Closed System Precepts in Systems Engineering for Artificial Intelligence - SE4AI

Niloofar Shadab

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Industrial and Systems Engineering

Peter A Beling, Co-chair

Alejandro Salado, Co-chair

Taylan G Topcu

Daniel Hoek

Tyler M Cody

December 8, 2023

Blacksburg, Virginia

Keywords: Systems Theory, Systems Engineering, Artificial Intelligence, Closed Systems Precepts, Core and Periphery, Outcome-based Engineering, Scenario-based Engineering

Closed System Precepts in Systems Engineering for Artificial Intelligence - SE4AI

Niloofar Shadab

(ABSTRACT)

Intelligent systems ought to be distinguished as a special type of systems that require distinctive engineering processes. While this distinction is informally acknowledged by some, practical systems engineering (SE) methodologies for intelligent systems remain primarily rooted in traditional SE paradigms centered around component aggregation.

Initially, this dissertation posits that the traditional approach is grounded in the notion of open systems as the fundamental precept, whereas engineering intelligent systems necessitates an alternative approach founded on the principles of closed systems. This dissertation endeavors to identify potential gaps within the current SE foundations concerning the accommodation of the unique characteristics of intelligent systems, such as continuous learning and sensitivity to environmental changes. Furthermore, it argues for the mitigation of these gaps through the formalization of closed systems precepts. It adopts a systems-theoretic perspective to elucidate the phenomena of closed systems and their intricate interplay with engineering intelligent systems. This research contends that, given the intricate coupling between intelligent systems and their environments, the incorporation of closed systems precepts into SE represents a pivotal pathway to construct engineered intelligence. In pursuit of this objective, this dissertation establishes a formal foundation to delineate closed systems precepts and other fundamental practices. Subsequently, it provides formalism to discern two important categories of closed systems, informationally and functionally closed systems, and their relevance in the domains of engineering and design across diverse levels of system

abstraction. Additionally, it explores the practical application of the closed systems precepts through the novel paradigm of *core and periphery*, followed by its examination within real-world contexts. This dissertation is organizes as follows:

Chapter 1 initiates the dissertation by presenting the problem formulation and motivation. It subsequently delves into a thorough literature review and outlines the research's scope and objectives, contributing to the essence of this work.

In Chapter 2, a narrative unfolds, elucidating the contributions of the provided papers to the objectives outlined in Chapter 1. This chapter illuminates how each paper aligns with and furthers the overarching goals set forth in the Chapter 1.

Chapter 3 serves as a culmination, offering a summary of the accomplishments, acknowledging limitations, and delineating potential avenues for future research within this domain.

Paper A is devoted to substantiating the closed notion of intelligence property. In the realm of artificial intelligence (AI), systems are often expected to exert influence upon their environments and, reciprocally, to be influenced by their surroundings. Consequently, a profound interdependence exists between the system and its environment, transcending the confines of conventional input-output relations. In this regard, Paper A postulates that the engineering of intelligent systems mandates an approach that elevates closed systems as foundational precepts for characterizing intelligence as a property contingent upon the system's relationship with its context. The ensuing discussion will juxtapose the viewpoints of open and closed systems, illustrating the limitations of the open system perspective in representing intelligence as a relational property. In response, this paper will advocate for the adoption of the closed system view to establish intelligence as an inherent relational property arising from the system's dynamic interactions with its environment.

Paper B is dedicated to the formalization of the closed systems paradigm within SE. In this paper, formalism is proffered for the closed systems precepts, drawing upon systems theory, cybernetics, and information theory. A comprehensive comparison of two closure types, informational and functional closure, within closed systems is presented, underpinned by a common systems-theoretic formal framework. This dissertation contends that by grounding these initiatives in the core and periphery concept, we can facilitate the design and engineering of intelligent systems across multiple levels of abstraction. These levels may span a spectrum from informational closure to a synthesis of informational and functional openness. It posits that this approach represents a versatile, method-agnostic solution to some of the principal challenges encountered when engineering multiple tiers of intelligence for complex systems.

