concentrated in a few countries [36, 73].

Table 1. Lithium-ion battery materials (% of weight at package level). Source: adapted from [73].

Aluminum	32%
Graphite	18%
Nickel	10%
Electrolyte	9%
Copper	6%
Plastic	5%
Manganese	3%
Cobalt	2%
Electronics	2%
Lithium	2%
Steel	1%
Residual	10%
	100%

Among the methods used to separate materials in battery recycling are melting (pyrometallurgy) and dissolving them with acids (hydrometallurgy). Direct recycling recovers battery cathodes using mechanical and chemical processes, without breaking them down into primary materials. It is a promising method, but still in early development stage [10, 46, 69, 70]. Battery recycling is complex, can be energy intensive, may emit GHGs, and may be hard to be compete economically with raw materials mining [69, 72].

Battery technology is in full ferment mode but reuse and recycling are in infant stages. It is important battery and vehicle manufacturers incorporate echelon utilization and recycling at design stage, making it easier to identify and to separate battery components and materials. Lithium-ion will continue to be the battery chemistry of choice for at least the next ten years [73]. But future batteries may require different metals in the cathode, making demand for those materials uncertain ten to fifteen years ahead [37, 72]. To find a sweet spot between standardization and freedom of innovation is a major ordeal [56, 58]. Machine learning techniques are being developed to improve battery echelon utilization, increasing service life, energy efficiency, and environmental benefits [68].

3.2. Hybrid transition

Existing firms challenged by radical innovations sometimes choose *hybrid* strategies to fill the gap between traditional and new technologies [74]. Hybrids contain features of the emerging innovation combined with others from the existing technology. For instance, hybrid cars combine electric motors and internal combustion engines. However, the effect of hybrids is controversial, and they are sometimes disliked as inelegant technology adaptations. But they can be useful to learn and to bridge transitions under the right circumstances [74].

It is imperious to understand why a hybrid technology is being adopted [75]. Hybrid strategies are employed to learn, to shape, to buy time, or to prevent a new technology from taking hold. In most cases, they are temporary and a major risk is sticking to a hybrid for too long, hoping it will be a permanent solution.

Biofuel hybrid electric vehicles (HEVs) are a compelling way to bridge the transition to battery vehicles in countries like Brazil. They are means to learn some elements of the EV technology, to bridge the transition while the new technology and infrastructure is not yet economically feasible, and to shape the transition process, accommodating the needs of both market and industry. But they should not be a way to block the transition [15].

Hybrid electrification is a significant step forward in energy efficiency, compared to the conventional internal combustion engine, while still employing most of the existing product and production competencies. They can smooth the transition until a battery vehicle dominant design emerges and complementary technologies – charging infrastructure, battery reuse and recycling, power grid etc. – are ready. Hybrids can be used to understand EV technology, value chain, distribution, and marketing [75].

3.2.1. Hybrid electric vehicles (HEVs)

A hybrid vehicle combines an internal combustion engine with at least one electric motor powered by a battery. Electrification improves the efficiency of an internal combustion engine vehicle by recovering energy from braking and storing it in batteries, to assist engine start and acceleration [47]. The combination of combustion engines and electric motors in hybrids broadens the optimal operation range of speed and loads, reducing both energy usage and emissions [41]. The battery is charged either by brake energy recovery or by a generator driven by the combustion engine. If the combustion engine propels the vehicle with mechanical link to the wheels, it is a *parallel* hybrid. In a *series* hybrid, the combustion engine works just as a battery recharger and it is not connected directly to the wheels [40, 76].

Plug-in hybrid electric vehicles (PHEVs) can charge the batteries from the grid, besides being charged by the combustion engine or by brake regeneration [40]. Their batteries are usually larger than in non-plug-in hybrids and they can operate as pure battery electric vehicles for short distances [47]. Plug-in hybrid electric vehicles can also have parallel or series configurations. A series PHEV is also called a *range extender electric vehicle* (REEV), as the combustion engine supplements energy from the power grid for range increase, operating as an energy generator only, without driving the vehicle wheels [40].

Parallel hybrids usually have larger combustion engines than series hybrids. Conversely, series hybrids tend to have larger electric motors than parallel hybrids. In a parallel hybrid, the main source of propulsion is still the combustion engine, while in the series hybrid, the electric motor is supplemented by a combustion engine. Parallel hybrids tend be more efficient to operate at high speeds in highways and series hybrids are cleaner and smoother in urban environments [77].

3.2.2. Fuel cell electric vehicles (FCEVs)

A fuel cell electric vehicle can be understood as a range extender vehicle (series hybrid) that uses a fuel cell instead of a combustion engine to power the electric motor. It uses electrical energy generated by the chemical reaction between hydrogen and oxygen (from the air) in fuel cells, charging batteries and powering electric motors that propel the vehicle [40]. The most common fuel cell in vehicles is the proton exchange membrane (PEM) type, which uses hydrogen from fuel tanks. The ethanol powered solid oxide fuel cell electric vehicle (SOFCEV) extracts hydrogen from ethanol using a device called *reformer*, eliminating the need for hydrogen production and supply infrastructure [78]. The electrodes in solid oxide fuel cells are separated by a rigid oxygen ion conducting ceramic membrane [79]. There are no hydrogen tanks and no connection to the power grid in solid oxide fuel cell vehicles. Batteries are charged by the fuel cells and are smaller than in pure electric car batteries, making solid oxide fuel cell vehicles potentially more cost competitive [79].

If current development obstacles – durability and reliability – are overcome, ethanol fuel cell vehicles may reach the market by 2030 [79]. However, direct hydrogen proton exchange fuel cell vehicles (PEMFCEVs) are uncompetitive in Brazil. Direct hydrogen fuel cell vehicles are considerably more expensive than battery electric vehicles, due to the need of precious metals – e.g., platinum – to catalyze reactions (precious metal catalyzers are not needed in solid oxide fuel cells because of their higher operating temperatures). Hydrogen production and distribution are even more complex, challenging, and expensive than the battery charging infrastructure, and Brazil is unlikely to mobilize in that direction [79, 80].

4. Discussion

In isolation, neither market nor technology can explain the emergence of a dominant design. The selection process is strongly influenced by political and social dynamics, making prediction of the exact shape of a dominant design difficult, if not impossible [22, 24]. However, it is possible to understand the evolution of the core technologies and the dominant design emergence process, to estimate both its probability and timing [81]. It is important to evaluate the capabilities and resources needed to accomplish the transition.

The hybrid strategy life cycle must be mapped to understand the technology transition, remembering most hybrids are stopgap solutions and resisting the temptation to stick to them for too long. Firms (and countries) that strive to learn and to embrace the future are more successful than those that are recalcitrant [75]. Table 2 summarizes major facts and events affecting the transition to battery vehicles.

Table 2. Main events affecting the transition to electric cars in Brazil.

Event	References	Time Estimate
COP26, zero tailpipe emissions declaration	[3,65,82]	2035/2040
Dominant design emergence, from the study of patents	[19,20,27]	2030
Dominant design emergence, from the automobile history pattern	[15,24,83]	2032–2050
Battery high energy density (~400 Wh/kg)	[2,46,84]	2030
Battery cost competitiveness	[46,84]	2035–2040
Vehicle acquisition cost parity	[17,64]	2035–2040
Adequate charging infrastructure in developing countries	[4,14,17,18]	2035
Emission regulations impact on pure gasoline and diesel	[4,10,18]	2035
Zero emissions traffic in urban perimeters	[4,77]	2040
Global industry ceasing internal combustion engine production	[3,4,61,65,82]	2045
Ethanol fuel cell electric vehicle maturity	[8,85]	2030–2035
Small pure electric car diffusion in developing countries	[17,19,83,86]	2040–2050

4.1. Lessons from history

The competition among battery and combustion engine cars is not new. It happened at the beginning of the automobile history, in late nineteenth and early twentieth century [43, 83, 87]. Electric cars lost because batteries were unable to provide adequate driving range, recharging infrastructure was more complex than supplying liquid fuel, and oil was cheap (in the United States), electric cars were more expensive, and electricity was not even available in certain regions – especially in rural areas [61].

The introduction of the steel closed body in the 1920's increased vehicle comfort, room, practicality, and safety, creating the concept of a *touring car*, and the aspiration for motorized traveling on holidays [43, 83]. The touring car was largely responsible for the victory of gasoline cars. Touring was not possible with early twentieth century batteries, due to range and lack of infrastructure. In 1911, Charles Kettering invented the electric starter, eliminating the dangerous operation of hand cranking to start a combustion engine, and removing a major disadvantage of the internal combustion engine car [43]. From that event, battery technology became subordinate to the combustion engine and used to engine start, ignition, and lighting [83].

The Ford Model T, made from 1908 to 1927, introduced mass personal motorization to both rural and urban populations, and eventually set the dominant design for automobiles [43, 83]. The automobile was no longer a toy for rich customers. The model T was comfortable, useful, versatile, reliable, safe, reasonably powerful, and efficient – a respectable automobile that could be used for both commuting and travelling – and still affordable [15, 83]. At the end of its life, it incorporated the core elements that defined the dominant design – gasoline powered engine, steel closed body, steering wheel (some cars were steered using a tiller), electric starter, and electric lighting [83].

From the study of 2.6 million patents in a vast range of industries, Brem et al. [20] concluded dominant designs take between fifteen and twenty years on average (the mode was eighteen years) to emerge from their first applications. It took twenty-two years, between the first internal combustion engine car (the 1886 Benz Patent-Motorwagen) and the 1908 Ford Model T [43]. But the Model T incorporated the whole set of core elements – like closed steel body and electric starter – only eighteen years later, in 1926 [83].

New attempts to make the electric car viable happened in the 1970's, 1990's and early 2000s [59]. After the 2009 great recession, electric cars finally gained traction, in no small degree due to achievements in battery technology – lithium-ion chemistry was developed by the consumer electronics sector [87]. The first modern era series production BEV was the Nissan Leaf, launched in 2010 [25]. Tesla launched the model S in 2012 [43].

Innovations are driven not only by cost and performance, but from emotional factors such as status and luxury [17, 88]. Some electric vehicle trials of the past, like the Norwegian Th!nk (2008-2012), did not take emotional factors into consideration and failed. In contrast, Tesla vehicles – largely responsible for the widespread interest in electric cars – sell on attributes like performance and style, while still making their owners feel both intelligent and good about themselves, due to smaller carbon footprints than in combustion vehicles [88, 89]. Products like automobiles and smartphones create experiences that establish emotional connections with people's ethos and culture [83]. It is not possible to change behavior solely with technology and policy, without consumers' acceptance. Research shows the aspiration for personal transportation will not vanish [55, 83, 86]. Electric vehicles will need to have the same basic capacities as combustion engine vehicles, like range, cost, performance, comfort, space, safety, and convenience [83].

An intriguing assessment from the patent study by Brem et al. [20] is, once a dominant design is established, it lasts in the original form for a maximum of only six years. The increasing acceleration of technology change in some industries may make the emergence of a dominant design ever more difficult and, when it happens, shorten their life spans.

4.2. Stricter emission regulations in Brazil

Emission regulations in Brazil (and other emerging countries) will be ever more rigorous, following benchmarks in developed countries, albeit with some delay [18]. Progressively more stringent emission limits on greenhouse gases and other pollutants will make gasoline, diesel, and even flex fuel vehicles struggle to comply and increase the need for electrification [4, 10]. Non-flex gasoline and diesel-powered light vehicles may be phased out by the middle of next decade, with flex fuel and hybrid vehicles remaining for a little longer.

Although ethanol powered combustion engine vehicles present low carbon emissions, they release other air pollutants (CO, NO_2, C_XH_Y, particulate matter). Legislation in municipalities may forbid using combustion engines within urban perimeters, allowing traffic in electric mode only [4, 77]. This will effectively ban internal combustion engine vehicles and conventional hybrids around 2040. Plug in hybrid vehicles will be electronically controlled to operate in local zero emissions mode in urban areas and be able to fire combustion engines only outside city boundaries. To meet consumer needs, electric mode range must increase from current 30~50 km to about 90 km to be practical [77].

4.3. The global automotive industry

Some of world's major automobile manufacturers have announced plans to stop selling internal combustion engine light vehicles between 2030 and 2040, especially in leading countries [3, 4, 61, 65]. Although they will continue to offer combustion engines in emerging markets for a few additional years, their demise (including hybrids) is most likely by midcentury. Although electric car adoption is a challenge in countries with large territorial extension, limited purchasing power and modest charging infrastructure, battery technology is expected to mature in power, energy content, convenience, and cost to replace other types of energy by that time.

All large automobile manufacturers in Brazil are either subsidiaries of transnational companies or local companies that license foreign technology. Most core products are designed in engineering centers abroad from global architectures. A few existing derivatives, like small SUVs, pickups, and light cargo vans are locally developed, usually based on global architectures. Most incumbent global manufacturers currently adapt battery vehicles from existing combustion engine vehicle architectures, to reduce investments, risk, and time to market. In the future, they will be developed from dedicated architectures and R&D activities tend to be even more centralized than they are today, seeking optimal returns on investment [87]. It will be hard for countries in the periphery to engage in the core technologies of the battery vehicle industry.

4.4. Developing countries

Competitive advantage is the set of skills, knowledge, and resources to create value that is superior to what competitors are offering [30, 83]. It is unlikely that two companies or countries starting at different points in time will achieve the same results in the battery electric vehicle industry. The United States, Europe, China, Japan, and Korea are likely to concentrate the technology as they have a significant head start and have been investing heavily in research and development. It will be difficult for emerging countries to catch up without government intervention. Access to the knowledge is likely to happen only after a dominant design is consolidated globally [17, 19, 86].

Brazil may consider the development of region-specific applications. One venue is the exploration of biofuel hybrids to bridge the transition to battery vehicles [15]. Another is to develop vehicles for emerging markets based on global battery vehicle architectures (small SUVs, light duty small pickup trucks, and cargo vans), as it currently happens in a few combustion engine vehicle cases. However, designing small and affordable electric cars with bona fide driving range, comfort, safety, and practicality attributes is challenging because of the cost, size, and weight of batteries.

Mass market electric cars will need to accommodate four people with luggage, offer a real-world driving range over 400km (about 250 miles), and to be priced around $20,000. Currently battery vehicles are almost three times more expensive than equivalent combustion engine vehicles [90]. For instance, an electric Peugeot e-208 (a small hatchback) costs R$265,900 (US$51,700; 16 March 2022 exchange rate) in Brazil – a comparably equipped combustion engine version costs R$90,990 (US$18,100) [91]. Electric car prices will need to decrease a lot to be affordable to current vehicle buyers [64]. Considering batteries may account for half the cost of a small car [41, 55], electric vehicles will take significantly longer than the 2030's (projected for the leading countries) to disseminate in emerging markets. In the meantime, hybrids will be needed to fill the gap (figure 6), helping to mitigate climate change.

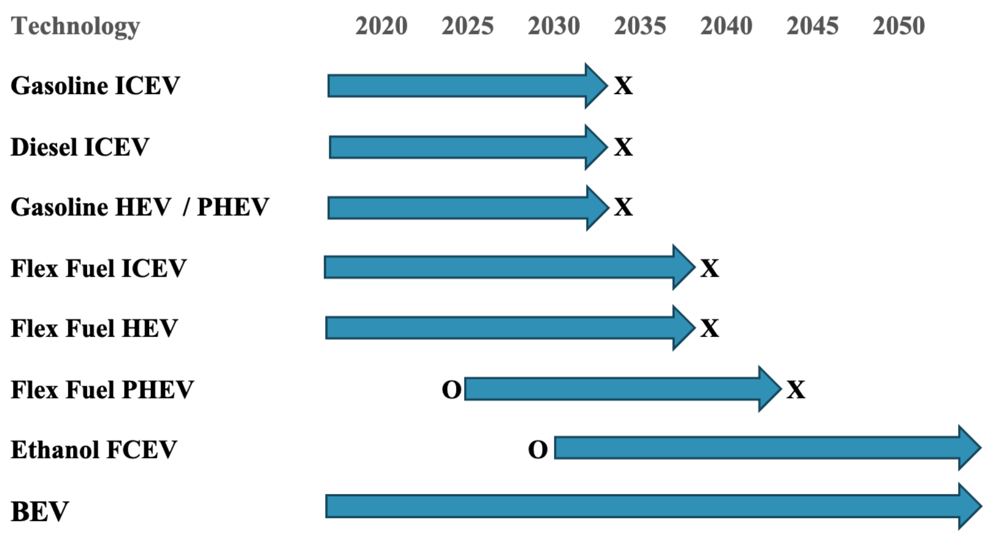


Figure 6. Plausible technology transition map of Brazilian light vehicles. Source: the authors.

5. Conclusions

The objective of this investigation was to analyze current policy, legislation, market, and technology developments in light vehicles, to explain the transition to electric cars in Brazil. It was accomplished with a case study, analysis and interpretation of both academic literature and secondary documents. Flex fuel hybrids can bridge the transition process from combustion engines to fully electric vehicles. Ever more stringent emission standards may demand hybrids to operate only in electric mode in urban settings. As batteries improve in both energy density and cost, and infrastructure is built, pure electric vehicles will eventually take over, but with a considerable delay compared to advanced economies.

Mass diffusion of electric cars in developing countries is unlikely to happen before the emergence of dominant designs in core vehicle technologies – batteries, electric motors, control systems, and architectural integration – and the advent of affordable and practical vehicles that can replace current vehicles in the full spectrum of needs and usages. Dominant designs are not necessarily performance optimal in all core components, but they are satisficing value propositions that accommodate technical, economic, and sociopolitical requirements. The success of a design is determined not by its technical performance but by cost. Brazil can support the development and manufacturing of relatively affordable light electric vehicles (small SUVs, light pickup trucks and cargo vans) based on global architectures, catering to both domestic and export emerging markets.

The transition involves both biofuels and gradual electrification, but it should not lose sight of electrification for the sake of biofuels. Although biofuels will bridge the transition, simultaneous development of both series hybrids and battery vehicles will be needed to build technical knowledge and competencies. Hybrids and parallel plug-in hybrids may be adapted from combustion engine vehicle platforms, but a battery vehicle is better deployed from a native architecture. The onboard use of biofuels to extract hydrogen and to power fuel cells is also being actively investigated and shows promise.

The emergence of dominant design may take about twenty years (if it emerges at all) but may last in its original form for only six years. That is approximately a single life cycle or generation of an automobile. It is also close to an individual product development cycle. It means once a dominant design emerges, its successor is already in gestation. Among the promising emerging core technologies are solid state batteries, axial motors, structural architectures, direct battery recycling, and machine learning technologies to optimize both charging network management and battery echelon utilization. The supply chain crisis triggered by the Covid 19 pandemic exacerbated some challenges such as strategic dependence on some materials sourcing for batteries (cobalt, lithium) and permanent magnets in electric motors (rare earth metals), as the industry moves from fuel intense to materials intense. The vehicle electrification process is extremely dynamic and hard to follow. The window of opportunity may be very narrow. Clear government policy, support, and investment are essential to develop the skills and resources needed to do a successful transition to electric cars.

Future research

Since the objective of innovation is to deliver value to customers, market acceptance is a natural mirror image of technology. Dominant designs may emerge in certain regional markets without converging to a global dominant design. It is equally plausible dominant designs emerge at specific market segment level, reflecting customers' economic and psychological profiles. As the global automobile industry transitions to battery electric vehicles, to understand how dominant designs are ramified and delineated in both geographic regions and market segments – and their impact on transport decarbonization – is an important and fertile field for further investigation.

Batteries are the heaviest and more expensive components in electric vehicles, and they need specific attention. Battery echelon utilization and recycling are in early experimental stages and achieving a balance between standardization and flexibility is a distant and uncertain goal. Despite intense ongoing research, it is still an issue waiting for answers, especially in developing countries.

References

1. UNFCC – United Nations Framework Convention on Climate Change. Paris Agreement, **2015**. Available online: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement (accessed on 10 June 2021).

2. IEA – International Energy Agency. Net zero by 2050 – a roadmap for the global energy sector, **2021**. Available online: https://iea.blob.core.windows.net/assets/7ebafc81-74ed-412b-9c60-5cc32c8396e4/NetZeroby2050-ARoadmapfortheGlobalEnergySector-SummaryforPolicyMakers_CORR.pdf (accessed on 18 May 2018).

3. U.K. Government. COP26 declaration on accelerating the transition to 100% zero emission cars and vans. Policy paper, 10 Nov **2021**. Available on line: https://www.gov.uk/government/publications/cop26-declaration-zero-emission-cars-and-vans (accessed on 11 Nov 2021).

4. Leal, A.C.B., Consoni, F.L. Emissões poluentes dos veículos: impacto dos combustíveis utilizados e potencialidades da mobilidade elétrica (pollutant emissions from vehicles: impact of fuels and potentialities of electric mobility). Brazilian Federal Senate, Legislative Consulting Studies and Research Center, discussion text No. 293, **2021**. Available online: https://www12.senado.leg.br/publicacoes/estudos-legislativos/tipos-de-estudos/textos-para-discussao/td293 (accessed on 30 June 2021).

5. ANFAVEA – Brazilian Automotive Industry Association. *Brazilian Automotive Industry Yearbook* **2022**. Available online: https://anfavea.com.br/anuario2022/2022.pdf (accessed on 25 March 2022).

6. OECD-FAO. Agricultural outlook 2020-2029, **2021**. Available online: https://www.oecd-ilibrary.org/agriculture-and-food/oecd-fao-agricultural-outlook-2020-2029_1112c23b-en (accessed on 29 May 2021).

7. ABVE – Brazilian Association of the Electric Vehicle. *Eletrificados batem todas as previsões em 2021* (electric cars exceed all predictions in 2021), January **2022**. Available online: http://www.abve.org.br/eletrificados-batem-todas-as-previsoes-em-2021/ (accessed on 24 January 2022).

8. EPE – Energy Research Company. *Brazilian Energy Balance, 2020.* Brazilian Ministry of Mines and Energy, **2021**. Available online: https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-601/topico-596/BEN2021.pdf (accessed on 22 July 2021).

9. Ferreira, A.L., Tsai, D.S., Boareto, R. *The Brazilian automotive industry transition*. IEMA (Institute for Energy and the Environment), **2021**. Available online: https://energiaeambiente.org.br/produto/the-brazilian-automotive-industry-transition.

10. ICCT - International Council on Clean Transportation. *Brazil PROCONVE L-7 and L-8 emission standards for light-duty vehicles,* **2020**. Available online: https://theicct.org/sites/default/files/publications/Brazil_L7L8_policy_update_01302020.pdf (accessed on 30 jan 2020).

11. Mello, A.M., Marx, R., Souza, A. Exploring scenarios for the possibility of developing design and production competencies of electrical vehicles in Brazil. *Int. J. Automot. Technol. Manag*. **2013**, *13*(3), 289-314. [CrossRef]

12. Marx, R., De Mello, A.M. New initiatives, trends and dilemmas for the Brazilian automotive industry: the case of Inovar Auto and its impacts on electromobility in Brazil. *Int. J. Automot. Technol. Manag***. 2014**, *21,14*(2), 138-157. [CrossRef]

13. Machado, C.A.S., Takiya, H., Yamamura, C.L.K., Quintanilha, J.A. Berssaneti, F.T. Placement of infrastructure for urban electromobility: a sustainable approach*. Sustainability* **2020**, *12*(16), 6324. [CrossRef]

14. Consoni, F.L., Oliveira, A., Barassa, E., Martinez, J., Marques, M.C., Bermudez, T. Estudo de governança e políticas públicas para veículos elétricos (study on electric vehicles governance and public policy). Bilateral technical cooperation project between the Brazilian Industrial Development and Competitiveness Secretariat and the German Sustainable Development Cooperation (PROMOB-e), **2018**. Available online: http://www.pnme.org.br/biblioteca/estudo-de-governanca-e-politicas-publicas-para-veiculos-eletricos/ (accessed on 27 Mar 2022).

15. Masiero, G., Ogasavara, M.H., Jussani, A.C., Risso, M.L. The global value chain of electric vehicles: A review of the Japanese, South Korean and Brazilian cases. *Renew. Sust. Energ. Rev*. **2017**, *80*, 290-296. [CrossRef]

16. Pompermayer, F. M. Etanol e veículos elétricos: via de mão única ou dupla? (Ethanol and electric vehicles: one or two-way street?). *Radar*, *7*, **2010**. Available online: https://www.ipea.gov.br/radar/temas/industria/308-radar-n-07-etanol-e-veiculos-eletricos-via-de-mao-unica-ou-dupla (accessed on 20 Mar 2022)

17. Costa, E., Horta, A., Correia, A., Seixas, J., Costa, G., Sperling, D. Diffusion of electric vehicles in Brazil from the stakeholders' perspective. *Int. J. Sustain. Transp.* **2020**, *15*(11), 865-878. [CrossRef]

18. ANFAVEA – Brazilian Automotive Industry Association. O caminho da descarbonização do setor automotivo no Brasil (the way to decarbonization in the Brazilian automotive sector), **2021**. Available online: https://anfavea.com.br/docs/apresentacoes/APRESENTAÇÃO-ANFAVEA-E-BCG.pdf (accessed on 25 November 2021)

19. Nylund, P. A., Brem, A., Agarwal, N. Enabling technologies mitigating climate change: the role of dominant designs in environmental innovation ecosystems. *Technovation* **2021**, 102271. [CrossRef]

20. Brem, A., Nylund, P.A., Schuster, G. Innovation and de facto standardization: The influence of dominant design on innovative performance, radical innovation, and process innovation. *Technovation* **2016**, *50*, 79-88. [CrossRef]

21. Argyres, N., Bigelow, L., Nickerson, J. A. Dominant designs, innovation shocks, and the follower's dilemma. *Strateg. Manag. J.* **2015**, *36*(2), 216-234. [CrossRef]

22. Chen, T., Qian, L., Narayanan, V. Battle on the wrong field? Entrant type, dominant designs, and technology exit. *Strateg Manag. J.* **2017,** *38*(13), 2579-2598. [CrossRef]

23. Brem, A., Nylund, P. A. Maneuvering the bumps in the new Silk Road: open innovation, technological complexity, dominant design, and the international impact of Chinese innovation. *R.D. Manag.* **2021,** *51*(3), 293-308. [CrossRef]

24. Anderson, P., Tushman, M. L. Technological discontinuities and dominant designs: A cyclical model of technological change. Adm. Sci. Q. **1990**, *35*(4), 604-633. [CrossRef]

25. Cecere, G., Corrocher, N., Battaglia, R. D. Innovation and competition in the smartphone industry: Is there a dominant design? *Telecommun Policy* **2015**, *39*(3-4), 162-175. [CrossRef]

26. Murmann, J.P., Frenken, K. Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. *Res. Policy* **2006**, *35*(7), 925-952. [CrossRef]

27. Brem, A., Nylund, P.A. Home bias in international innovation systems: The contingent role of central technologies in the emergence of dominant designs in the electric vehicle industry. *J. Clean. Prod.* **2021**, *321*, 128964. [CrossRef]

28. Tushman, M.L., Murmann, J.P.. Dominant Designs, Technology Cycles, and Organization Outcomes. *Acad. Manag. Proc*. **1998**, *1,* A1-A33.

29. Yamamura, C.L.K., Santana, J.C.C., Masiero, B.S., Quintanilha, J.A., Berssaneti, F.T. Forecasting new product demand using domain knowledge and machine learning. *Res. Technol. Manag*. **2022**, submitted.

30. Adner, R. Match your innovation strategy to your innovation ecosystem. *Harv. Bus. Rev.* **2006**, *84*(4), 98. [CrossRef]

31. Adner, R., Kapoor, R. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strat. Manag. J.* **2010**, *31*(3), 306-333. [CrossRef]

32. Adner, R., Euchner, J. Innovation ecosystems. *Res. Technol. Manag.* **2014,** *57*(6), 10-14. [CrossRef]

33. Adner, R., Kapoor, R. Innovation ecosystems and the pace of substitution: Re-examining technology S-curves. Strat. Manag. J. **2016**, *37*(4), 625-648. [CrossRef]

34. Adner, R., Kapoor, R. Right tech, wrong time. *Harv. Bus. Rev.* **2016**, *94*(11), 60-67. [CrossRef]

35. Adner, R. Ecosystem as structure: An actionable construct for strategy. *J. Manage.* **2017**, *43*(1), 39-58. [CrossRef]

36. IEA – International Energy Agency, 2021. The role of critical minerals in clean energy transitions, **2021**. Available online: https://www.iea.org/reports/the-role-of-critical-minerals-in-clean-energy-transitions (accessed on 10 April 2022).

37. Castelvecchi, D. Electric cars and batteries: how will the world produce enough? *Nature* **2021**. Available online: https://www.nature.com/articles/d41586-021-02222-1 (accessed on 17 Aug 2021)

38. Eisenhardt, K.M., Graebner, M.E., Sonenshein, S. Grand challenges and inductive methods: Rigor without rigor mortis. *Acad. Manage. J.* **2016**, *59*(4), 1113-1123. [CrossRef]

39. CONAMA – National Environment Council, 2018. Resolution No. 492, December 20, **2018**. Available online: https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/56643907/do1-2018-12-24-resolucao-n-492-de-20-de-dezembro-de-2018-56643731 (accessed on 27 Mar 2020).

40. Teixeira, A.C.R., da Silva, D.L., Neto, L.D.V.B.M., Diniz, A.S.A.C., Sodré, J.R. A review on electric vehicles and their interaction with smart grids: the case of Brazil. *Clean Tech. Environ*. **2015**, *17*(4), 841-857. [CrossRef]

41. Yu, X., Sandhu, N. S., Yang, Z., Zheng, M. Suitability of energy sources for automotive application – a review. *Appl. Energy* **2020**, *271*, 115169. [CrossRef]

42. MacDuffie, J.P. Response to Perkins and Murmann: Pay attention to what is and isn't unique about Tesla. *Manag. Organ. Rev.* **2018**, *14*(3), 481-489. [CrossRef]

43. Enge, P., Enge, N., Zoepf, S. *Electric Vehicle Engineering*. McGraw-Hill: New York, USA, 2021.

44. Yoshimoto, K., Hanyu, T. Nissan e-Power: 100% electric drive and its powertrain control. *IEEJ J. Ind. Appl.* **2021**, *10*(4), 411-416.. [CrossRef]

45. Xiong J., Zhao S., Meng Y., Xu L., Kim S. How latecomers catch up to build an energy-saving industry: the case of the Chinese electric vehicle industry 1995-2018. *Energy Policy* **2022**, *161*, 112725. [CrossRef]

46. MIT – Massachusetts Institute of Technology. *Insights into future mobility*. MIT Energy Initiative, *2019*. Available online: https://energy.mit.edu/wp-content/uploads/2019/11/Insights-into-Future-Mobility.pdf. (accessed on 20 March 2020)

47. Bieker, G. *A global comparison of the life-cycle greenhouse gas emissions of combustion engine and electric passenger cars*. The International Council on Clean Transportation, July **2021**. Available online: https://theicct.org/sites/default/files/publications/Global-LCA-passenger-cars-jul2021_0.pdf (accessed on 20 July 2021).

48. Xu, C., Dai, Q., Gaines, L., Hu, M., Tukker, A., Steubing, B. Future material demand for automotive lithium-based batteries. *Commun. Mater*. **2020**, *1*(1), 1-10. [CrossRef]

49. Yang, X.G., Liu, T., Wang, C.Y. Thermally modulated lithium iron phosphate batteries for mass-market electric vehicles. *Nat, Energy* **2021**, *6*(2), 176-185. [CrossRef]

50. Agathie, C. The main types of Li-Ion batteries explained and what is the best for electric vehicles. *Autoevolution*, **2022**. Available online: https://www.autoevolution.com/news/the-main-types-of-li-ion-batteries-explained-and-what-is-the-best-for-electric-vehicles-181498.html (accessed on 12 Feb 2022).

51. Arribas-Ibar, M., Nylund, P.A., Brem, A. The risk of dissolution of sustainable innovation ecosystems in times of crisis: the electric vehicle during the COVID-19 pandemic. *Sustainability* 2021, *13*(3), 1319. [CrossRef]

52. Arora, S., Shen, W., Kapoor, A. Review of mechanical design and strategic placement technique of a robust battery pack for electric vehicles. *Renew. Sust. Energ. Rev*. **2016**, *60*, 1319-1331. [CrossRef]

53. Tesla, 2020. Battery Day presentation. Sept. 20, **2020**. Available online: https://www.tesla.com/2020shareholdermeeting (accessed on 21 Sept 2020).

54. Brereton, P. Should we standardize electric vehicle batteries? *Electric and Hybrid Vehicle Technology International*, **2020**. Available online: https://www.electrichybridvehicletechnology.com/opinion/should-we-standardize-electric-vehicle-batteries.html (accessed on 7 May 2020).

55. Jetin, B. Who will control the electric vehicle market? *Int. J. Automot. Technol.* **2020**, *20*(2), 156-177. [CrossRef]

56. Mosquet, X., Arora, A., Xie, A., Renner, M. *Who Will Drive Electric Cars to the Tipping Point?* Boston Consulting Group, **2020**. Available online: https://www.bcg.com/en-br/publications/2020/drive-electric-cars-to-the-tipping-point (accessed on 2 Jan 2020).

57. Motor Trend. The green issue – the future of electric vehicles. *Motor Trend* **2021**, *June*, 10-71.

58. Jenkins, J. A closer look at axial flux motors. *Charged – Electric Vehicles Magazine*, **2021**. Available online: https://chargedevs.com/features/a-closer-look-at-axial-flux-motors/ (accessed on 19 May 2021).

59. Anderson, B. New electric motor technology will push EVs to the next level. Carscoops, **2022**. Available online: https://www.carscoops.com/2022/02/new-electric-motor-technology-will-push-evs-to-the-next-level/ (accessed on 3 Feb 2021).

60. Erriquez, M., Schäffer, P., Schwedhelm, D., Wu, T. *How to drive winning battery- electric-vehicle design: lessons from benchmarking ten Chinese models*. McKinsey & Company, **2020**. Available online: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-to-drive-winning-battery-electric-vehicle-design-lessons-from-benchmarking-ten-chinese-models (accessed on 10 July 2020).

61. Guan, M., Gao, P., Peng, B., Zhou, T., Hsu, A. *The race to win*: *How automakers can succeed in a post-pandemic China*. McKinsey China Auto Consumer Insights **2021**. Available online: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/the-race-to-win-how-automakers-can-succeed-in-a-post-pandemic-china (accessed on 13 Aug 2021).

62. Hussain, S., Kim, Y. S., Thakur, S., Breslin, J. G. Optimization of waiting time for electric vehicles using a fuzzy inference system. *IEEE T. Intell. Transp. Syst*. **2022**. [CrossRef]

63. ABVE – Brazilian Association of the Electric Vehicle. Rede de recarga aumenta 50% em quatro meses (recharging network grows by 50% in four months), **2021**. Available online: https://www.abve.org.br/eletropostos-no-brasil-crescem-50-em-quatro-meses/ (accessed on 20 March 2022).

64. Schiavo, F.T., Calili, R.F., de Magalhães, C.F., Fróes, I.C., 2021. The Meaning of Electric Cars in the Context of Sustainable Transition in Brazil. *Sustainability* **2021**, *13(*19), 11073. [CrossRef]

65. Singla, A., Bansal, R. Sustainability of electric vehicles: a short study of the Indian electric vehicle market. Shiv Nadar University, **2022**. Available online: https://www.researchgate.net/publication/357648886_SUSTAINABILITY_OF_ELECTRIC_VEHICLES_A_short_study_of_the_Indian_Electric_Vehicles_Market (accessed on 1 May 2022).

66. Hussain, S., Ahmed, M. A., Kim, Y. C. Efficient power management algorithm based on fuzzy logic inference for electric vehicles parking lot. *IEEE Access* **2019**, *7*, 65467-65485. [CrossRef]

67. Hussain, S., Ahmed, M. A., Lee, K. B., Kim, Y. C. Fuzzy logic weight based charging scheme for optimal distribution of charging power among electric vehicles in a parking lot. *Energies* **2020**, *13*(12), 3119. [CrossRef]

68. Li, C., Wang, N., Li, W., Li, Y., Zhang, J. Regrouping and echelon utilization of retired lithium-ion batteries based on a novel support vector clustering approach. *IEEE T. Transp. Electr.* **2022**. [CrossRef]

69. Beaudet, A., Larouche, F., Amouzegar, K., Bouchard, P., Zaghib, K. Key challenges and opportunities for recycling electric vehicle battery materials. *Sustainability* **2020**, 12(14), 5837. [CrossRef]

70. Parajuly, K., Ternald, D., Kuehr, R. *The future of electric vehicles and material resources: a foresight brief*. The United Nations University, **2020**. Available online: http://collections.unu.edu/eserv/UNU:7820/n2020_Future_of_Electric_Vehicles_Foresight_Brief.pdf (accessed on 11 April 2022).

71. Passos, E. R. Reciclagem de automóveis (Automobile recycling). Instituto Mauá de Tecnologia, **2013**. Available online: https://maua.br/files/monografias/completo-reciclagem-automoveis-161657.pdf (accessed on 9 April 2021).

72. Morse, I. A dead battery dilemma. *Science* **2021**, 373(6544), 780-783. [CrossRef]

73. Backhaus, R. Battery raw materials – where from and where to? *ATZ Worldw*. **2021**, *123*(9), 8-13. [CrossRef]

74. Christensen, C.M., McDonald, R., Altman, E.J., Palmer, J.E. Disruptive innovation: An intellectual history and directions for future research. *J. Manag. Stud*. **2018**, *55*(7), 1043-1078. [CrossRef]

75. Furr, N., Snow, D. The Prius Approach. *Harv. Bus. Rev*. Nov. **2015**, 102-109. [CrossRef]

76. Tran, M.K., Bhatti, A., Vrolyk, R., Wong, D., Panchal, S., Fowler, M., Fraser, R. A review of range extenders in battery electric vehicles: current progress and future perspectives. *World Electr. Veh. J.* **2021**, *12*(2), 54. [CrossRef]

77. Plötz, P., Moll, C., Li, Y., Bieker, G., Mock, P. *Real-world usage of plug-in hybrid electric vehicles: Fuel consumption, electric driving, and CO2 emissions*. International Council on Clean Transportation, **2020**. Available online: https://theicct.org/publications/phev-real-world-usage-sept2020 (accessed on 28 Aug 2021)

78. Silva, E.P. Etanol e hidrogênio: uma parceria de futuro para o Brasil (ethanol and hydrogen: a promising partnership for Brazil), **2021**. Available online: http://cienciaecultura.bvs.br/pdf/cic/v60n3/a15v60n3.pdf (accessed on 1 June 2021)

79. Vargas, J.E.V., Seabra, J.E.A. Fuel-cell technologies for private vehicles in Brazil: Environmental mirage or prospective romance? A comparative life cycle assessment of PEMFC and SOFC light-duty vehicles. *Sci. Total Environ*. **2021**, *798*, 149265. [CrossRef]

80. Plötz, P. Hydrogen technology is unlikely to play a major role in sustainable road transport. *Nat. Electron*. **2022**, *5*(1), 8-10. [CrossRef]

81. Srinivasan, R., Lilien, G.L., Rangaswamy, A. The emergence of dominant designs. *J. Mark*. **2006**, *70*(2), 1-17. [CrossRef]

82. Sacchi, R., Bauer, C., Cox, B., & Mutel, C. When, where and how can the electrification of passenger cars reduce greenhouse gas emissions? *Renew. Sust. Energ. Rev*. **2022**, *162*, 112475. [CrossRef]

83. Bladh, M. Origin of car enthusiasm and alternative paths in history. *Environ. Innov. Soc. Transit*. **2019**, *32*, 153-168. [CrossRef]

84. Rathi, A., Murray, P., Dottle, R. The Hidden Science Making Batteries Better, Cheaper and Everywhere. Bloomberg, **2021**. Available online at: https://www.bloomberg.com/graphics/2021-inside-lithium-ion-batteries/ (Acessed on 7 May 2022).

85. Dogdibegovic, E., Fukuyama, Y., Tucker, M.C. Ethanol internal reforming in solid oxide fuel cells: a path toward high performance metal-supported cells for vehicular applications. *J. Power Sources* **2020**, *449*, 227598. [CrossRef]

86. Fujimoto, T. The long tail of the auto industry life cycle. *J. Prod. Innov. Manage*. **2014**, *31*(1), 8-16. [CrossRef]

87. Midler, C., Beaume, R. Project-based learning patterns for dominant design renewal: The case of Electric Vehicle. *Int. J. Proj. Manag*. **2010**, *28*(2), 142-150. [CrossRef]

88. Sovacool, B.K., Axsen, J. Functional, symbolic and societal frames for automobility: Implications for sustainability transitions. *Transp. Res. A: Policy Pract*. **2018**, *118*, 730-746. [CrossRef]

89. Noel, L., Sovacool, B.K., Kester, J., de Rubens, G.Z. Conspicuous diffusion: Theorizing how status drives innovation in electric mobility. *Environ. Innov. Soc. Transit.* **2019**, *31*, 154-169. [CrossRef]

90. Arora, A., Niese, N., Dreyer, E., Waas, A., Xie, A. *Why electric cars can't come fast enough*. Boston Consulting Group, **2021**. Available online: https://www.bcg.com/pt-br/publications/2021/why-evs-need-to-accelerate-their-market-penetration. (accessed on 22 Dec 2021).