Paper C delves into the rise of the concept of core-periphery from some cybernetics principles, such as variety and "The Law of Requisite Variety" and provides a formalism that is a derivation of the mentioned principles in Cybernetics. Later, it elaborates on the practical implications of such concepts in intelligent systems from biological systems and entails an engagement with a CNN model to explore the core and periphery concept within AI-enabled systems.

Paper D proposes the practical implementation of the closed systems doctrine in SE, offering frameworks that rigorously define the boundaries between closed systems and their environment. These frameworks are meticulously designed to account for stakeholder requirements and the inherent design constraints of the system. This paper illustrates practical applications of informational and functional closure within SE processes, leveraging a hypothetical example for elucidation. It focuses on two aspects of engineering intelligence, scope and scale to provide a platform for the utilization of closed systems precepts.

Closed System Precepts in Systems Engineering for Artificial Intelligence - SE4AI

Niloofar Shadab

(GENERAL AUDIENCE ABSTRACT)

There has been a longstanding call within the Systems Engineering (SE) community for the development of a comprehensive SE theory. This endeavor seeks to bestow upon the field of SE the requisite credibility to stand autonomously as an engineering discipline, capable of addressing the contemporary engineering challenges that confront us. In the pursuit of establishing SE as a distinct engineering field, it becomes imperative to furnish precise and formal definitions for the fundamental concepts that underpin SE processes. Presently, the absence of concrete formalism and clear distinctions surrounding certain core concepts introduces ambiguity into various SE practices. Until recently, the immediate necessity for such foundational formalism was not universally acknowledged or appreciated, as engineers predominantly relied on established practices to design traditional engineered systems. These conventional SE practices had withstood the test of time, until the emergence of a new generation of complex systems characterized by distinctive features. Among these emergent systems, Artificial Intelligent (AI) systems have garnered significant attention, bearing unique attributes that call into question the adequacy of the current SE practices in supporting their development.

Consequently, it has been asserted that intelligent systems necessitate the incorporation of new characteristics that render them incompatible with conventional SE practices. This assertion underscores the need for a thorough reevaluation of SE, potentially entailing an expansion of the formalism underpinning its fundamental principles. However, despite these

pressing concerns, SE currently lacks a solid theoretical foundation capable of facilitating a paradigm shift away from current practices. The primary objective of this dissertation is to identify the existing gaps responsible for the misalignment between the characteristics of AI systems and prevailing SE practices. Additionally, it seeks to propose innovative methodologies to bridge these gaps effectively. In alignment with this objective, the dissertation provides formalism for these methodologies. Finally, this dissertation aims to provide practical implication of such formalism to validate their applicability.

In summary, the central research question, along with the ensuing objectives of this dissertation, can be articulated as follows:

- What aspects of SE are insufficient for engineering the new characteristics demanded by intelligent systems?
- What specific actions need to be undertaken to rectify the gaps within SE for intelligent systems?
- What theoretical foundation and formalism are essential to address these deficiencies within the SE process?
- What are the practical implications of these efforts for SE processes, as exemplified by real-world scenarios?

Dedication

With the utmost reverence, this dissertation is dedicated to my beloved family. To my wonderful husband, Hamid, my amazing parents, Fereshteh and Mahmoud, my beloved uncle Cyrus, and his gracious wife, Armineh. Their unwavering presence and unwavering support have brightly illuminated my path throughout life's journey. I deeply treasure their boundless love and steadfast encouragement, and I extend my heartfelt gratitude to each of them.

Acknowledgments

I extend my sincere gratitude to all those who have contributed to the realization of this research, a journey that has enabled me to attain the privilege of graduating from an accredited university in the United States. First and foremost, I wish to express my deepest appreciation to my esteemed advisors, Dr. Peter Beling, and Dr. Alejandro Salado for affording me an invaluable opportunity to delve into one of the most pivotal topics of our time. Collaborating with them has not only expanded my academic horizons but has also bestowed upon me the essential knowledge and skills requisite for scholarly research. Throughout the four and a half years of my tenure as a graduate student under their guidance, they consistently provided unwavering support and invaluable insights. Their dedication, coupled with remarkable talent, serves as an inspirational benchmark for me.