91. Peugeot. New Peugeot 208, **2022.** Available online: https://carros.peugeot.com.br/compre/tenha-um-peugeot/ofertas/ofertas-peugeot-208.html?gclid=Cj0KCQjw3IqSBhCoARIsAMBkTb1LeoNfPVu90uvViLVJl0puSlc9ndtWzj4wwxKUbESWTxKXW31kHT0aAub8EALw_wcB (accessed on 16 March 2022)

APPENDIX B: Forecasting new product demand using domain knowledge and machine learning

Journal	Research-Technology Management – Scopus CiteScore 2021 = 4.7 (80% percentile)
Authors	Charles Lincoln Kenji Yamamura; José Carlos Curvelo Santana; Bruno Sanches Masiero; José Alberto Quintanilha; Fernando Tobal Berssaneti
Reference	Yamamura, C. L. K., Santana, J. C. C., Masiero, B.S., Quintanilha, J. A., & Berssaneti, F. T. (2022). Forecasting New Product Demand Using Domain Knowledge and Machine Learning. *Research-Technology Management*
DOI	**https://doi.org/10.1080/08956308.2022.2062553**

FEATURE ARTICLE

Forecasting New Product Demand Using Domain Knowledge and Machine Learning

A proposed method uses machine learning and an expert's domain knowledge to enhance the accuracy of new product predictions.

Charles Lincoln Kenji Yamamura, José Carlos Curvelo Santana, Bruno Sanches Masiero, José Alberto Quintanilha, and Fernando Tobal Berssaneti

OVERVIEW: Forecasting demand for new products is a challenging task, as it involves capturing relations of complex variables in markets where little or no historical data exist. Managers usually rely on surveys, intuition, and heuristics to forecast new products. Linear statistical tools used to predict demand for existing products are not suitable because there are not enough data to capture complex nonlinear relations in yet-to-be launched products. Other tools are appropriate for aggregate new categories but not for incremental company-specific products. Machine learning can capture complex nonlinear relations, but it usually requires significant amounts of data. Using an expert's domain knowledge can circumvent the need for vast training datasets. To support product development activities, we propose a method that combines domain knowledge and machine learning to forecast market share of complex incremental new products. An experiment from the automobile industry shows the approach yields expressive results (82 percent forecast accuracy).

KEYWORDS: New product forecasting; Domain knowledge; Machine learning; Product concept; Artificial neural network

For the last 50 years, automobiles have experienced rapid incremental innovations—in energy efficiency, costs, safety, performance, comfort, convenience—that have kept them from experiencing an industry life-cycle decline (Fujimoto 2014). It takes three to five years to develop a new automobile (Ulrich, Eppinger, and Yang 2020), and it must remain competitive in the market for six to eight years after it is launched. Rather than reacting to customer demands,

Charles Yamamura has over 20 years' experience in strategic planning and product development in the automobile industry. He is a former chief portfolio strategist at Fiat Chrysler, technical director at JAC Motors, product planning manager at Nissan and Yamaha, marketing manager at Volkswagen, and product planner at Toyota. He is currently completing a PhD in industrial engineering at the Polytechnic School of the University of São Paulo, Brazil. He holds an MBA from Yale University and an MSc from the Getulio Vargas Foundation. His research focuses on vehicle electrification, machine learning applied to product strategy, innovation, and small data. charles.yamamura@usp.br

José Santana is a professor of industrial engineering at the Federal University of ABC in Brazil and a researcher at the Polytechnic School of the University of Sao Paulo, Brazil. He holds MSc and PhD degrees in chemical engineering from the State University of Campinas, and a bachelor's degree in industrial chemistry at the Federal University of Sergipe. He has experience in the fields of industrial and chemical engineering, focusing on biochemical and chemical processes, effluent treatments, sustainability development process simulation and optimization, statistical control of quality, factorial planning, accreditation of methods, quality assurance, product planning, and development. jccurvelo@yahoo.com.br

Bruno Masiero is a professor of electrical and computing engineering at the University of Campinas, Brazil. He has a bachelor's degree in electrical engineering and an MSc in electrical engineering from the University of São Paulo, and a PhD in electrical engineering from the RWTH Aachen University, Germany. His research focuses on the application of modern digital signal processing techniques in several audio and acoustic applications, such as spatial sound acquisition and reproduction, acoustic imaging, characterization of acoustic materials, and development of audiological assessment tools. He is a member of the board of the International Commission for Acoustics. masiero@unicamp.br

José Quintanilha is a senior associate professor at the Energy and Environment Institute of the University of São Paulo, Brazil. He has a bachelor's degree in statistics from the Institute of Mathematics and Statistics of the University of São Paulo, an MSc in remote sensing and geoprocessing from the National Institute of Space Research, and PhD in transportation engineering from the University of São Paulo. He has experience in geostatistical analysis, with emphasis on environment, city, and transport applications. jaquinta@usp.br

Fernando Berssaneti is a professor of industrial engineering at the Polytechnic School of the University of São Paulo, Brazil, and a counseling member of the Vanzolini Foundation. He holds a bachelor's degree, a MSc, and a PhD in industrial engineering from the Polytechnic School of the University of São Paulo. He is also a business consultant in industrial engineering, with emphasis on product and innovation management, mobility, and sustainability. He has several articles and books published in those fields. fernando.berssaneti@usp.br

DOI: 10.1080/08956308.2022.2062553

Forecasting New Product Demand Using Domain Knowledge and Machine Learning

Overview: Forecasting demand for new products is a challenging task, as it involves capturing relations of complex variables in markets where little or no historical data exist. Managers usually rely on surveys, intuition, and heuristics to forecast new products. Linear statistical tools used to predict demand for existing products are not suitable, as there is not enough data to capture complex nonlinear relations in yet-to-be launched products. Other tools are appropriate for aggregate new categories but not for incremental company-specific products. Machine learning can capture complex nonlinear relations, but it usually requires significant amounts of data. Using an expert's domain knowledge can circumvent the need for vast training datasets. To support product development activities, we propose a framework that combines domain knowledge and machine learning to forecast market share of complex incremental new products. An experiment from the automobile industry shows the approach yields expressive results (82 percent forecast accuracy).

Keywords: New product forecasting, Domain knowledge, Machine learning, Product concept, Artificial neural network

For the last 50 years, automobiles have experienced rapid incremental innovations—in energy efficiency, costs, safety, performance, comfort, convenience—that have kept them from experiencing an industry life-cycle decline (Fujimoto 2013). It takes three to five years to develop a new automobile (Ulrich, Eppinger, and Yang 2020), and it must remain competitive in the market for six to eight years after it is launched. Rather than reacting to customer demands, automotive companies must anticipate what customers want and do so usually with little or no historical data since the new product is not in the market yet (Kahn 2002).

Given the scarcity of both data and adequate analytical tools to forecast demand for yet-to-be developed products, managers usually rely on surveys, intuition, and heuristics to make decisions. The decision-maker's knowledge and perception of the business environment (customers, competitors, technology), and expected changes over time (Dane and Pratt 2007) drive intuition. Intuitive decision-making strives to extract meaning from available information, but that process is usually vague and subjective (Spangler 1991).

Among the existing analytical methods, linear models are easy to interpret and explain, but they rarely capture the complexity of real-world problems with accuracy (Shrestha et al. 2020). Some analytical approaches rely on diffusion theory (Linton 2002; Ching-Chin et al. 2010; Lee et al. 2014; Yin et al. 2020) from microeconomics. For instance, a Bass model contemplates the initial stages of new product adoption by a population (Linton 2002). Although diffusion models are applicable to aggregate product categories, they are not adequate for incremental company-specific products (Kahn 2002), which are the scope of this study.

Machine learning algorithms can capture patterns that are difficult to delineate mentally or heuristically. But those algorithms are data driven (Yu 2007; Yu et al. 2010) and require significantly large amounts of training data. Lack of sufficient data is a major challenge in data analytics (Ng 2016; Yang et al. 2019). This study proposes using the manager's domain knowledge to partially mitigate the lack of data for machine learning simulations (Vapnik and Izmailov 2019, 2020). Domain knowledge can accelerate the development of solutions.

Our study offers a framework practitioners can use to predict market share of new product concepts (sets of features), and thereby support and enhance new product development. Our method combines domain knowledge and machine learning, precluding the need for vast amounts of training data. The domain expert (manager or market specialist) synthesizes a database from his/her knowledge, combined with some competitor sales data. An artificial neural network processes the database, simulating the market by capturing the nonlinear relations among the variables, which are expressed in the weights of the connections in the stable machine learning algorithm (an artificial neural network models the market).

We demonstrate our method with an experiment—namely, to predict the market share of small SUVs (sport utility vehicles) in the Brazilian market. The experiment shows a machine learning algorithm is useful to support new product development. The trained artificial neural network captured the complex nonlinear relations in the variables, producing meaningful results: a forecast accuracy of 82 percent, which is to be compared to the standard accuracy of 58 percent achieved by most firms (Kahn 2002, 2010). We conducted an additional smaller scale experiment on medium-size pickup trucks. Machine learning usually requires tens of thousands of observations to capture patterns in data (Allen 2019). In our study, the combination of machine learning and domain knowledge was able to reduce data size requirements to a much smaller number of observations (Shrestha et al. 2020) while getting superior results. The SUV dataset contains 1,563 observations and we achieved a correlation factor (R) of 0.96 between predictions and real market share.

New Product Forecasting

A product is a bundle of features (attributes and characteristics) that customers perceive as having "value"—that is, which satisfy their needs and aspirations (Woodruff 1997; Ulrich, Eppinger, and Yang 2015). A product concept is a set of the most relevant customer features, as interpreted by a company. Effective product decision-making implies good forecasting (Weber 2009). Successful new product forecasting predicts the most likely outcome of a product concept given a set of assumptions (Kahn 2006).

Companies use statistical techniques, like time series and regression methods, to forecast existing products. By contrast, new product forecasting is predominantly judgmental and seek meaningfulness as the performance criterion, instead of statistical accuracy, because little or no historical data exist (Kahn 2014). The forecasting process for new products relies on surveys, managerial judgment, and sometimes on extrapolations made by assessing similarities with competitor products (Sharma et al. 2020). A meaningful forecast can be used for decision-making because it provides deeper understanding of the problem (Kahn 2002). Due to the lack of data, new product forecasting usually aims to produce realistic and operationally useful approximations in predictions, instead of numerical precision. Our method proposes using machine learning and domain knowledge to enhance the accuracy of new product predictions.

Kahn (2014) identifies seven possible new product categories: cost reduction, product improvement (replaces an existing product), product line extension (addition to the existing product line), new usage, new (geographical) market, "new-to-the-company" products, and "new-to-the-world" products. Products that are new to the world are also called radical products, as they disrupt existing industries, technologies, and customer perceptions of the market (Henderson and Clark 1990; Rice et al. 1998; Kahn 2018; Veryzer 1998; Leifer et al. 2001). We do not consider radical products in this analysis. The other six categories fall within incremental or continuous product innovation (Veryzer 1998). Note that product innovation requires the convergence of both technical capabilities and customer perceptions to succeed (Veryzer 1998; Leifer et al. 2001) and those modest technical innovations may sometimes produce significant competitive advantage (Henderson and Clark 1990).

Companies usually need to balance a mix of incremental and radical new products for a healthy business (Rice et al. 1998; Leifer et al. 2001; Veryzer 1998; Kahn 2018). Forecasting becomes harder and less accurate as products move from the incremental to radical scale, and applied techniques tend to shift from statistical to intuitive (Kahn 2014). Machine learning performs best when simulating relatively stable phenomena (Shrestha et al. 2020). Our proposed framework is not suitable to predict radical products, as no data exist to extract patterns and trends. But "new-to-the-world" products are rare (Ching-Chin 2010) and forecasting them is still an exercise of vision and brainstorming (Linton 2002).

From a survey of 168 industrial companies (Kahn 2002, 2014), average prediction accuracies are 72 percent (cost reduction), 65 percent (improvement), 63 percent (line extension), 54 percent (market extension), 47 percent (new-to-the-company), and 40 percent (new to the world). The average accuracy was 58 percent (Kahn 2002, 2010; Armstrong 2002), but an acceptable level should be around 76 percent (Kahn 2010; Ching-Chin et al. 2010).

Machine Learning

A machine learning algorithm is a computer program that improves its performance with data inputs, without being explicitly programmed (Mitchell 1997; Goodfellow, Bengio, and Courville 2016; Yu 2007; Shrestha et al. 2020). It distinguishes itself from conventional algorithms by its capacity to learn or to improve with new data. Most current achievements in machine learning are of the *supervised learning* type, where algorithms improve by processing *labeled* data (Ng 2018; Goodfellow, Bengio and Courville 2016).

An artificial neural network is a machine learning algorithm that maps functions of complex feature variables to target values in nonlinear hierarchical relations (Goodfellow, Bengio and Courville 2016; Aggarwal 2018). Machine learning methods with multiple levels of representation are also called *deep learning* methods (LeCun, Bengio and Hinton 2015). Data are processed by inputting feature data and adjusting the weights in the connections through layers of hidden nodes. The artificial neural network calculates errors by successive forward and backward iterations, with the goal of minimizing the derivatives of errors (figure 1). As data move into higher layers of representation, the algorithm processes more abstract relations through the network (Aggarwal 2018).

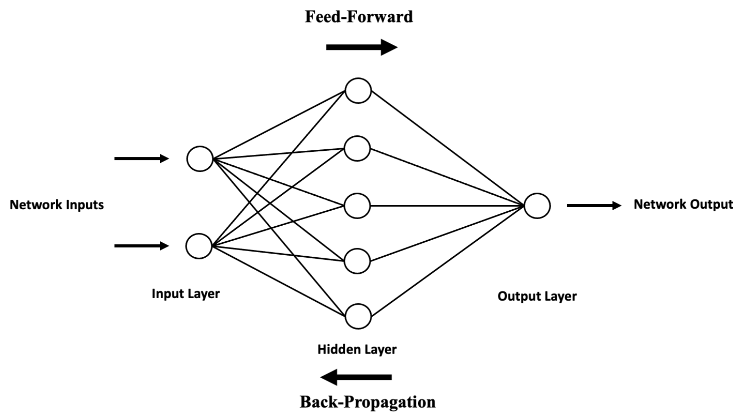


Figure 1. An artificial neural network

Domain Knowledge

People possess knowledge in the forms of information on facts, ideas, perceptions, and processes, that enable them to take effective action (Alavi and Leidner 2001). Domain knowledge is the knowledge pertaining to a particular specialized field, in opposition to general knowledge (Yu 2007). In this study, domain knowledge is the sum of the manager's experience and perceptions about customers, technologies, and competitors related to certain markets and products. Domain knowledge is usually informal and ill-structured (Yu 2007; Yu et al. 2010).

The initial conditions, assumptions, and context make the invariants of a problem (Gil et al. 2019). Invariants do not change when an object suffers a transformation (Aleksandrov, Kolmogorov, and Lavrent'ev 1999). In the product design context, invariants are the variables the expert takes for granted in a problem – they assume those variables have fixed values in a given context – and does not revisit them every time they handle that problem. Isolating invariants is a difficult task and requires an expert's skill and ingenuity (Wigner 1949; 1960). Due to their use of intelligence, human learning requires far fewer examples compared to machine learning, which uses brute force in data analytics (Vapnik and Izmailov 2019, 2020). Many potential variables in a problem are irrelevant to a task at hand. Based on their knowledge and experience, experts identify the invariants latent in the problem and put them aside, leaving the relevant variables. Vapnik and Izmailov (2019) exemplify jocosely with the *duck test*: "if it looks like a duck, swims like a duck and quacks like a duck, it probably is a duck"—just three features to qualify a duck, instead of many possible variables that characterize birds in general.

As an industry matures, innovation becomes predominantly incremental and experts accumulate knowledge on solving specific problems that emerge from established designs and architectures, which are not revisited every time a new task arises. Innovation occurs at component level, not in the overall design (the way components are coordinated as a system) (Henderson and Clark 1990). Fixed features in a product design are examples of invariants, although not the whole set of invariants, which is further skimmed by the expert by crafting the partial product hypothesis.

An increase in invariants diminishes the size of training data and leads to more accurate predictions, while an increase in variables requires more training data (Vapnik and Izmailov 2019, 2020).

Proposed forecasting method

Conventional data analytics is data driven, and an algorithm discovers patterns in that data (Yu 2007; Yu et al 2010). The domain expert only provides project guidelines and interprets the results from the trained model. The domain expert formulates a market hypothesis only after reviewing simulation results (Figure 2a). In the proposed method, the manager is involved in the entire process, including the formulation of an initial hypothesis and synthesizing data, not only in inference and judgment (Shrestha et al. 2020). A partial hypothesis using the manager's domain knowledge precedes the simulation (Figure 2b).

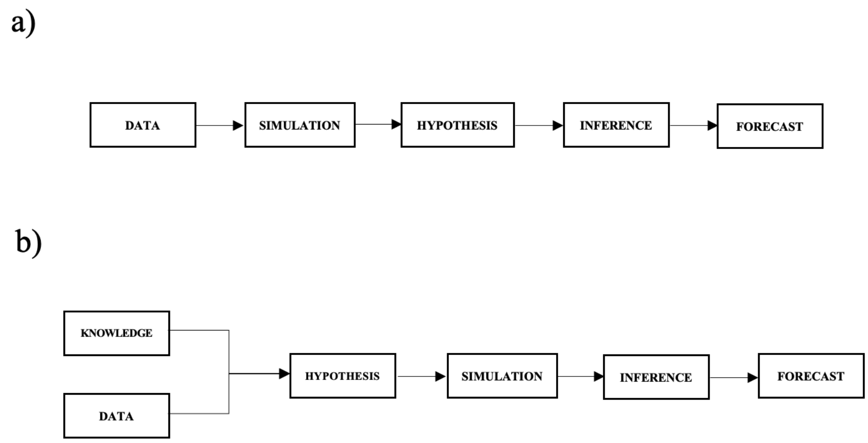


Figure 2. New product forecasting: a) conventional data analytics; b) proposal integrating DK and ML.

The proposed forecasting tool is based on algorithmic machine learning simulation guided by an expert's domain knowledge, which they use to formulate the initial hypothesis and to pre-process the training examples (Yu 2007; Yu et al. 2010). The initial hypothesis comes from the expert's knowledge and experience related to a specific market. Pre-processing involves selecting, cleaning, formatting, and synthesizing the training data. The simulation refines the hypothesis and converges to an optimal model by capturing the relations (weights) among multiple nonlinear variables. The stable set of variables and weights in the algorithm is used to predict the market share of new product concepts (feature sets). Domain knowledge guides the algorithmic training process, improving its performance, and making it more explainable (Yang et al. 2019).

The quality of new product forecasting depends on the quality of assumptions—coming from managerial judgment—and the quality of data (Kahn 2014). The initial hypothesis is a set of the relevant product features (Simon 1996) from the customer point-of-view (Woodruff 1997); it contains the independent variables of the machine learning model. A company usually relies on staff expertise, market research, static and dynamic testing, competitor benchmarking, distributor feedback, literature, and other sources to elaborate the hypothesis.

A manufacturer of complex products handles and defines hundreds (sometimes thousands) of feature variables, as they are elements of the overall product design and development (Weber 2009). But fewer variables comprise the key determinants of customer value (Weber 2009), and too many redundant features introduce noise and can harm the machine learning algorithm's performance (Yu et al. 2010). Identifying a compact set of relevant features to model the simulation is a challenge and prior domain knowledge is required to set the partial hypothesis (initial hypothesis space). Features that get left out are invariants in the model. The set of feature variables is a "partial" hypothesis, but it still misses important information—namely, weights estimated by the machine learning algorithm. The simulation converges the hypothesis space (potentially infinite sets of possible weights) to an optimum hypothesis (a stable set of variables and weights in the trained model).

Database construction starts with the hypothesis of product features. Then data on company and competitor product sales is collected by the expert and converted to market share. A relatively modest amount (in the order of a few hundred observations) of sales data are sufficient. Sample data are synthesized by evaluating the features of each product in the sample. A machine learning algorithm (an artificial neural network) processes the features and target values (market share calculated from sales), capturing patterns and estimating the weights of the variables. The trained model can then be used to forecast the market share of potential new products.

Experiment Using Data from the Brazilian SUV Market

To demonstrate the method, we present an experiment on small sport utility vehicles (SUVs) in the Brazilian market. Although there is no global consensus on the definition of an SUV, it refers to a closed passenger vehicle with a "two-box" styling bodywork, which is taller than a conventional automobile. By "small," we mean vehicles shorter than 4.5 meters (177 in) in overall length.

In this small-scale experiment, we formulated a hypothesis of key product features using industry knowledge and experience, comparing the feature set with specialized magazines and websites (Quatro Rodas 2020a; UOL Carros 2020), and validating them with prescriptions from an automobile development text (Weber 2009). We identified 16 primary product attributes: novelty, brand, style, robustness, comfort, space, trunk, convenience, finish, equipment, infotainment, economy, performance, agility, safety, and price. In reality, a large manufacturer would rely on multiple resources, including the experience of its executives and staff, market research, product testing, distributor surveys, and competitor benchmarking.

We collected monthly sales data of twenty-two small SUV products—from March 2003 to December 2020—from the National Federation of Brazilian Vehicle Distributors website (FENABRAVE 2021) and converted it to market share. It covers the entire small SUV history in Brazil and generated 1,563 observations (Yamamura 2020). We used market share to mitigate seasonal effects. As companies regularly forecast overall market volumes, it is easy to convert market share forecasts back to sales figures.

We used reviews from an automobile magazine (Quatro Rodas 2020a) and customer evaluations from a specialized website (UOL Carros 2020) to measure product features. Some vectors (observations) are repetitive; their feature values are constant over a certain period (like a few months). However, other vectors differ for the same product at different times in its product life. For example, *novelty* decreases as time passes: a new product scores five for *novelty* in its launch year, but one point is deducted every year after. We based *brand strength* on a survey from a Brazilian automobile magazine (Quatro Rodas 2020b). Although some variables are quantitative—like price, fuel economy, and performance—others are subjective, like brand and style. To normalize all features, we evaluated them using a one-to-five point Likert scale, a level of granularity that was sufficient to capture relations in the database (Table 1).

Table 1. A sample section of the database.

New	Brand	Style	Robust	Comf	Space	Trunk	Conv	Finish	Equip	Info	Econ	Perf	Agile	Safe	Price	Share
3	5	5	5	5	4	3	4	5	5	5	1	3	4	4	1	3.16
1	5	5	5	4	2	1	4	5	4	5	1	1	3	4	3	3.00
3	2	5	4	5	5	5	5	5	3	1	1	4	3	5	2	0.08
4	3	5	3	5	4	5	5	5	2	3	1	5	3	4	1	0.13
2	3	3	2	1	2	4	2	2	2	1	4	4	4	3	4	1.85
2	1	1	1	3	1	2	1	1	3	3	3	1	4	3	4	0.31
2	2	4	5	1	5	4	2	1	2	3	1	1	2	3	4	0.44
3	2	2	5	2	5	5	2	2	2	4	1	2	2	2	5	0.69
5	4	4	4	4	4	2	4	4	5	5	4	5	5	5	3	3.72
5	4	5	3	3	2	3	3	4	5	5	4	5	5	5	4	1.24

We split the whole dataset in 70:15:15 ratios among training, test, and validation sets—shuffled, drawn stochastically, and set apart from the beginning. The artificial neural network input layer contains 16 "neurons," corresponding to the feature variables. The output layer has just one node, predicting the target value. The model in the experiment presents a single hidden layer, adequate for a relatively small dataset (Goodfellow, Bengio, and Courville 2016).

No consensus exists regarding an appropriate number of units in a hidden layer. A common heuristic is between half and twice the input layer size (between n/2 and 2n), from 8 to 32 in the experiment. The neural network achieved the highest correlation between real and predicted data (R=0.96) with 10 hidden units. The algorithm estimates the weights of hidden and output layer neurons using *activation* functions. The model optimizes the weights using a function that minimizes the output (market share estimation) errors (Goodfellow, Bengio, and Courville 2016; Aggarwal 2018). We ran the algorithm on MATLAB™ R2019b, in a MacBook with 1.2 GHz Intel Core M processor.

We tested the method using real data, predicting market share of the six most significant (largest sales volume) new small SUVs launched in Brazil between 2016 and 2020. We used mean absolute percent error (MAPE) to measure the accuracy of those predictions. MAPE is an indicator commonly used in companies to measure the accuracy of predictions (Ching-Chin et al. 2010). MAPE is calculated by subtracting actual demand from the forecast; dividing by actual demand; taking its absolute value; and multiplying by one hundred:

$$MAPE = \left|\frac{fcst - actl}{actl}\right| * 100$$

Overfitting tends to be a major issue in machine learning algorithms (Shrestha et al. 2018). It consists of low bias but high variance in errors. In other words, an algorithm could perform well on the dataset used for training but could struggle to generalize to new data (Shrestha et al. 2020). Increasing the size of the sample or decreasing the size of the model (number of variables) are procedures used to handle overfitting. Machine learning algorithms have built-in procedures like regularization (penalization of variables with small effect) and cross-validation (splitting data in random subsamples for test and validation) to mitigate overfitting (Yu 2007; Yu et al. 2010). Since overfitting is not an absolute issue, but relative to comparison (Hawkins 2004), it can also be handled by testing, calibrating, and comparing new dataset results, and by interpreting and understanding the circumstances, limitations, and contingencies of the predictions (Hawkins 2004; Ng 2018; Shrestha et al. 2020).

In our experiment, we limited the number of feature variables (sixteen), using domain knowledge, and made sure the sample size of the original dataset (1,563 observations) was sufficient to generalize to new data. We conducted the MAPE analysis of six real products using data prior to each product's launch time, to assess the capacity of the algorithm to generalize to new data.

Results

In the six real products case, predictions were made with neural networks using the portion of the dataset prior to a new product launch. The results come from artificial neural networks trained only once (not after several trials), which is a reason why the errors were spread out for new data not in the training set. For instance, the neural network did a remarkable job at predicting the Renault Duster facelift (product improvement), which was likely easier because it was an improvement and not strictly new. In other words, the facelift was an enhanced version of an existing product, with some features values that were common to the previous version (i.e., present in the training dataset). The neural network did a relatively poor job of predicting the Captur, a completely new product also from Renault, with more new feature values.

Still the model presented meaningful predictions overall, being able to fit new data not present in the training dataset. The estimated MAPE ranged from 2 percent to 48 percent—with average of 18 percent, or accuracy of 82 percent (Table 2). According to Kahn (2010, 2002), the average forecasting accuracy commonly achieved by firms is 58 percent. Ching-Chin et al. (2010) suggest an adequate performance should be about 76 percent. As mentioned, overfitting is relative to understanding the circumstances of the prediction. In this case, a benchmark from the literature was used to reference and to confirm the model generalizes to new data.

Table 2. Six new product market share predictions, using ANNs trained with data prior to launch date.

Make	Model	Launch date	Mkt share (%)	Prediction	MAPE	1 - MAPE
Nissan	Kicks	Aug '16	1.254	1.312	4.6%	95.4%
Hyundai	Creta	Feb '17	1.952	1.879	3.7%	96.3%
Renault	Captur	Apr '17	0.732	1.080	47.5%	52.5%
Volkswagen	T Cross	Aug '18	2.806	3.526	25.7%	74.3%
Chevrolet	New Tracker	Apr '20	3.175	3.928	23.7%	76.3%
Renault	Duster facelift	Apr '20	1.377	1.350	2.0%	98.0%
Average					**17.9%**	**82.1%**

Second Experiment on Brazil's Medium-Size Pickup Market

For confirmation, we conducted an additional experiment on Brazil's medium-size pickup market in Brazil. The model is less elaborate as there were some limitations. For example, there is a smaller number of players (seven) and product life cycles tend to be longer than the small SUV segment—more than 10 years in some cases, like the Volkswagen Amarok, which was launched in 2010 and is still in the market).

The dataset contains the history of medium-size pickup trucks between January 2015 and December 2020, with 491 observations. We identified seven key feature variables—brand, robustness, style, engine power, cabin, price, and size—once gain based on industry experience from the authors and validation from specialized press documents and Weber's textbook (2009). We present a database extract (Table 3).

Table 3. A section of the pickup database.

Year	Month	Make	Model	Brand	Tough	Style	Power	Cabin	Price	Size	Mkt share
2020	12	Chevrolet	S10	4	4	4	4	3	3	5	1.38
2020	12	Fiat	Toro	2	1	3	2	2	5	3	3.00
2020	12	Ford	Ranger	4	3	4	3	4	2	5	1.03
2020	12	Mitsubishi	L200	4	4	4	3	3	2	5	0.30
2020	12	Nissan	Frontier	3	4	3	3	2	2	5	0.41
2020	12	Toyota	Hilux	5	5	4	4	3	1	5	1.42
2020	12	Volkswagen	Amarok	3	2	3	5	3	1	5	0.39

There were no recent product launches in the segment, but historical data includes product facelifts, content changes, and engine improvements. Since there were no "new-to-the-company" products in the given time period, we used data from three consecutive years to predict the market share in the following year. For instance, we predicted products in 2020 with a neural network trained with data from 2017 to 2019. The average forecast accuracy yielded was 78 percent (Table 4), confirming useful predictive power.

Table 4. Prediction of market share using ANNs trained with data from three previous years.

		2020				2019				2018			
Make	**Model**	**Mkt %**	**Pred**	**MAPE**	**1 - MAPE**	**Mkt %**	**Pred**	**MAPE**	**1 - MAPE**	**Mkt %**	**Pred**	**MAPE**	**1 - MAPE**
Chevrolet	S10	1.02	1.26	23.6%	76.4%	1.21	1.88	55.8%	44.2%	1.29	1.33	3.3%	96.7%
Fiat	Toro	2.77	2.40	13.3%	86.7%	2.47	2.28	7.5%	92.5%	2.37	2.48	4.6%	95.4%
Ford	Ranger	1.02	0.81	20.2%	79.8%	0.84	0.83	1.3%	98.7%	0.83	0.66	21.1%	78.9%
Mitsubishi	L200	0.49	0.42	13.5%	86.5%	0.38	0.46	18.8%	81.2%	0.44	0.39	9.9%	90.1%
Nissan	Frontier	0.41	0.31	25.2%	74.8%	0.30	0.06	80.4%	19.6%	0.26	0.18	28.3%	71.7%
Toyota	Hilux	1.66	1.55	6.6%	93.4%	1.52	1.67	9.7%	90.3%	1.59	1.57	1.1%	98.9%
Volkswagen	Amarok	0.54	0.78	44.2%	55.8%	0.71	0.76	7.3%	92.7%	0.76	1.35	77.5%	22.5%
Average				20.9%	79.1%			25.8%	74.2%			20.8%	79.2%
										Overall Average		22.5%	77.5%

Discussion and Managerial Implications

New product forecasting with small data—Our method demonstrates it is possible to forecast market share of incremental but complex new products using expert domain knowledge and small datasets. The manager uses domain knowledge to set the initial conditions for the product hypothesis, by pre-selecting the relevant product feature variables (and isolating the invariants). The machine learning algorithm simulates the market and estimates the weights of feature variables. The proposed method builds on strengths of the managerial profession—market knowledge and experience—to enhance new product development. It sharpens decision-making, as good decisions imply good forecasting.

By providing understanding and reducing uncertainty (Ganguly and Euchner 2018), it is also a learning tool. As new knowledge is acquired or market conditions change, the model needs to be updated and improved. For instance, market, technology, or legislation may dictate the emergence of new relevant customer features, like powertrain electrification (hybridization), barely present in current small SUVs in Brazil. The diffusion of that technology will require the introduction of a new feature variable in the database and hence in future algorithmic model. On the other hand, "infotainment" technology may become so common and undifferentiated that customers simply take it for granted and it is no longer a relevant variable for the method. In that case, it becomes an invariant and it is no longer explicit in the model. Those changes will demand both redesigning the database and re-training the algorithm, which simultaneously generate new learning for decision-makers and raise product development performance.

Constructing the hypothesis with domain knowledge—Constructing an appropriate set of features to frame the new product concept is a major challenge. Regrettably, there is no established formula to do it. As Wigner (1949) suggests, setting the initial conditions of an experiment is an art. Vapnik (2020) suggests it is the intelligence part of learning. But experienced managers have competence, understanding, and insight into their businesses. They have knowledge and intuition of what is relevant in those markets, and their skills are routinely used in their work, albeit not always in a structured way.

The initial hypothesis is a set of the most relevant feature variables in the market under consideration. The domain expert (manager) creates a hypothesis by separating key feature variables from invariants (unchanging features) in the pool of potentially infinite variables. The compact set of feature variables left is the product concept to be modeled and refined. The model is complete when the algorithm approximates the weights of the variables and is stable. To avoid potential bias, it is important the domain expert validates the partial hypothesis using surveys, press information, literature, and maybe a second expert's opinion.

Forecasting new products designs—The trained algorithm, with a fixed set of variables and weights, models the market to be investigated for new product development. New product concepts—vectors with evaluated features—can be inputted to predict market share. Our method can test both alternative company product proposals and potential competitor responses. Market experts can predict a new product demand by introducing a new feature vector in the trained algorithm. They may feed alternative variable sets (product concepts), testing their market sensitivity. Or they can input a set of features from a hypothetical competitor product, based on competitive intelligence, predicting its potential market strength and guiding the company's future pre-emptive actions.

Our proposed method is robust and reproducible for incremental new products in complex but relatively stable markets, but we do not recommend it to forecast "new-to-the-world" products because as the machine learning algorithm identifies nonlinear patterns in data, it assumes probability distributions that are constant over time, which is not the case for radical products.

Domain and data experts—Usually, domain and data experts are different people (Crews 2019; Amersh et al. 2014). Pure managerial judgment relies on intuition and experience, and a data expert is usually not involved. In data analytics, however, the algorithm is responsible for discovering patterns. The domain expert (manager) participates initially to provide general requirements for the project, and returns later in the process, when the model is ready. The domain expert's involvement is collateral and subsidiary (Amersh et al. 2014).

Our proposed approach requires that the domain expert and data scientist work closely throughout the process; otherwise, the domain specialist must possess some systems knowledge to handle both database and algorithm development. The domain expert is responsible for creating the initial product hypothesis (set of feature variables mapped to target indicators) and guiding data synthesis. Once the model is trained, the domain expert will interpret the results and explain them for new product development action.

With recent commercially available software, the manager does not need to be a full-fledged data expert (Gil et al. 2019; Amersh et al. 2014). In the next few years, using machine learning to support business decisions may become as common as manipulating spreadsheets and databases is today (Allen 2019, Euchner 2019).

Limitations

One limitation is that this is a single-industry study. The logic of the proposed approach (machine learning algorithms just read numbers and capture relations among them) is extensible to other sectors and contexts. Still, studies exploring the application in different domains could be valuable contributions. Also, further research exploring alternative machine learning techniques, such as regression SVMs or *extreme learning machine* (ELM), to new product strategy is a promising venue.

Conclusion

New product forecasting supported by machine learning is an emerging and fertile field of investigation. Our proposed forecasting method, which uses a machine learning algorithm combined with domain knowledge, can yield meaningful predictions. Our method precludes the need for huge datasets, which is important for new product development where little or no historical data exist. Narrowing the range of potential research variables increases efficiency and speed. Our method also reduces training data size and trades data quantity for quality. Our model explains and predicts, thereby supporting new product development and creating knowledge that can be abstracted and transferred to other projects.

References

Aggarwal, C. C. 2018. *Neural Networks and Deep Learning: A Textbook*. New York: Springer.

Alavi, M., and Leidner, D. E. 2001. Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly* 25(1): 07–136. doi: 10.2307/3250961

Aleksandrov, A.D., Kolmogorov, A.N., and Lavrent'ev, M.A. 1999. *Mathematics: Its Contents, Methods and Meaning*. New York: Dover Publications.

Allen, R. 2019. Five lessons for applying machine learning. *Research-Technology Management* 62(3): 38–44. doi:10.1080/08956308.2019.1587330.

Amershi, S., Cakmak, M., Knox, W.B., and Kulesza, T. 2014. Power to the people: the role of humans in interactive machine learning. *AI Magazine* 35(4): 105–120. doi: 10.1609/aimag.v35i4.2513.

Armstrong, J. S. 2002. Benchmarks for new product forecast errors. *Journal of Product Innovation Management* 19: 133–143. https://www.researchgate.net/profile/J-Armstrong/publication/237409816_Benchmarks_for_New_Product_Forecast_Errors/links/02e7e52814b26c3423000000/Benchmarks-for-New-Product-Forecast-Errors.pdf.

Ching-Chin, C., Ieng, A. I. K., Ling-Ling, W., and Ling-Chieh, K. 2010. Designing a decision-support system for new product sales forecasting. *Expert Systems with Applications* 37(2): 1654–1665. doi: 10.1016/j.eswa.2009.06.087

Crews, C. 2019. What machine learning can learn from foresight: A human-centered approach: for machine learning–based forecast efforts to succeed, they must embrace lessons from corporate foresight to address human and organizational challenges. *Research-Technology Management* 62(1): 30–33. doi: 10.1080/08956308.2019.1541725

Dane, E., and Pratt, M.G. 2007. Exploring intuition and its role in managerial decision making. *Academy of Management Review* 32(1): 33–54. doi: 10.5465/amr.2007.23463682

Daugherty, P. and Euchner, J. 2020. Human + machine: collaboration in the age of AI. *Research-Technology Management* 63(2): 12–17. doi: 10.1080/08956308.2020.1707001

Euchner, J. 2019. Little ai, big AI–good AI, bad AI. *Research-Technology Management* 62(3): 10–12. doi: 10.1080/08956308.2019.1587280.

FENABRAVE (National Federation of Brazilian Vehicle Distributors) 2021. *New Registrations*. http://www.fenabrave.org.br/Portal/conteudo/emplacamentos. Accessed on Jan 12th, 2021.

Fujimoto, T. 2014. The long tail of the auto industry life cycle. *Journal of Product Innovation Management* 31(1): 8–16. doi: 10.1111/jpim.12076

Ganguly, A., and Euchner, J. 2018. Conducting business experiments: Validating new business models well-designed business experiments can help validate assumptions and reduce risk associated with new business models. *Research-Technology Management* 61(2): 27–36. doi: 10.1080/08956308.2018.1421381

Gil, Y., Honaker, J., Gupta, S., Ma, Y., D'Orazio, V., Garijo, D., Gadewar, S., Yang, Q., and Jahanshad, N. 2019. Towards human-guided machine learning. *Proceedings of the 24th International Conference on Intelligent User Interfaces*: 614-624. doi: 10.1145/3301275.3302324.

Goodfellow, I., Bengio, Y., and Courville, A. 2016. *Deep Learning*. Cambridge: MIT press.

Hawkins, D.M. 2004. The problem of overfitting. *Journal of Chemical Information and Computer Sciences* 44(1): 1–12. doi: 10.1021/ci0342472

Henderson, R. M., and Clark, K. B. 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* 35(1): 9–30. Doi: 10.2307/2393549

Hobday, M. 1998. Product complexity, innovation and industrial organisation. *Research Policy* 26(6): 689–710. doi: 10.1016/S0048-7333(97)00044-9

Kahn, K. B. 2002. An exploratory investigation of new product forecasting practices. *Journal of Product Innovation Management* 19(2): 133–143. doi: 10.1111/1540-5885.1920133

Kahn, K.B. 2006. *New Product Forecasting*. New York: M.E. Sharpe.

Kahn, K. B. 2010. New product forecasting. Wiley International Encyclopedia of Marketing. doi: 10.1002/9781444316568.wiem05011

Kahn, K. B. 2014. Solving the problems of new product forecasting. *Business Horizons* 57(5): 607–615. doi: 10.1016/j.bushor.2014.05.003

Mitchell, T. M. 1997. *Machine Learning*. New York: McGraw-Hill.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. *Nature* 521: 436–444. doi: 10.1038/nature14539

Lee, H., Kim, S. G., Park, H. W., and Kang, P. 2014. Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. *Technological Forecasting and Social Change* 86: 49-64. doi: 10.1016/j.techfore.2013.08.020

Leifer, R., O'Connor, G. C., Rice, M. 2001. Implementing radical innovation in mature firms: The role of hubs. *Academy of Management Perspectives* 15(3): 102–113. doi: 10.5465/ame.2001.5229646

Linton, J. D. 2002. Forecasting the market diffusion of disruptive and discontinuous innovation. *IEEE Transactions on Engineering Management* 49(4): 365–374. doi: 10.1109/TEM.2002.806723

Ng, A. 2016. What artificial intelligence can and can't do right now. *Harvard Business Review*, November 9. https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now

Ng, A. 2018. Machine Learning Yearning. Electronic book. https://www.deeplearning.ai/programs/

Quatro Rodas. 2020a. *Vehicle Tests*. https://quatrorodas.abril.com.br/testes/. Accessed on Nov. 2020.

Quatro Rodas. 2020b. Top brands. *Quatro Rodas*, January 2020, 729: 69–78.

Rice, M. P., O'Connor, G. C., Peters, L. S., and Morone, J. G. 1998. Managing discontinuous innovation. *Research-Technology Management* 41(3): 52–58. doi: 10.1080/08956308.1998.11671210

Sharma, A.K., Goel, N., Rajput, J., and Bilal, M. 2020. *An intelligent model for predicting the sales of a product*. IEEE. https://ieeexplore.ieee.org/abstract/document/9057880.

Shrestha, Y. R., He, V. F., Puranam, P., and von Krogh, G. 2020. Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science* 32(3):856-880. doi: 10.1287/orsc.2020.1382

Simon, H.A. 1996. *The Sciences of the Artificial*. 3rd ed. Cambridge: MIT Press.

Spangler, W. E. 1991. The role of artificial intelligence in understanding the strategic decision-making process. *IEEE Transactions on Knowledge and Data Engineering* 3(2): 149–159. doi: 10.1109/69.87995

Ulrich, K.T., Eppinger, S.D., and Yang, M.C. 2020. *Product Design and Development*. New York: Mc Graw-Hill.

UOL Carros 2020. *Tests and Evaluations*. https://www.uol.com.br/carros/testeseavaliacoes/capa-a.jhtm. Accessed on Nov. 2020.

Vapnik, V. 2019. Complete statistical theory of learning. *Automation and Remote Control* 80(11): 1949–1975. doi: 10.1134/S000511791911002X

Vapnik, V., and Izmailov, R. 2019. Rethinking statistical learning theory: learning using statistical invariants. *Machine Learning* 108: 381–423. doi:10.1007/s10994-018-5742-0.

Vapnik, V., and Izmailov, R. 2020. Complete statistical theory of learning: learning using statistical invariants. *Conformal and Probabilistic Prediction and Applications* 2020 Aug 16: 4-40. http://proceedings.mlr.press/v128/vapnik20a.html.

Veryzer, R. W. 1998. Discontinuous innovation and the new product development process. Journal of Product Innovation Management, 15(4): 304–321. doi: 10.1111/1540-5885.1540304.

Weber, J. 2009. Automotive Development Process. Berlin: Springer-Verlag.

Wigner, E.P. 1949. Invariance in physical theory. Proceedings of the American Philosophical Society 93(7): 521–526. http://www.jstor.org/stable/3143140

Wigner, E.P. 1960. The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics* 13: 1–14. doi: 10.1142/9789814503488_0018

Woodruff, R.B. 1997. Customer value: the next source for competitive advantage. *Journal of the Academy of Marketing Science* 25(2): 139–153. doi:10.1007/BF02894350

Yamamura, C. 2020. Small SUVs Brazil 2003-2020. Harvard Dataverse. doi: 10.7910/DVN/DKCLGA

Yang, Z., Al-Bahrani, R., Reid, A.C., Papanikolaou, S., Kalidindi, S.R., Liao, W.K., Choudhary, A. and Agrawal, A. 2019. Deep learning based domain knowledge integration for small datasets: Illustrative applications in materials informatics. *2019 International Joint Conference on Neural Networks (IJCNN)*: 1–8. doi: 10.1109/IJCNN.2019.8852162

Yin, P., Dou, G., Lin, X., and Liu, L. 2020. A hybrid method for forecasting new product sales based on fuzzy clustering and deep learning. *Kybernetes* 49(120): 3099–3118. doi: 10.1108/K-10-2019-0688

Yu, T. 2007. Incorporating prior domain knowledge into inductive machine learning. PhD diss., University of Technology Sidney. https://opus.lib.uts.edu.au/handle/2100/385

Yu, T., Simoff, S., and Jan, T. 2010. VQSVM: a case study for incorporating prior domain knowledge into inductive machine learning. *Neurocomputing* 73(13-15): 2614–2623. doi: 10.1016/j.neucom.2010.05.007

APPENDIX C: The minority game in product strategy

Journal	Journal of Evolutionary Economics – Scopus CiteScore 2021 = 4.0 (76% percentile)
Authors	Charles Lincoln Kenji Yamamura; Cintia Isabel de Campos; Cira Souza Pitombo; José Carlos Curvelo Santana; José Alberto Quintanilha; Fernando Tobal Berssaneti
Reference	Yamamura, C. L. K., Campos, C.I., Pitombo, C.S., Santana, J. C. C., Quintanilha, J. A., & Berssaneti, F. T. (2022). The Minority Game in Competitive Product Strategy. *Journal of Evolutionary Economics*
DOI	Submitted

Journal of Evolutionary Economics

The minority game in product strategy

--Manuscript Draft--

Manuscript Number:		
Full Title:	The minority game in product strategy	
Article Type:	regular article	
Corresponding Author:	Charles Yamamura Universidade de Sao Paulo BRAZIL	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	Universidade de Sao Paulo	
Corresponding Author's Secondary Institution:		
First Author:	Charles Lincoln Kenji Yamamura	
First Author Secondary Information:		
Order of Authors:	Charles Lincoln Kenji Yamamura	
	Cintia Isabel de Campos, PhD	
	Cira Souza Pitombo, PhD	
	José Carlos Curvelo Santana, PhD	
	José Alberto Quintanilha, PhD	
	Fernando Tobal Berssaneti, PhD	
Order of Authors Secondary Information:		
Funding Information:	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (0001)	Mr. Charles Lincoln Kenji Yamamura
	Conselho Nacional de Desenvolvimento Científico e Tecnológico (305987/2018-6)	Prof. José Carlos Curvelo Santana
Abstract:	The competitive interaction among product strategies in complex markets is an evolutionary game, where players engage using strategies that are forced to differ and there is no optimum solution. It is a network of multiple nonlinear relations, operating at the edge between stability and instability. Information is incomplete, settings change continuously, and outcomes in the long run are unpredictable. However, it is possible to understand patterns of behavior to make better decisions. Winners are the minority players, exhibiting differing strategies but also common patterns – despite flickering settings and constraints. This paper proposes a framework to understand competitive product strategy behavior, illustrating with a case from the automobile industry, validated by an artificial neural network model. Categorical principal component analysis (CATPCA) was used to analyze relations among the attributes. The experiment yielded meaningful results, showing the framework provides valuable insights for deciding how and where to focus in product strategy.	
Suggested Reviewers:		

The minority game in competitive product strategy

Abstract

The competitive interaction among product strategies in complex markets is an evolutionary game, where players engage using strategies that are forced to differ and there is no optimum solution. It is a network of multiple nonlinear relations, operating at the edge between stability and instability. Information is incomplete, settings change continuously, and outcomes in the long run are unpredictable. However, it is possible to understand patterns of behavior to make better decisions. Winners are the minority players, exhibiting differing strategies but also common patterns – despite flickering settings and constraints. This paper proposes a framework to understand competitive product strategy behavior, illustrating with a case from the automobile industry in Brazil, validated by an artificial neural network model. Categorical principal component analysis (CATPCA) was used to analyze relations among the attributes. The experiment yielded meaningful results, showing the framework provides valuable insights for deciding how and where to focus on product strategy.

Keywords

Product strategy; evolutionary game; competitive strategy; minority rule; machine learning; CATPCA.

1. Introduction

In Cost Rica, choruses of male frogs sing to attract females, despite the risk of being caught by frog-eating bats. It is a competitive game, by advertising to attract mates away from rival males. If a frog sings louder than its competitors, it increases its chance of mating but decreases its odds of survival; the reverse is true also. If most males are quiet, a male is better off being louder. If most rivals are loud, it benefits from being quiet. Over time, the strategies tend to converge towards the same volume for all males.

(Adapted from Vincent and Brown, 2005)

Product strategy in dynamic markets is a challenging endeavor, with uncertainty about both customers and competitors (Sorenson, 2000a). Competitive dynamic markets are characterized by multiple players, complex strategies, and incomplete information, with variables changing over time. Most firms have leeway to choose among different product strategies. However, there are many constraints – from the market, technology, economics, corporate policy, history, tradition, legislation – making it difficult to yield optimal strategies. In addition, the best strategy does not depend only on what a firm does, but also on what competitors do. The *minority rule* states winners play differing strategies, achieving a superior balance in attributes despite the constraints (Devetag et al. 2014). This paper explains how the minority rule works in competitive product strategies and how it can enhance decisions in complex environments. The framework is illustrated with an automobile industry case. A database is engineered as the repository of market history, contemplating customers, product attributes, competitors, pricing, and brands. The strategy space is hard to map mentally or heuristically, hence an artificial neural network traces those relations. The experiment confirms winners adopt distinctive strategies, command higher than average market share and enjoy stability; but they also exhibit common patterns. The proposed framework and method are original contributions to understand competitive product markets and provide practical insights for managerial decision-making.

2. Theoretical background

Evolutionary game theory was formulated by Smith and Price (1973, 1976, 1982) and elaborated further by Vincent and Brown (1985, 1987, 2005, 2011). A game is a model of the interaction among players using strategies, in pursuit of payoffs (Sharif and Heydari 2013). Evolution implies changes in the strategies over time, under selection and adaptation to the current situation (Li et al. 2019). Evolutionary games are composed of players, strategies, payoffs, and a solution concept (Vincent and Brown 2005; Vincent et al. 2011). Players are sets of individuals with the same strategies and payoffs. A strategic group is a set of players who perceive themselves as being part of the same game (Camerer 1991). Strategies are unique and inheritable sets of attributes; different players act with different strategies. A *product strategy* is a set of attributes, mapped to customer value (Simon 1996, Woodruff 1997). Value is the expression of benefits perceived by customers (Woodruff 1997).

A strategy set comprises all evolutionary strategies that are feasible for a player. Possible strategies are constrained by genetic, physical, and environmental factors. Other constraints can be derived from inter-dependencies among the attributes. Some attributes are fixed, or *invariants*; others are changeable. Companies are constrained in the range of feasible product strategies; but they still enjoy significant leeway, depending on perceptions and resources. Ratneshwar et al. (1999) recommend a broad vision of strategic groups, since product innovation and changes in pricing policies may reshape the market and affect consumer considerations.

Payoffs are expressions of fitness – the relative presence of a strategy at a particular time (Vincent et al. 2011). While classical games focus on strategies to optimize payoffs, evolutionary games look for persistence in time. The *solution concept* is a function that maps the strategies to fitness. Since fitness results from the interaction of strategies, evolution by natural selection is an emerging evolutionary game (Vincent and Brown 2005). Natural selection is an optimization problem, seeking the best fitness in a given time and space. But optimum is an unstable state. By seeking persistence of strategies, evolution is a game – not just a problem of optimization. Players may disappear, but strategies persist through inheritance. New strategies occasionally emerge through mutation or invasion. The former is a heritable change in the attributes; the later occurs when a new player enters the strategic group.

Sorenson (2000b) observed competition and legitimation are the two forces that shape a market; the number of players (fitness) is their measurement. Legitimacy happens when a player is taken for granted in an industry, the degree of acceptance by both customers and industry peers. It determines a player's resilience in the market; once established legitimate players tend to exhibit a certain inertia in the market – expressed by a steady market share. New entrants have a hard time to acquire legitimacy. The complexity (number of interactions) of a market grows by the square of its density (number of players). Competition will increase with the number of players, unless the market expands at the same rate as the number of interactions (Sorensen 2000b).

The "El Farol bar" model is a game of adaptive players with incomplete information. Players win by being in the *minority*: going to the bar, if it is relatively empty; staying home, if it is crowded (Arthur 1994, 1999). Nobody knows what others are doing but they have information about past attendance. Players learn by inductive reasoning. Attendance converges to a threshold value over time; but it is dynamically unstable, continuously fluctuating around that value. The *El Farol* problem was mathematically solved by Challet, Zhang and Marsili (1997, 1998, 2001, 2004, 2005), in the *minority game*. Convergence itself is a consequence of the law of large numbers (Challet et al. 2004). But there is no winning strategy – if there were such a strategy, all players would adopt it and then they would all lose (Casti 1996; Ranadheera et al. 2017). Strategies are forced to differ, to fluctuate, and to evolve quickly (Casti 1996; Bischi and Merlone 2017). A larger minority or a variety of differing strategies imply smaller fluctuations and a more efficient overall system (Bottazzi and Devetag 2007; Ranadheera et al. 2017). Success in real life situations, as in financial trading, is often a composite of majority and minority decisions (Bischi and Merlone 2017). Up to a point, it is convenient to follow trends – players tend to be conservative and there are benefits from collective experience (Sharif and Heydari 2013). But deviating at the right time is a successful strategy (Devetag et al. 2014; Challet, Marsili and Zhang 2005).

Game theory models the competition among alternative strategies, attempting to predict the actions of rival players (Kleindl 1999). Competition is the pursuit of market space by players with strategies contending for similar target customers (Hoffmann et al. 2018). Despite its name, game theory is a set of tools for disciplining model construction, not a substantive theory by itself (Postrel 1991; Kleindl 1999). It is neither normative nor descriptive; it is analytic, seeking the implications of strategic situations (Camerer 1991). Strategic settings are complex nonlinear networks that operate at the edge of stability and instability, exhibiting both at the same time. They are unpredictable in the long-run and the task of strategy making is to understand behavior patterns, not outcomes by themselves (Stacey 1995). Business strategy is a practical field and game theory comes into existence by being applied to real strategy situations (Camerer 1991; Saloner 1991; Stacey 1995). Chatterjee (1986) urged the confluence of abstraction from game theory and empiricism of marketing science, to design robust competition models with practical reach. To gain legitimacy, game theory needs a corroborating case-specific theory. It links supporting theory and managerial actions, providing a decision framework for choosing strategies based on predictions of competitor actions. Since strategy implies a view of the future, some form of prediction is necessary (Wilson 1999). Prediction should not be understood as forecasting, but as the construction of hypotheses and the analysis of their sensitivity to theory and data (Simon 1996). From Mintzberg (1994), strategy making is a process of learning from all available resources – including experience and data – and synthesizing them into a vision of the future.

Porter's *five forces* – customers, suppliers, existing competitors, new entrants, and substitutes – are a well-known framework (1979, 2008). The *minority game* implies intense rivalry and positioning for differentiation from competitors. Among the reasons for intense rivalry, Porter cites multiple committed competitors and imperfect information. This investigation elaborates on the interactions among heterogenous and complex strategies, contending in evolving markets. It confirms winners are the minority players. Although all players are forced to differ in their strategies, minority players exhibit some common patterns. The framework assists decision-makers to both understand the product strategy game and to guide their actions.

3. Method

This study is grounded on theoretical review, conceptual framework, and a case study with simulation. Key constructs were defined, and a conceptual framework was built. In the experiment, an artificial neural network (ANN) simulated and validated the competitive strategy dynamics of an automobile segment. Attribute dimensions were identified using a *nonlinear principal components analysis* (PCA) technique. The investigation is based on a single industry and does not seek statistical generalization. However, from Eisenhardt (1989), a case study can provide in-depth understanding of a particular phenomenon, providing explanation and prediction. Slater (1995) extolls the control of market influences in single industry research, enhancing internal validity. It is preferable to go narrow and deeper in the investigation, relating it to a specific performance criterion.

Case study

Small SUVs are the fastest growing automobile segment in Brazil, currently accounting for 30% of the market (Fig. 1). Although there were small *off-road* vehicles in the market before, the Ford Ecosport is recognized as the product that inaugurated the segment in 2003 (Bazanini and Berton 2011). In its first year of production, the Ecosport held 2% of the overall automobile market, growing to almost 3%, a couple of years later – significant figures for a single model. It remained without a strong direct competitor until 2011, when Renault launched the Duster. Currently, there are twenty-seven players in the small SUV segment.

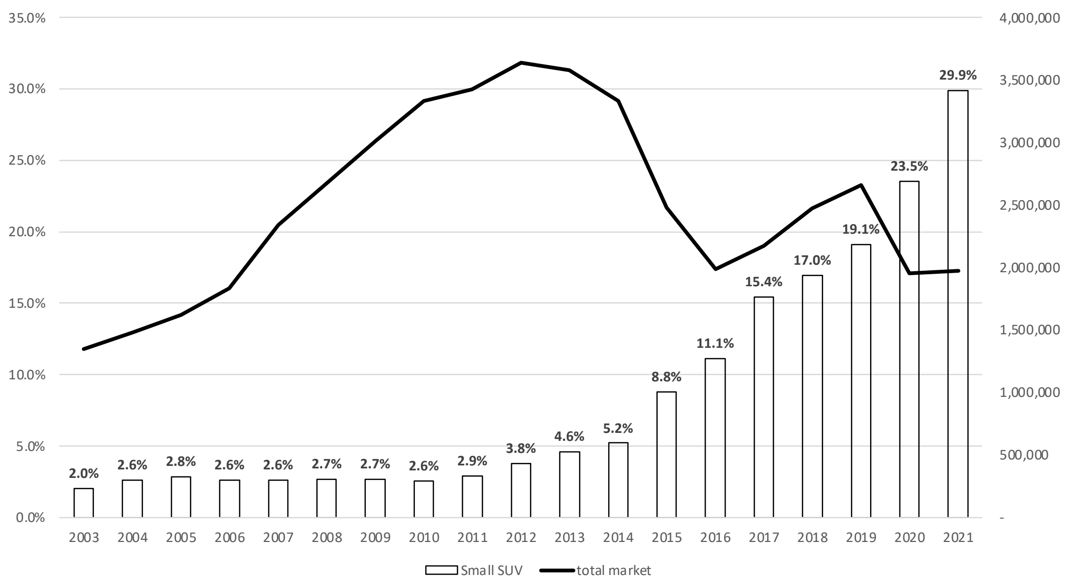


Fig. 1 Total automobile sales and small SUV market share in Brazil. Source: FENABRAVE.

Modeling and simulation

A database summarizing product strategy in the small SUV segment, from years 2003 to 2020, was crafted. Year 2021 was not included in the model due distortions from the Covid-19 pandemic, which affected the supply of electronic components and sales of several vehicles. We relied on specialized publications, customer evaluation websites (Quatro Rodas 2021; UOL Carros 2021), and industry experience to identify the key product features (customer attributes). The initial feature assessment yielded sixteen product attributes, which were then confronted with the prescriptions from Weber's automobile development text (2009) - for validation. *Novelty, brand, style, robustness, space, trunk, comfort, agility, convenience, finish, equipment, infotainment, performance, economy, safety, and price* were the selected features. There is a parallel between strong product attributes and USPs - *unique selling propositions* (distinctive innovations for product marketing) (Weber 2009). Although USPs are tools for advertising and attributes are elements for strategy, they are aligned since the later are mapped to customer value.

Market share is the fitness indicator adopted in this study to anchor the model to reality and to validate its accuracy. It is a proxy for *customer value*, as customers vote with their pockets. Monthly sales data (vehicle registration figures) were collected from the Brazilian Federation of Vehicle Distributors website (FENABRAVE 2021), generating 1,571 observations. Market share is the ratio between product sales and market size. Using market share instead of straight sales figures purges market seasonality and effects of fluctuations in the economy, when comparing data from different moments. As the SUV market has been growing continuously and it is receiving new entrants, share of the overall automobile market was adopted, instead of share of the SUV segment.

A product strategy is a vector of features (product attributes). Each attribute in the observations was evaluated to synthesize the vectors in the matrix database. Most sample vectors are recursive as their attribute values are constant over time. However, some attributes may differ for the same product in different moments of time. For instance, one point was deducted in *novelty* for every year from launch time to account for product ageing. It recovers one or two points if it receives a styling update, depending on the intensity of the refreshening action. Product specifications may also be updated, receiving improvements along its life cycle. An automobile manufacturer would rely on sources like expert knowledge, proprietary research, static and dynamic testing, competitor *benchmarking* and *tear down*, to evaluate products. For this experiment, customer evaluations from a website (UOL Carros 2021) and reviews from a specialized automobile magazine (Quatro Rodas 2021) were used. The objective was to find *satisficing* solutions, not necessarily optimal ones (Simon 1996), not numerical precision but adequate representations for information processing. A discreet scale of values from one to five was used to evaluate the attributes.

The second stage of the process is to find a *fitness generating function* (Vincent and Brown 2005), modeling game rules and minimizing the conditional probabilities of error. A shallow artificial neural network (ANN) was chosen because of its capacity to handle variables in complex, nonlinear and hierarchical levels of abstraction. It is a two-stage hierarchical function that maps linear combinations of input features into nonlinear representations of target values. Data is processed by inputting feature data and adjusting the parameter weights in the connections through layers of hidden nodes. Errors are calculated in a loss function and learning occurs by *feed-forward* and *back propagation* iterations, minimizing errors by seeking global minima through stochastic gradient descent. As data moves through higher layers, more abstract representations are processed through the network.

Data was split among training, validation, and testing sets in 70:15:15 ratios. About the number of neurons in the hidden layer, the best result was achieved with ten units - activated by the hyperbolic tangent function $g(z) = tanh(z)$. Target values are activated by the *rectified linear unit* (ReLU) function $f = max\ (0, x)$. For optimization, the Levenberg-Marquardt method was used to find the minimum of the sum of squared errors in the functions (Marquardt 1963). Predictions were compared with real market share data (target values) and a correlation R = 0.96 was obtained (Fig.2), showing a high level of explanation. Those figures suggested the model was an adequate representation of competitive product strategies mapped to fitness. It was run on MATLAB™ R2021a, in a MacBook™ with 1.2 GHz Intel™ Core M processor.

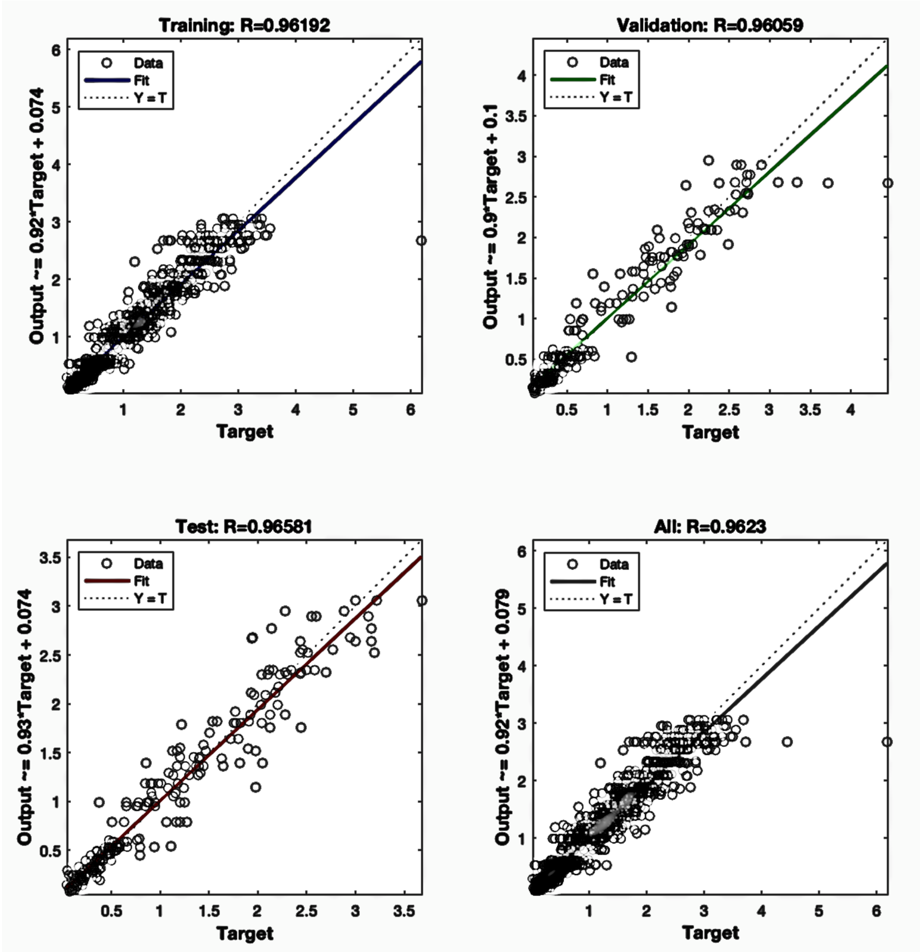


Fig. 2 Correlation between real data and predictions - 0.96 (MATLAB™).

For the analysis of attribute clusters, the *categorical principal component analysis* (CATPCA) was performed, which reduces data dimension assigning metric values to non-metric variables in a dataset with different levels of measures and nonlinear relations. The *ordinal level of optimal scaling*, v*ariable principal normalization* and *varimax rotation* methods were selected for the analysis (IBM 2020). Dimensions were defined according to the *Kaiser Criterion* (eigenvalue greater than 1), total variance explained and interpretability. For technical information, see Kuroda et al. (2013), Linting and Van Der Kooij (2012), Campos et al. (2020) and Linting et al. (2007). Analysis was run in the IBM SPSS 25 statistical software package.

The framework used for evolutionary product strategy analysis is summarized in Fig. 3.

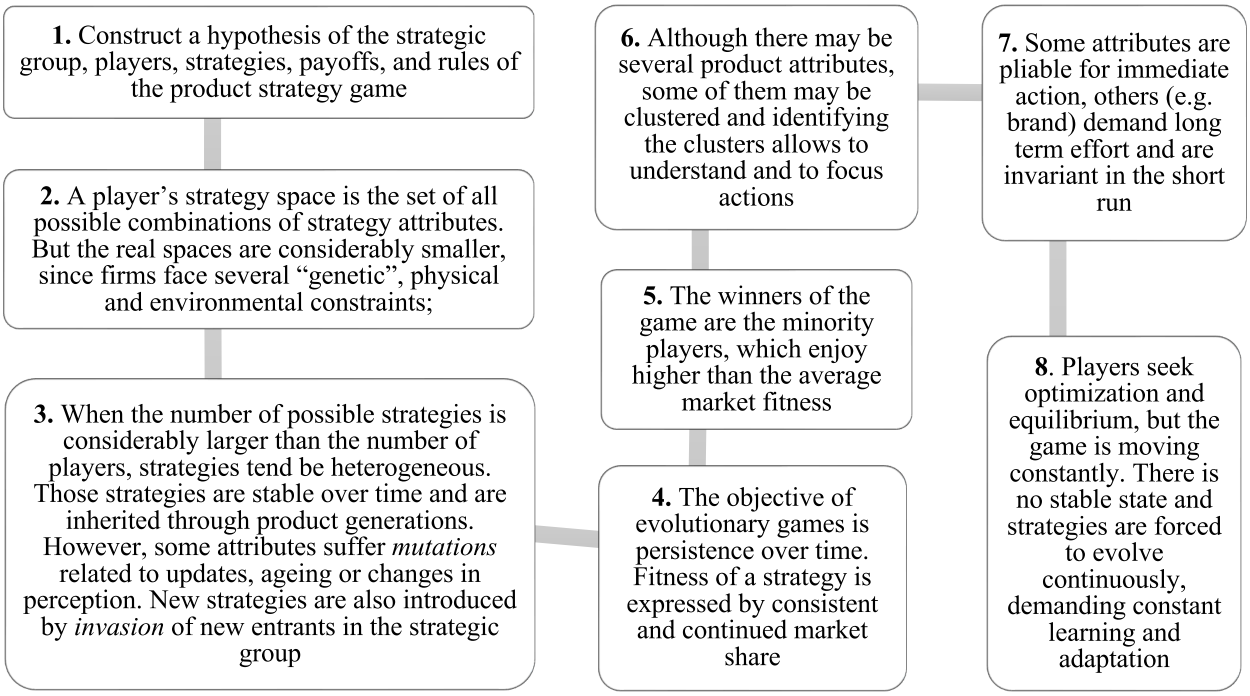


Fig. 3 Evolutionary product strategy framework.

4. Results

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	avg
Ecosport	**2.02**	**2.61**	**2.80**	**2.38**	**2.01**	**1.65**	**1.45**	**1.29**	**1.12**	**1.05**	**1.85**	**1.63**	**1.37**	**1.41**	**1.44**	**1.40**	**1.29**	**1.23**	**1.67**
Tucson			0.03	0.19	0.53	0.74	0.96	0.84	0.53	0.58	0.45	0.55	0.46	0.56	0.23	0.21	0.13	0.12	0.44
Sportage			0.01	0.02	0.09	0.27	0.26	0.22	0.24	0.25	0.26	0.31	0.27	0.23	0.17	0.23	0.16	0.08	0.19
iX35								0.20	0.40	0.30	0.29	0.46	0.59	0.51	0.47	0.35	0.22	0.13	0.36
ASX								0.01	0.32	0.30	0.28	0.36	0.39	0.24	0.23	0.22	0.14	0.06	0.23
Duster									0.27	**1.29**	**1.40**	**1.47**	**1.39**	**1.28**	0.81	0.95	0.98	1.00	**1.08**
Tracker											0.07	0.43	0.45	0.43	0.56	1.06	0.61	2.53	0.77
HRV													**2.07**	**2.81**	**2.20**	**1.94**	**1.86**	**1.67**	**2.09**
Renegade													**1.58**	**2.60**	**1.76**	**1.88**	**2.58**	**2.91**	**2.22**
2008													0.25	0.54	0.49	0.39	0.33	0.24	0.37
Kicks														0.54	**1.54**	**1.89**	**2.11**	**1.87**	**1.59**
Compass															**2.26**	**2.44**	**2.27**	**2.71**	**2.42**
Creta															**1.92**	**1.98**	**2.16**	**2.45**	**2.13**
WRV															0.71	0.60	0.46	0.54	0.58
Captur															0.63	1.07	1.08	0.56	0.84
Tiggo																0.21	0.24	0.24	0.23
Cactus																0.14	0.62	0.49	0.41
TCross																	**1.78**	**3.08**	**2.43**
Tiggo5X																	0.32	0.45	0.39
Tiggo7																	0.15	0.13	0.14
Eclipse																	0.09	0.13	0.11
Nivus																		0.83	0.83
Territory																		0.08	0.08
avg	2.02	2.61	0.95	0.86	0.88	0.88	0.89	0.51	0.48	0.63	0.66	0.74	0.88	1.01	1.03	1.00	0.93	1.02	**0.94**

Table 1 Market share (%) by model, between 2003 and 2019. Source: FENABRAVE.

The small SUV market model is an evolutionary game with twenty-three distinct players in 2020 (Table 1). Strategies are vectors of sixteen product attributes, each with five score levels; hence each player has a potential strategy space of 5^{16}, or S = $\sim 1.5 \times 10^{11}$ possible strategies. In principle, the probability of someone playing a certain strategy is $\frac{1}{1.5\times 10^{11}}$. However, effective strategy space is smaller because of firm and environment constraints. Still, the probability of two players playing the exact same strategy is very low, if the strategy space is considerably larger than the strategic group size ($S \gg N = 23$). The examination of strategies in the small SUV model, confirms there are no players with identical strategies (Table 2).

model	new	brand	style	robust	comf	space	trunk	conv	finish	equip	info	econ	perf	agile	safe	price	mkt%
TCross	**5**	**4**	**4**	**4**	**4**	**4**	2	**4**	**4**	**5**	**5**	**4**	**5**	**5**	**5**	3	**2.43**
Compass	3	**5**	**5**	**5**	**5**	**4**	3	**4**	**5**	**5**	**5**	1	3	**4**	**4**	1	**2.42**
Renegade	1	**5**	**5**	**5**	**4**	2	1	**4**	**5**	**4**	**5**	1	1	3	**4**	3	**2.22**
Creta	2	**5**	3	**4**	**4**	3	**4**	3	3	3	**4**	2	2	**4**	2	**4**	**2.13**
HRV	2	**4**	**4**	3	3	3	**4**	**5**	3	1	2	**4**	3	**5**	5	3	**2.09**
Ecosport	1	3	2	2	3	1	2	3	**4**	**4**	**5**	3	2	**5**	3	**5**	**1.67**
Kicks	2	3	3	2	1	2	**4**	2	2	2	1	**4**	**4**	**4**	3	**4**	**1.59**
Duster	3	2	2	**5**	2	**5**	**5**	2	2	2	**4**	1	2	2	2	**5**	**1.08**
Captur	2	2	4	5	1	5	4	2	1	2	3	1	1	2	3	4	0.84
Nivus	5	4	5	3	3	2	3	3	4	5	5	4	5	5	5	4	0.83
Tracker	5	4	4	3	4	2	3	4	3	5	5	3	5	5	4	3	0.77
WRV	3	4	1	3	2	2	2	5	2	1	1	4	3	4	4	5	0.58
Tucson	3	5	4	4	5	5	5	5	5	5	4	3	5	3	5	1	0.44
Cactus	4	1	2	1	4	2	1	2	3	4	3	2	2	2	2	5	0.41
Tiggo5X	4	1	3	1	3	3	1	2	3	4	4	3	4	1	2	3	0.39
2008	2	1	1	1	3	1	2	1	1	3	3	3	1	4	3	4	0.37
iX35	1	5	3	3	4	4	5	2	4	3	4	1	3	1	2	2	0.36
ASX	1	3	2	5	2	4	3	3	2	2	3	1	3	2	3	3	0.23
Tiggo	3	1	2	1	1	1	3	1	1	2	2	1	1	2	1	5	0.23
Sportage	3	2	5	4	5	5	5	5	5	3	1	1	4	3	5	2	0.19
Tiggo7	4	1	4	1	5	4	3	3	4	5	5	3	4	1	2	2	0.14
Eclipse	4	3	5	3	5	4	5	5	5	2	3	1	5	3	4	1	0.11
Territory	5	3	2	2	3	5	3	3	4	5	4	2	3	3	3	1	0.08
																avg	**0.94**

Table 2 Product strategies and average market share. Sources: FENABRAVE; Quatro Rodas; UOL Carros; the authors.

In biology, strategies are called *phenotypes* and they are constrained by genetic, physical, and environmental factors. In business, "genetic" constraints are linked to a firm's history, culture, brand policy and image. For instance, styling decisions are influenced by a firm's *DNA* (identity) and history. Physical constraints are related to the possession of material, technical, and economic resources. Environmental restraints can have legal, social, ecological and market origins. Some attributes are harder to change in the short run. For instance, brand building demands continuous and long-lasting technical and marketing efforts – it is a short-term invariant. On the other hand, prices can be adjusted in a relatively short time.

Strategic attributes are elements of a product's *phenotype,* they tend to be stable over its life cycle, and are transmitted by *inheritance*. But some attributes can suffer mutations, like modifications or updates in product content. Some mutations are almost imperceptible – e.g., model-year detail updates, small running changes; others have significant impact in market fitness, like mid-life actions with substantial changes in styling and content. The Ecosport, launched in January 2003, was strong in attributes like *novelty, styling, agility*, and relatively affordable *price*. It received a mid-life facelift in October 2007, changing its front styling. In September 2012, a second generation was launched – a descendant that superseded the previous Ecosport. Most attributes, like *comfort, convenience, agility,* and *price* positioning, were inherited. The first generation disappeared as a player, but strategy persisted in the next generation. However, the styling was *mutated,* and its impact was enough for a 75% increase in market share, from 2012 to 2013 (table 1). A mid-life update in August 2018 brought a significant mutation in several previously weak attributes – *equipment, finish*, and *infotainment* technology – to face the *invasion* of new entrants in the strategic group (Honda HRV, Jeep Renegade, Nissan Kicks, and others). The Ecosport was discontinued in 2021, becoming extinct in the marketplace, as Ford Motor Company ceased automobile production in Brazil.

Fitness is the relative density of a product strategy and is expressed by stable and persistent market share. Competition for customer preference is the *natural selection* mechanism that determines fitness. The small SUV segment has been growing continuously and the entrance of new players have expanded the segment size - participation in the automobile market has been roughly proportional to the number of players (Fig. 4). However, players do not have an "easy life". The intensity of competition is not a function of the number of players, but of the number of interactions – i.e., it increases by the square of the number of players (N^2) (Sorensen 2000b). This explains some players experiencing significant decline in market share, despite the expansion in the overall segment size (table 1).

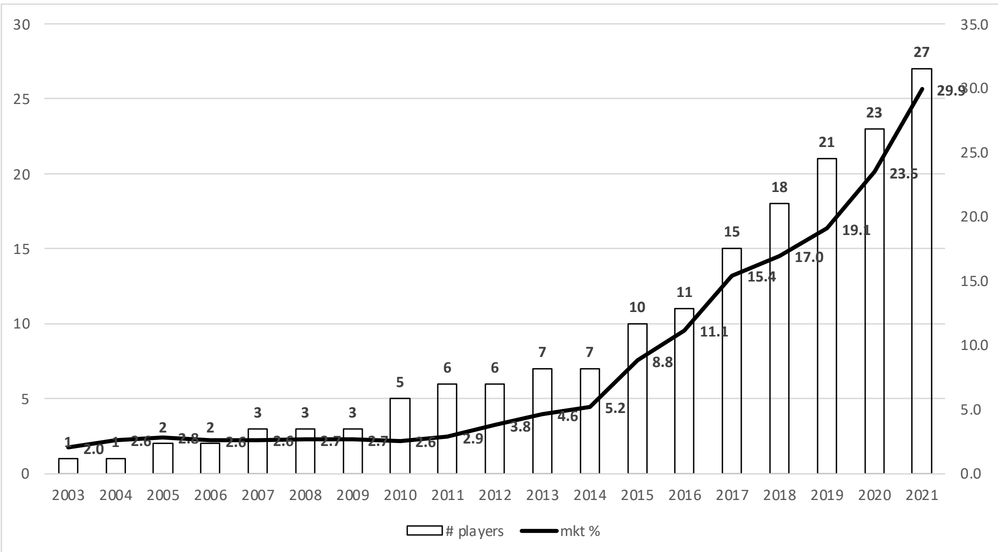


Fig. 4 Number of small SUV models vs market share. Source: FENABRAVE.

If all players adopted the exact same strategies (under the same conditions), market share should be the same for each player. They would all be equal in the *majority*, with individual market share $ms = \frac{m}{N}$ (*ms* = market share; *m* = market size; *N* = number of players). Historical data shows a small SUV holds 0.94% of the automobile market on average (Table

1). Players acting with different and advantageous strategies should enjoy $ms > 0.94\%$. From table 1, winners in the small SUV game (the *minority*) are Ecosport, Duster, HRV, Renegade, Kicks, Compass, Creta, and TCross.

Strong attributes are scored four or five points in the simulation. Close examination of minority strategies reveals differentiation, but also common traits. The CATPCA analysis resulted in attributes clustered into three major dimensions (Fig. 5), explaining 76,0% of data variance. Minority players are strong in attributes like *comfort*, *convenience*, *finish*, *equipment*, and *infotainment* (table 2), which are clustered in dimension 1. Equipment and infotainment technology are closely linked and related to price – more expensive vehicles tend to be better fitted. The same is valid for comfort, finish, and safety – distinction in those attributes tends to be followed by higher prices. *Space* and *trunk* go in direction opposite to *economy* and *agility* (dimension 2): small vehicles tend to be more fuel efficient and agile, while losing in passenger and luggage space in comparison to larger products. Those attributes are not particular to minority players. However, combinations of affordable prices, space, and trunk (e.g., Duster), or economy and agility (Ecosport, Kicks), seem to be winning propositions. A strong brand is an expression of legitimacy, and a long time is needed to gain recognition (Sorensen 2000a). Five of the eight minority players exhibit strong brands; small market share is related to less endowed brands (Table 2). *Robustness* goes beyond physical sturdiness, being also related to *brand* image (dimension 3). It is not just resistance to wearing and breaking, or off-roading capability, but it is linked to *reliability* (although the Jeep brand is strongly associated with off-road vehicles).

The analysis could go further. However, the above explanation should be sufficient as the objective of the automobile case study was merely to exemplify and to illustrate how the evolutionary game and minority rule frameworks can assist to understand strategic patterns and behavior.

	Dimension		
	1	2	3
finish	**0.961**	-0.074	0.097
comfort	**0.892**	-0.229	-0.083
safety	**0.880**	0.176	0.120
convenience	**0.831**	-0.006	0.286
style	**0.810**	-0.050	0.137
infotain	**0.792**	-0.040	-0.031
equipment	**0.789**	0.104	-0.295
price	**-0.774**	0.411	-0.088
perform	0.543	0.041	0.224
novelty	0.438	0.181	-0.238
space	0.133	**-0.938**	0.193
agile	0.130	**0.933**	0.180
trunk	0.070	**-0.889**	0.193
economy	0.012	**0.841**	-0.160
brand	0.082	-0.127	**0.969**
robust	0.096	-0.160	**0.964**
Cronbach Alpha	0.898	0.784	0.664
Eigenvalue	6.235	3.598	2.335
Variance Explained (%)	38.968	22.489	14.593

Variable Principal Normalization.

a. Rotation Method: Varimax with Kaiser Normalization. Rotation failed to converge in 4 iterations. (Convergence = 0.000).

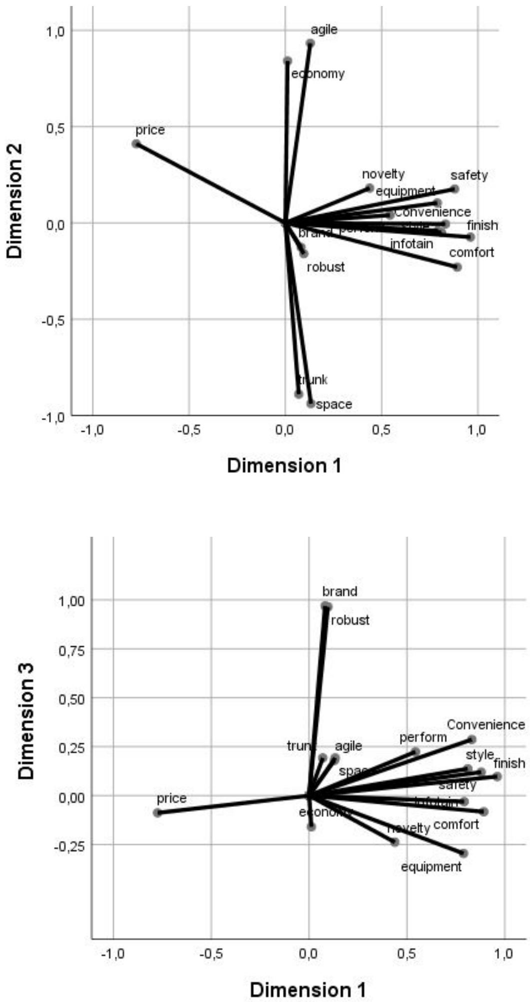


Fig. 5 Categorical Principal Component Analysis (CATPCA) rotated component loadings (left) and dimensions biplot (right).

5. Conclusion and future research

The evolutionary game framework applied to an automobile market case provided valuable insights. Among the several product strategy features, some can be enhanced in short or medium terms, like pricing. Others are much harder – they can be invariant for a single product generation, like styling, interior space, and others. Brand is an expression of legitimacy, not easy to be acquired. Once possessed, it tends to persist. It is worth to invest in an image of robustness – which tends to be associated with reliability – for a firm's long-term product success. Multiple attributes tend to be related in clusters (e.g., comfort, convenience, equipment, and infotainment technology) and – by selecting and focusing on them – firms benefit from synergies and efficient use of resources.

The case study was an illustration of both evolutionary games and the minority rule. It showed the framework is a useful tool to analyze an industry systematically, extracting focal points for product strategy making. It highlights the attributes a firm should pay attention to, in different time frames. Some actions can be specific to a single product, but others will require a continued effort that spans several product generations. It is worth remembering the aim of evolutionary games is the long-run persistence of market fitness.

The investigation was based on a single industry, limiting its capacity of generalization. But it showed the approach – identifying the elements from an evolutionary game perspective and constructing the theoretical framework – is meaningful to assist strategy decisions. Investigations involving other business settings and a variety of other algorithmic tools to generate the fitness functions would be valuable future research contributions.

References

Arthur, WB (1994) Inductive reasoning and bounded rationality. Am Econ Rev 84(2):406-411. http://www.jstor.org/stable/2117868

Arthur, WB. (1999) Complexity and the economy. Science 284(5411):107-109. http:/doi.org/10.1126/science.284.5411.107

Bazanini R, Berton R (2011) Estratégias e geração de vantagem competitiva no mercado de veículos off-road: análise fenomenológica do lançamento do modelo Ford Ecosport no Brasil (Strategies and competitive advantage generation in the off-road vehicle market: phenomenological analysis of the Ford Ecosport launch in Brazil). Revista Brasileira de Marketing 10(1):123-150.

Bischi GI, Merlone U (2017) Evolutionary minority games with memory. J Evol Econ 27(5):859-875. https://doi.org/10.1007/s00191-017-0526-4

Bottazzi G, Devetag G (2007) Competition and coordination in experimental minority games. J Evol Econ 17(3):241-275. https://doi.org/10.1007/s00191-006-0040-6

Brown JS, Vincent TL (1987) A theory for the evolutionary game. Theor Popul Biol 31(1):140-166. https://doi.org/10.1016/0040-5809(87)90026-8

Camerer CF (1991) Does strategy research need game theory? Strateg Manag J 12(S2):137-152. https://doi.org/10.1002/smj.4250121010

Campos CI, Pitombo CS, Delhomme P, Quintanilha JA (2020). Comparative analysis of data reduction techniques for questionnaire validation using self-reported driver behaviors. J Safety Res 73:133-142. https://doi.org/10.1016/j.jsr.2020.02.004

Casti JL (1996) Seeing the light at El Farol: a look at the most important problem in complex systems theory. Complexity 1(5):7-10. https://doi.org/10.1002/cplx.6130010503

Challet D, Zhang YC (1997) Emergence of cooperation and organization in an evolutionary game. Phys A: Stat Mech Appl 246(3-4):407-418. https://doi.org/10.1016/S0378-4371(97)00419-6

Challet D, Zhang YC (1998) On the minority game: Analytical and numerical studies. Phys A: Stat Mech Appl 256(3-4):514-532. https://doi.org/10.1016/S0378-4371(98)00260-X

Challet D, Chessa A, Marsili M, Zhang YC (2001) From minority games to real markets. Quant Finance 1(1):168-176. https://doi.org/10.1080/713665543

Challet D, Marsili M, Ottino G (2004) Shedding light on el farol. Phys A: Stat Mech Appl 332:469-482. https://doi.org/10.1016/j.physa.2003.06.003

Challet D, Marsili M, Zhang YC (2005). Minority games: interacting agents in financial markets. Oxford University Press.

Chatterjee K, Lilien GL (1986) Game theory in marketing science uses and limitations. Int J Res Mark 3(2):79-93. https://doi.org/10.1016/0167-8116(86)90012-1

Devetag G, Pancotto F, Brenner T (2014) The minority game unpacked. J Evol Econ 24:761–797. https://doi.org/10.1007/s00191-014-0351-y

Eisenhardt KM (1989) Building theories from case study research. Acad Manage Rev 14(4):532-550. https://doi.org/10.5465/amr.1989.4308385

FENABRAVE – National Federation of Brazilian Vehicle Distributors. (2020). New registrations. http://www.fenabrave.org.br/Portal/conteudo/emplacamentos

Hoffmann W, Lavie D, Reuer JJ, Shipilov A (2018) The interplay of competition and cooperation. Strat Manag J 39(12):3033-3052. https://doi.org/10.1002/smj.2965

IBM (2020) Categorical Principal Component Analysis (CATPCA) in IBM SPSS Categories 25. https://www.ibm.com/support/pages/ibm-spss-statistics-25-documentation

Kleindl B (1999) A game theoretic perspective on market-oriented versus innovative strategic choice. J Strateg Mark 7(4):265-274. https://doi.org/10.1080/096525499346378

Kuroda M, Mori Y, Masaya I, Sakakihara M (2013) Alternating least squares in nonlinear principal components. Comput Stat 5(6):456–464. https://doi.org/10.1002/wics.1279

Li J, Jiao J, Tang Y (2019) An evolutionary analysis on the effect of government policies on electric vehicle diffusion in complex network. Energy Policy 129:1-12. https://doi.org/10.1016/j.enpol.2019.01.070

Linting M, Meulman JJ, Groenen PJ, Van Der Koojj AJ (2007) Nonlinear principal components analysis: Introduction and application. Psychol Methods, 12(3):336. https://doi.org/10.1037/1082-989X.12.3.336

Linting M, Van Der Koojj A (2012). Nonlinear principal components analysis with CATPCA: A tutorial. J Pers Assess 94(1):12–25. https://doi.org/10.1080/00223891.2011.627965

Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM J Appl Math 11(2):431-441. https://doi.org/10.1137/0111030

Mintzberg H (1994) The fall and rise of strategic planning. Harv Bus Rev 72(1):107-114.

Porter ME (1979) How competitive forces shape strategy. Harv Bus Rev 57(2):137-145.

Porter ME (2008) The five competitive forces that shape strategy. Harv Bus Rev 86(1):25-40.

Postrel S (1991) Burning your britches behind you: can policy scholars bank on game theory? Strateg Manag J 12(S2):153-155. https://doi.org/10.1002/smj.4250121011

Quatro Rodas (2021). Vehicle tests. https://quatrorodas.abril.com.br/testes

Ranadheera S, Maghsudi S, Hossain E (2017) Minority games with applications to distributed decision making and control in wireless networks. IEEE Wirel Commun 24(5):184-192. http://doi.org/10.1109/MWC.2017.1600351WC

Ratneshwar S, Shocker AD, Cotte J, Srivastava RK (1999) Product, person, and purpose: putting the consumer back into theories of dynamic market behaviour. J Strateg Mark 7(3):191-208. https://doi.org/10.1080/096525499346422

Saloner G (1991) Modeling, game theory, and strategic management. Strateg Manag J 12(S2):119-136. https://doi.org/10.1002/smj.4250121009

Sharif P, Heydari H (2013) An introduction to multi-player, multi-choice quantum games: Quantum minority games & kolkata restaurant problems. In: Aberjel F et al. (ed) Econophysics of Systemic Risk and Network Dynamics. Springer, Milano, pp. 217-236.

Simon HA (1996). The Sciences of the Artificial. MIT Press, Boston.

Slater SF (1995) Issues in conducting marketing strategy research. J Strateg Mark 3(4):257-270. https://doi.org/10.1080/09652549500000016

Smith JM (1976) Evolution and the Theory of Games: in situations characterized by conflict of interest, the best strategy to adopt depends on what others are doing. Am Sci 64(1):41-45. https://www.jstor.org/stable/27847040

Smith JM (1982) Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Smith JM, Price GR (1973). The logic of animal conflict. Nature 246(5427):15-18. https://doi.org/10.1038/246015a0

Sorenson O (2000a) Letting the market work for you: an evolutionary perspective on product strategy. Strateg Manag J 21:577-592. https://doi.org/10.1002/(SICI)1097-0266(200005)21:5<577::AID-SMJ105>3.0.CO;2-C

Sorenson O (2000b) The effect of population-level learning on market entry: The American automobile industry. Soc Sci Res 29(3):307-326. https://doi.org/10.1006/ssre.1999.0671

Stacey, R. D. (1995). The science of complexity: An alternative perspective for strategic change processes. Strateg Manag J 16(6):477-495. https://doi.org/10.1002/smj.4250160606

UOL Carros (2021). Tests and evaluations. https://autolivraria.com.br/bc/teste-do-leitor/

Vincent TL (1985) Evolutionary games. J Optim Theory Appl 46(4):605-612. https://doi.org/10.1007/BF00939163

Vincent TL, Brown JS (2005) Evolutionary game theory, natural selection, and Darwinian dynamics. Cambridge University Press, Cambridge.

Vincent TL, Vincent TLS, Cohen Y (2011) Darwinian dynamics and evolutionary game theory. J Biol Dyn 5(3): 215-226. https://doi.org/10.1080/17513758.2010.526306

Weber J (2009) Automotive Development Processes. Springer-Verlag, Berlin.

Wilson DF (1999) Competitive marketing strategy in a volatile environment: theory, practice and research priorities. J Strateg Mark 7(1):19-40. https://doi.org/10.1080/096525499346512

Woodruff RB (1997) Customer value: the next source for competitive advantage. J Acad Mark Sci 25(2):139-153. https://doi.org/10.1007/BF02894350

APPENDIX D: The front-end of product development as systems thinking and predictive learning

Journal	Procedia Manufacturing
Authors	Charles Lincoln Kenji Yamamura; Celma de Oliveira Ribeiro; Denise Dantas; José Alberto Quintanilha; Fernando Tobal Berssaneti
Reference	Yamamura, C. L., Ribeiro, C. O., Dantas, D., Quintanilha, J. A., & Berssaneti, F. T. (2019). The Front-End of Product Development as Systems Thinking and Predictive Learning. *Procedia Manufacturing*, *39*, 1346-1353.
DOI	https://doi.org/10.1016/j.promfg.2020.01.323

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 00 (2019) 000–000

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

25th International Conference on Production Research Manufacturing Innovation:
Cyber Physical Manufacturing
August 9-14, 2019 | Chicago, Illinois (USA)

The front-end of product development as systems thinking and predictive learning

Charles L.K. Yamamura*[1], Celma O. Ribeiro[1], Denise Dantas[2], Jose A. Quintanilha[1], Fernando T. Berssaneti[1]

[1]*Polytechnic School of the University of São Paulo, São Paulo 05508-070, Brazil*
[2]*College of Architecture and Urbanism, University of São Paulo, São Paulo 05508-080, Brazil*

Abstract

Designing a new product, to be attractive to customers several years from its inception, is a major business challenge. The front-end of product development (the set of exploratory activities prior to project approval) is usually vague, unstructured and based on heuristics. Most supporting tools and methods interpolate data and are anchored in the present and the past. However, new product decisions are made for the future: predictive knowledge should be extrapolated from current information. This paper proposes the front-end should be akin to a scientific inquiry, with systems thinking and predictive learning as mainstays to create customer value and reduce its "fuzziness".

Keywords: *fuzzy front-end, predevelopment, concept design, systems thinking, prediction, statistical learning, machine learning, strategic decision making, set-based concurrent engineering, lean product development.*

1. Introduction

The front-end of product development (FEPD) comprises the activities preceding official approval and funding of a project - opportunity identification, idea generation and concept development [1]. It is often qualified as "fuzzy" because, in contrast to product development per se (clear, specific, systematic and deterministic), the front-end is vague, ambiguous, unstructured and probabilistic [2]. They are also called *predevelopment* activities and are often neglected, even if the fate of new products may be decided at this early stage [3].

Concept design is a key challenge in the FEPD. Product concept is a description of the technology, performance, features, and form of a product, or how it will satisfy customer needs [4]. It is an exclusive set of features (instances) - collectively, they

* Corresponding author. Tel.: +11-55-95062-0515.
E-mail address: charles.yamamura@usp.br

Abstract

Designing a new product, to be attractive to customers several years from its inception, is a major business challenge. The front-end of product development (the set of exploratory activities prior to project approval) is usually vague, unstructured, and based on heuristics. Most supporting tools and methods interpolate data and are anchored in the present and the past. However, new product decisions are made for the future: predictive knowledge should be extrapolated from current information. This paper proposes the front-end should be akin to a scientific inquiry, with systems thinking and predictive learning as mainstays to create customer value and reduce its "fuzziness".

Keywords: fuzzy front-end, predevelopment, concept design, systems thinking, prediction, statistical learning, machine learning, strategic decision making, set-based concurrent engineering, lean product development.

1. Introduction

The front-end of product development (FEPD) comprises the activities preceding official approval and funding of a project – opportunity identification, idea generation and concept development [1]. It is often qualified as "fuzzy" because, in contrast to product development per se – clear, specific, systematic, and deterministic – the front-end is vague, ambiguous, unstructured, and probabilistic [2]. They are also called *predevelopment* activities and are often neglected, even if the fate of new products may be decided at this early stage [3].

Concept design is a key challenge in the FEPD. Product concept is a description of the technology, performance, features, and form of a product, or how it will satisfy customer needs [4]. It is an exclusive set of features (instances) – collectively they do not belong to other sets (concepts) [5]. It is difficult to map variables (concept features) and to make sense of them into coherent and integrated concepts. Product strategy is usually performed based on heuristics and businesses struggle to keep track of knowledge and to use it effectively. Relying on heuristics can sometimes bring reasonable judgment but may also produce systematic errors [6]. There is variability in the staff's individual knowledge, skills, and experience. Often, a firm's knowledge may not be available, be inaccessible or even be forgotten and lost [7].

It is not uncommon for a recently launched product to be matched or surpassed by a new competitor, raising the bar, and leaving the former behind in customer value, sales, and profitability. There may be a contradiction in the product strategy process. Most supporting tools are rooted in the present and the past – competitor benchmarking, QFD (Quality Function Deployment), conventional market research, brainstorming and others. They are analytical techniques to interpolate data. But new product development needs to create value for the future: it should be predictive and extrapolate from data [8].

Although there is an extensive literature on the FEPD, there are few studies on how knowledge is created in the process [9]. Most articles tend to focus on organizational aspects, tools, process models and competencies [10] [11]. Knowledge creation is usually assumed to be implicit in the formal process. This paper investigates the FEPD as a set of activities to create knowledge to understand and to predict customer needs and aspirations, using systems thinking and data-based simulation.

2. Method

This study is grounded in literature review, theoretical model construction and simulation. With the question "knowledge creation in the front-end of product development" in mind, a literature review on *fuzzy front-end*, *front-end, predevelopment, front-end of innovation* and variations was performed. A Scopus query yielded 920 documents and Web-of-Science, 362. Initial screening, based on title and abstract content, reduced the set to 308 items. Additional review on *systems engineering, decision theory, knowledge creation, theory construction, set-based concurrent engineering, lean product development* and *statistical learning* added 134 potentially relevant works. Detailed content analysis, according to rigor, relevance and alignment to the research topic, lead to the documents on which this work is based.

Once literature was scrutinized, key concepts related to the front-end knowledge creation and decision-making were identified and defined. Definitions were compiled for theoretical fit, rigor and conciseness. A central proposition was made and a theoretical model showing both conventional and systemic approaches were built. The usefulness of a theory is subsumed by its capacity to predict a behavior. A computer experiment using an artificial neural network was conducted to simulate predictive learning in an automobile development case. As previously stated, the aim was to understand the role of systematic predictive learning in the front-end of product development and to explain why and how it can lead to customer value creation.

3. Theoretical background

The expression *fuzzy front end* (FFE) was coined by Reinertsen in 1985 [12]. FFE is the phase from the emergence of a new product need to the decision to commit resources in the development [13]. Its purpose – in economic terms – is to maximize the expected value (EV) of betting in the project [12], which is determined by its probability of success *p*, the upside of success U and the downside of failure *D*: $\boxed{EV = p * U - (1 - p) * D}$. The FFE can increase EV by increasing the probability of success, increasing the upside, and reducing the downside. The latter can be accomplished by reducing uncertainty, getting better information about possible outcomes from existing information. Upside is increased by augmenting customer value.

Woodroof [14] defines customer value as the evaluation of product features and performance, leading to the satisfaction of customer needs. Customer learning is the process of reducing the gap between what the company knows about customers and what they really value. The scholar remarks new learning tools are needed for predicting customer value. Tools such as QFD and conjoint analysis are too narrowly focused on current design decisions and are lacking in predictive power. Cooper and Edgett [15] also identified compelling customer value as the most important reason for new product development success, others being "front loading" (anticipating problems early), spiral development (learning loops), holistic approach, continuous learning, portfolio management and an evolved stage-gate system.

Geyer, Lehnen, and Herstatt [16] conducted a survey to determine the most used FEPD support methods and why companies prefer or avoid certain methods. The selection criteria were more related to familiarity and ease of use than to pure efficiency. Companies tend to stick to traditional methods, such as internal data (e.g., sales reports), external reports (from consulting, technical associations etc.) and customer observation, whereas they should consider different methodologies to seek optimal results.

Khurana and Rosenthal [17] cautioned product concept is a hard-to-reach moving target: a prediction challenge. Most companies do not have a clear understanding of customer needs. However, success in the marketplace depends on how well the product concept matches customer value. Kim and Wilemon [18] remark well-defined concepts are essential to understand project requirements and risks, as well as to avoid ill-informed decisions. Since FFE conditions are vague and ambiguous, it is better to consider a comprehensive set of alternative solutions and to build a good information system on customers, competitors, technology, and market conditions. Murphy and Kumar [19] point out the dynamic and unstructured nature of the front end as a challenge to generalize research findings. A project investment decision is still largely guided by “gut feel”, suggesting firms are not fully aware of the evaluation methods available.

Jetter [8] also remarked companies relied on guessing instead of structured methods for handling the fuzzy front-end. The author accentuates a product concept will affect future customer experience not only at a still coming launch time but till the end of its disposal (a time span of about twenty years for some products, like automobiles). There is also uncertainty about future competitor moves. Tools such as QFD and market surveys are insufficient because they rely on historic knowledge and experience. She cites scenario building, knowledge mapping and system thinking as alternatives.

In a subsequent article, Schroeder and Jetter [20] continued to elaborate on FFE supporting tools. Among requirements for effective tools: capacity to handle uncertain, imprecise, and changing information; to process diverse information, turning tacit into explicit knowledge; to enhance information processing, avoiding oversimplification and decision bias. They emphasize a holistic systems-view as fundamental to provide critical understanding of dynamic relations and to encourage systemic learning and knowledge transfer among projects. The authors remark it is almost impossible to attribute an observed result (such as sales) to a certain decision (like a concept choice). Also impossible is to learn from decisions not chosen. Hence the role of simulation techniques, which support systemic learning by evaluating different decisions and their parameters. They encourage early information collection and problem solving through a wide range of hypotheses testing. They proposed systemic and holistic tools, capable of mapping fuzzy dynamic relations and extrapolating then into information about the future.

Eling and Herstatt [21] conducted a comprehensive literature review on the FFE state-of-art and concluded it still lacks a well-established conceptual framework and vocabulary, despite over three decades of existence. They identified the need for “a more holistic view on the topic of formalization”.

4. Conceptual definitions

System: a set of elements, such as products, people, information, knowledge, and other assets, combined to meet a need (adapted from NASA – National Aeronautics and Space Administration [22]). Elements within a system have mutual relations among them and with the environment [23]. Complex systems exhibit properties absent in the mere sum of their parts, the *emergent phenomena* [23];

Theory: system of concepts integrated into propositions, within a defined domain [24]. The goal of theory is to reduce the complexity of the world (i.e., reduce uncertainty), based on explanation and prediction [25];

Systems theory: a level of theoretical model-building that lies between pure mathematical constructions and theories in specialized disciplines. It allows moving back and forth from both worlds of Platonic theory and fuzzy practice and facilitates interdisciplinary communication [26];

Systems engineering: a logical way of thinking to design, operate, manage and dispose a system (adapted from [22]);

Concept: formally defined idea, capturing the essence of its meaning [25]. A unique mapping of arguments into values of a function (adapted from [5]);

Proposition: causal relationship explaining how and why concepts are related [25];

Domain: context of application (when and where) of a concept [25];

Data: raw facts and numbers [7];

Information: interpreted data [7];

Knowledge: information assimilated by a person [7];

Uncertainty: gap between information needed to perform a task and information that is available. It is a function of number of outputs, number of inputs and the level of performance necessary for the task [27];

Judgment: process of determining payoffs of decisions in particular situations. Without prediction, the decision maker is forced to a blind decision (*heuristic*). In the absence of prediction, the value of judgment decreases. As the cost of predictions decreases with the development of statistical learning algorithms, demand for decision making will increase and so will the value of human judgment [28];

Prediction: rigorous extrapolation of new information from existing information [28]. Good predictions require sound theoretical understanding and reliable data on initial conditions. The objective of prediction is “not to forecast but to construct alternative scenarios for the future and to analyze their sensitivity to error in both theory and data” [5];

Simulation: technique for understanding and predicting the behaviour of a system. Simulations are only as good as their assumptions [5];

Decision: choice derived from a judgment [28];

Learning: reduction in the gap between needed and available knowledge [14];

Statistical learning: techniques for inferring, modeling, and predicting from complex data sets [29] [30]. It covers the disciplines of statistics, data mining, and predictive analytics [31] [32];

Supervised statistical learning: building statistical models for prediction, based on examples (data labels) [30];

Data mining: divided into two phases: exploratory, where data is used to create knowledge, to improve decision-making and to provide customer value; and testing, to confirm relationships discovered in the previous step [32];

Predictive analytics: uses relations confirmed from data mining to predict future customer behavior, an event involving many variables [32];

Value: capability to satisfy a stakeholder’s need, created by the interaction of system elements in addition to their individual contributions (adapted from [22]). Better understanding of data leads to higher decision-making accuracy and to superior customer satisfaction [32];

Product development: process of transforming market and technology data into information that reduces uncertainty about customer needs, competition, and technology, raising the likelihood of success [33].

5. Proposition and model

Scientific management is a systematic way or a set of methods for understanding and solving business problems [34]. It is the application of analytic techniques to processes, analysis of activities, splitting the problem into smaller blocks, elimination of unnecessary ones, grouping and sequencing those activities. In the axiomatic modality, research is driven by an idealized model. Assumptions, variables, derived solutions, and insights are delimited within the scope of a model. Axiomatic research is usually *normative*, aiming at creating strategies, policies, and actions, or to optimize a newly defined problem [34].

Fig.1. Conventional FEPD.

Managers usually interpret information and rely on "gut feel" to make product decisions [17] [19] (fig.1). They certainly strive to extract meaning from data and information on customers, competitors, and technology, but the process is usually vague, subjective, and heuristic. On the other hand, a systems approach can discipline the FEPD [8] [20] [35]. More than a single problem-solving event, it turns into theory-building, a knowledge creation activity [35] not unlike the work of a scientist in the laboratory [36]. Learning is an iterative process that produces knowledge that can be abstracted and transferred into future projects [37] [36]. A judgment without prediction is mere "guessing" [28]. Simulation and prediction (extrapolation of new information using simulation) increase the value of decision-making [28]. Fig. 2 shows the model integrating both theory and prediction in product strategy decision-making.

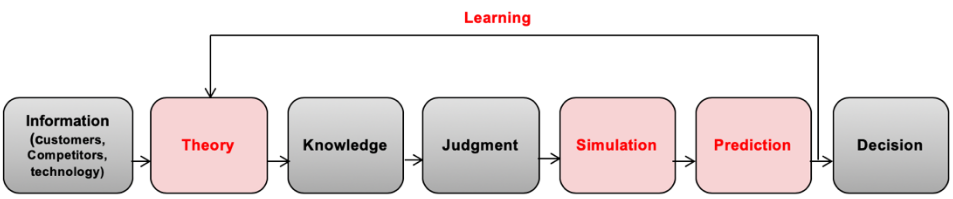


Fig.2. Systemic FEPD.

Proposition: *systems-thinking and predictive learning can reduce the uncertainty of the front-end and drive customer value creation in product development.*

6. Simulation

Vapnik and Izmailov [38] proposed using the intelligence of a "teacher" to accelerate statistical learning. Human learning requires far fewer examples than machine learning, due to the use of "intelligence" (in this case, teacher's intelligence) - in contraposition to brute force. A teacher provides *privileged information* - explanations, comments, and metaphors. The authors claim privileged information is ubiquitous and can be found in almost any problem. They propose using both data and *statistical invariants* to drive learning [39]. The idea of invariants finds an analogy in the role of a good instructor or expert to guide the learning process.

Many potential variables or features of a problem are irrelevant for the task at hand. An expert (teacher) helps to construct statistical invariants latent in the problem. Based on their knowledge and experience, experts offer "shortcuts", called *predicates*. Instead of initializing from a blank sheet of paper, the procedure starts by building on the expert's wisdom. Vapnik and Izmailov [39] exemplify (jocosely) with the "duck test": if it looks like a duck, swims like a duck and quacks like a duck, it probably is a duck (three predicates to qualify a duck, instead of many variables). The difference between invariants and features is that increasing invariants leads to more accurate predictions while the opposite is true for an increase in features, in this case requiring more training data.

For the experiment, a Brazilian automobile market segment was chosen: small SUVs (sport utility vehicles). Customer data collection started with semi-structured interviews and "storytelling". Six SUV owners were surveyed (three interviews, three storytelling). The interviews followed a predesigned script but were relatively flexible and lasted for about an hour each. In storytelling, users were asked to freely write an essay on their impressions and experiences. Interpretation of insights from those techniques yielded thirteen key value attributes, or independent variables (concept features): style, "ruggedness", space, trunk, comfort, "nimbleness", versatility, finish, features, connectivity, performance, economy, and safety. Although not explicitly mentioned, three features - novelty, brand, and price perception – were added. The selection of features, according to a researcher's domain knowledge and experience plays the role of the expert's invariant selection.

Usually, additional data implies better learning, provided enough processing power and adequate algorithms are available. But data hunger is a problem in many real-life applications. The volume of data required is roughly proportional to the complexity of the model (number of independent variables). Sometimes, enough samples may not be available to model the problem. Hence, handling data should start with strategy and intelligence, using invariants and predicates. First, identifying the problem; then, judiciously selecting the model features and determining which data is needed. Human judgment partially substitutes for data scarcity, by carefully selecting the features most likely to be relevant.

Market share was chosen as the product competitiveness indicator (dependent variable). Sales data (vehicle registrations) were collected from the FENABRAVE - National Federation of Brazilian Vehicle Distributors - website (www3.fenabrave.org.br). The seventeen most representative products were selected for monthly sales figures from January 2013 to December 2018. Market share is the ratio between product sales and total market size. Using market share instead of straight sales purges (or at least reduces) effects of market seasonality and economic fluctuations, when comparing data from different points in time. Market share is the target value *y*, to be compared to estimated value *y-prime* for calculating errors (residuals) in the functions.

To synthesize the sample vectors, each product was evaluated according to the sixteen attributes (independent variables). Most sample vectors are recursive since their feature values are constant over time. However, some vectors may differ for the same product in different moments of time. For instance, "novelty" will decrease as time passes by. A product may be upgraded or customer perception may change over its life cycle. In a large company (e.g., an automobile manufacturer), such evaluation would rely on several resources: expert knowledge, proprietary research, static and dynamic testing, "tear down" etc. For this small-scale experiment, reviews from a specialized automobile magazine (https://quatrorodas.abril.com.br) were used as a proxy for part of the "expert knowledge and judgment". A scale from one to five was used to evaluate attributes. Small SUV market share (target variable *y*) between 2013 and 2018 has ranged from about 0.002 to 0.029. For normalization, they were scaled by a factor of 200 to bring them close to the independent variable *x* range.

The configuration of vector $\boldsymbol{x}$, with features $\boxed{x_i}$ is the product concept proposition from the "expert". They represent a selected a priori finite set of predicates, which replace a potentially infinite set of functions (all possible product features, attributes, and combinations). Selected predicates are also invariants in the creation of customer value: they are not individually affected in importance for the customers (which is not the same as the technical compromises among product attributes that sometimes occur in practice).

Every concept feature must be clearly defined (e.g., "economy" means fuel autonomy in kilometers per liter, split between 70% of urban and 30% of road use; "trunk" stands for volumetric luggage capacity and easy loading access, etc.). A concept is given by $\boxed{f(x, x^*, \alpha)}$, where x is the feature variable, $\boxed{x^*}$ is the expert's privileged information and α , variable parameter weight (calculated by the algorithm). As far as experts study customers, analyze relations among predicates, understand how they meet customers' needs and aspirations and make those propositions clear and explicit, they are constructing a theory of customer value, to be tested in simulations, to predict the market acceptance (measured by market share) of a hypothetical product.

In the process, 765 vector samples with sixteen independent feature variables (concept attributes) and target dependent values were synthesized. They were split in an 80:10:10 proportion among training, test, and validation sets (they yielded slightly better performance than the default 70:15:15). Due to the relatively small sample size, a "shallow" neural network with one hidden layer was used.

There is no consensus in the literature concerning the number of units in a hidden layer. A common bounding heuristic is between the ratio of input and output layer sizes and less than twice the input layer size (p). In the experiment, architectures ranging from p/2 to 2p nodes (8, 10, 12, 16, 20, 30) were tried, yielding correlations between 0.89 and 0.95 - the best result was achieved with ten units. The hidden layer was activated by a hyperbolic tangent function $g(z) = tanh(z)$. Target values are activated by the ReLU (rectified linear unit) function $f = max (0, x)$.

As most nonlinear optimization techniques, the Levenberg-Marquardt method iterates to find the minimum of a multivariate function. It seeks to minimize the sum of squared errors in the nonlinear function and acts as a combination of a Taylor series approximation (Newton) and the steepest descent methods [40]. The experiment was run on MATLAB™ R2018a, in a MacBook™ with 1.2 GHz Intel™ Core M processor.

7. Results

The quality of result in a computer simulation is determined by its optimality – the extent a result achieved is the best possible [34]. The experiment yielded overall correlation coefficient R = 0.9485, after 15 epochs of 1,500 iterations. The correlation R for testing was 0.9612 and for validation, 0.9579. It was a small-scale experiment, with sixteen independent variables, 765 samples and a relatively coarse evaluation scale with discreet values. Nevertheless, it shows a customer-centered approach using expert knowledge and product concept simulation can produce significant results.

The combination of qualitative market research, evaluations from a specialized magazine and the researcher's judgment and experience were used to generate features $\boxed{(x, x^*)}$, which played the role of predicates and invariants ("privileged information" [39]) supplied by an expert. Features $\boxed{x_i}$ were combined into a vector design $\boldsymbol{x}$. The functions $\boxed{y = f(x, x^*)}$ are the set of hypotheses or the customer value theory in the project. The learning algorithm provided simulation and prediction based on that theory and yielded statistical validity. The acquired knowledge from the specific case can be abstracted to application in other settings (projects).

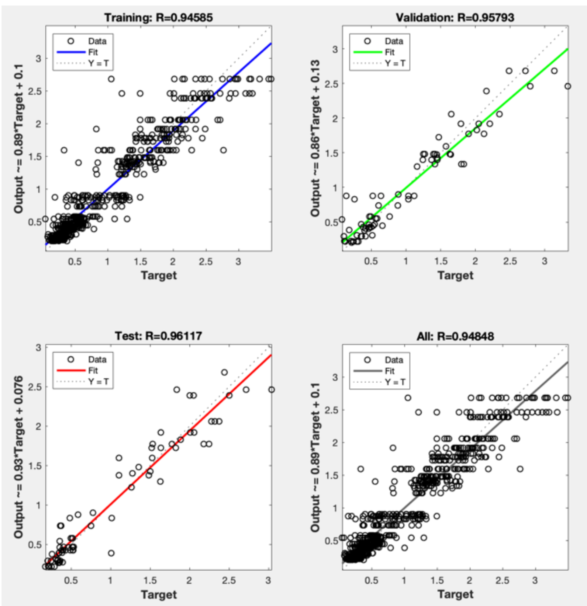


Fig 3. Correlation fits (MATLAB™).

8. Conclusion

Current product predevelopment practice may not be adequate. There is too much variability and loss of experience and knowledge. There is also lack of systems-view and integration. This article tackles a shortfall in the FEPD and suggests a systems approach supported by predictive learning. It integrates methodological rigor, human judgment, statistical learning, and customer-centered view. The human expert is an active participant in the learning process, neither a mere "guesser" nor just a data provider. A stable set of functions (the artificial neural network) becomes a knowledge repository and a tool for market simulation: customer value, product attributes, technical features and specs, sales, prices, competitive data, market history, technology evolution are contemplated.

Machine learning extracts relationships that are difficult to map mentally or heuristically, but usually needs large amounts of training data. Data hunger is the Achilles' heel of supervised learning [41]. Manufacturing turned-digital companies usually do not have access to the huge volume of data native digital companies do. This study suggests a way to partially reduce the substantial need for data, by substituting human judgment. By feeding qualitative information and expert domain knowledge, it is possible to preclude the need for massive amounts of data, which would be required in the case of blind data feeding. The objective was to pre-select key customer values and use statistical learning to weigh and to refine the relations among concept attributes. Narrowing the range of potential research features - which would otherwise be extensive and demand a proportionally larger amount of training data - leads to increased efficiency and speed. In a certain way, data quality was traded for quantity.

The proposed model is an extension of substantive existing literature. It contributes by offering a holistic problem-solving framework, integrating both systems thinking and statistical learning, to explain and to predict in the FEPD. The learning process produces knowledge that can abstracted and transferred to other projects. However, the simulation was based on a single industry (automobile manufacturing) – a limitation. An artificial neural network was chosen as the simulation algorithm. However, that does not mean other tools - such as SVMs (support vector machines) - could not be equally or more effective. Future studies comparing the efficiency of alternative methods could be useful contributions. Overall, there is a need for more research on the FEPD as a knowledge creation process, more systemic and less fuzzy: a customer value creation laboratory.

References

[1] P.A. Koen et al., Providing clarity and a common language to the "fuzzy front end, Research-Technology Management, 2(2001) 46-55.

[2] D. Reinertsen, P.G. Smith, The strategist's role in shortening product development, Journal of Business Strategy, 4(1991) 18-22.

[3] R. G. Cooper, Predevelopment activities determine new product success, Industrial Marketing Management, 17(1988) 237-247.

[4] K. Ulrich, S. Eppinger, Product design and development, Mc Graw-Hill, 6th ed., 2015.

[5] H.A. Simon, The sciences of the artificial, The MIT Press, 3rd ed., 1996.

[6] D. Kahneman, A. Tversky, The psychology of prediction, Psychological Review, 4(1973) 237-251.

[7] M. Alavi, D.E. Leidner, Knowledge management and knowledge management systems: conceptual foundations and research issues, MIS Quarterly, 1(2001) 107-136.

[8] A.J. Jetter, Educating the guess: strategies, concepts and tools for the fuzzy front end of product development, Management of Engineering and Technology, IEEE Portland International Conference, (2003) 261-273.

[9] H. Akbar, N. Tzokas, An exploration of new product development's front-end knowledge conceptualization process in discontinuous innovations, British Journal of Management, (2012) 1-19.

[10] L.M. Haase, L.N. Laursen, Reasoning in the fuzzy front end of innovation: framing the product DNA, International Journal of Innovation Management, 5(2018) 1-22.

[11] A.R. Pereira, J.J.P. Ferreira, A. Lopes, Front end of innovation: an integrative literature review, Journal of Innovation Management, 1(2017) 22-39.

[12] D. Reinertsen, Taking the fuzziness out of the fuzzy front end, Research-Technology Management, 6(1999), 25-31.

[13] P.G. Smith, D. Reinertsen, Shortening the product development cycle, Research-Technology Management, 3(1992) 44-49.

[14] R.B. Woodruff, Customer value: the next source of competitive advantage, Journal of the Academy of Marketing Science, 2(1997) 1139-153.

[15] R.G. Cooper, S.J. Edgett, Maximizing productivity in product innovation, Research-Technology Management, 2(2008) 47-58.

[16] F. Geyer, J. Lehnen, C. Herstatt, Customer need identification methods in new product development: what works "best"?, International Journal of Innovation and Technical Management, 1(2018) 1-26.

[17] A. Khurana, S.R. Rosenthal, Integrating the fuzzy front end of new product development, IEEE Engineering Management Review, (Feb 1997) 35-49.

[18] J. Kim, D. Wilemon, Focusing the fuzzy front-end in new product development, R&D Management, 4(2002) 269-279.

[19] S.A. Murphy, V. Khumar, The front end of new product development: a Canadian survey, R&D Management, 1(1997) 5-15.

[20] H-H. Schröeder, A.J.M. Jetter, Integrating market and technical knowledge in the fuzzy front end: an FCM-based action support system, International Journal of Technology Management, 5(2003) 517-539.

[21] K. Eling, C. Herstatt, Managing the front end of innovation – less fuzzy, yet still not fully understood, Journal of Product Innovation Management, 6(2017) 864-874.

[22] National Aeronautics and Space Administration – NASA, Systems engineering handbook, NASA, (2016).

[23] R. Dekkers, Applied systems theory, Springer, 2nd ed., (2017).

[24] J.G. Wacker, A theory of formal conceptual definitions: developing theory-building measurement instruments, Journal of Operations Management, 22(2004) 629-650.

[25] S.B. Bacharah, Organizational theories: some criteria for evaluation, Academy of Management Review, 4(1989) 496-515.

[26] K.E. Boulding, General systems theory: the skeleton of science, Management Science, 3(1956) 197-208.

[27] J. Galbraith, Designing complex organizations, Addison-Wesley, (1973).

[28] A. Agrawal, J. Gans, A. Goldfarb, Prediction, judgment and complexity: a theory of decision making and artificial intelligence, National Bureau of Economic Research, (Jan 2018).

[29] V.N. Vapnik, Statistical learning theory, John Willey and Sons, (1998).

[30] G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to statistical learning with applications in R, Springer, (2015).

[31] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Springer, 2nd ed., (2009).

[32] J.F. Hair Jr., Knowledge creation in marketing: the role of predictive analytics, European Business Review, 4(2007) 303-315.

[33] R.K. Moenaert, A. De Meyer, W.E. Souder, D. Deschoolmeester, R&D/Marketing communication during the fuzzy front-end, IEEE Transactions on Engineering Management, 3(1995) 243-258.

[34] J.W.M. Bertrand, J.C. Fransoo, Modeling and simulation: operations management research methodologies using quantitative modeling, International Journal of Operations Management Research, 2(2002) 242-264.

[35] T. Keikonen, R. Takala, Product concept design, Springer, (2006).

[36] A.C. Ward, D.K. Sobek II, Lean product and process development, Lean Enterprise Institute, (2007).

[37] S. Thomke, T. Fujimoto, The effect of "front-loading" problem solving on product development performance, Journal of Product Innovation Management, (Mar 2000),128-142.

[38] V.N. Vapnik, R.Z. Izmailov, Learning using privileged information: similarity control and knowledge transfer, Journal of Machine Learning Research, 16(2015) 2023-2049.

[39] V.N. Vapnik, R.Z. Izmailov, Rethinking statistical learning theory: learning using statistical invariants, Machine Learning, Springer, published online 18 July 2018.

[40] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, Journal of the society for Industrial and Applied Mathematics, 2(1963) 431-441.

[41] A. Ng, What artificial intelligence can and can't do right now, Harvard Business Review, Nov 9, 2016.