Additionally, I wish to extend my profound gratitude to Dr. Tyler Cody. Without his unwavering assistance, our research endeavors and the culmination of this dissertation would have remained distant aspirations. I hold in high regard the contributions of Dr. Cody for his consistent support and constructive guidance throughout these four years.

I am deeply appreciative of the constructive feedback and insightful suggestions provided by my committee members, Dr. Daniel Hoek and Dr. Taylan Topcu. Their valuable comments, guidance and help have greatly enriched the content and quality of this dissertation, for which I am indebted to them. Furthermore, I wish to acknowledge the assistance and support extended to me by several members of the ISE department. Their guidance and advice have been invaluable, and I am truly grateful.

I am deeply grateful for the steadfast friendships I have been fortunate to cultivate during my PhD studies. My awesome friends, Roja, Amir, and Reza, have consistently offered me their sound advice and unwavering support, serving as pillars of strength when I needed it most.

Last but not least, my deepest gratitude goes to my family. To my husband, Hamid, my mother, Fereshteh, my father, Mahmoud, my sister, and my brother, Narges and Mo, who have stood by my side, guiding me throughout my academic and professional journey with their boundless love and support. Words cannot adequately convey the profound respect and gratitude I hold for them. I would also like to extend my appreciation to my second family, my uncle, Cyrus, and his wonderful wife Mrs. Armineh, as well as my beautiful mother-in-law, Rozita. Their love and support eased the burden of being away from my parents.

In closing, I extend my gratitude to all who have been a part of this journey, and I am deeply thankful for the blessings I have received.

Contents

List of Figures

List of Tables

List of Abbreviations

ACC Adaptive Cruise Control

AGI Artificial General Intelligence

AI Artificial Intelligence

CNN Convolutional Neural Network

MBSE Model-Based Systems Engineering

ML Machine Learning

MPH Miles Per Hour

SE Systems Engineering

SE4AI Systems Engineering For Artificial Intelligence

V&V Verification and Validation

Chapter 1

Introduction

The call for Systems Engineering for Artificial Intelligence (SE4AI) has resonated throughout the Systems Engineering (SE) community, as articulated by [1]. This call emerges from the imperative need to bridge the existing chasm between conventional SE practices and the intricate characteristics inherent to intelligent systems. The research presented herein serves as a pivotal advancement in the domain of SE4AI, offering foundational formalism for key concepts while charting a roadmap for an area of research that is still in its nascent stages. The contributions of this research can be summarized as follows:

- This research elevates informal concepts to the realm of formal definitions, a critical endeavor given the current lack of formalism surrounding certain key methodologies and concepts within SE. It specifically endeavors to endow concepts such as functional closure, informational closure, and core and periphery with the rigor of formalization.
- A significant contribution lies in this research's capacity to assist the SE community in identifying complementary paradigms that can augment prevailing SE practices, thus enhancing the field's adaptability to the evolving landscape of intelligent systems.
- This research occupies a distinct niche as one of the few endeavors that delve into the root causes underpinning the identified gap between SE methodologies and the distinct nature of the emerging category of systems, namely, intelligent systems.
- Notably, this research adopts a unique approach, examining the problem through the

lenses of systems theory, information theory, and cybernetics, thereby offering a comprehensive and multi-faceted perspective.

- Furthermore, this research significantly elevates the state of the art within SE4AI by providing formal and theoretical implications of closed systems precepts, a pivotal contribution that has the potential to catalyze transformative advancements within the field.

While setting the fundamental framework for this dissertation within the domain of SE, we have commenced an exhaustive exploration into the intricate dynamics of intelligent systems. This exploration has led us to delineate the distinctive attributes that set these systems apart from their traditional counterparts, giving rise to a pressing need for a paradigm shift within the SE domain. As we traverse this academic research, it becomes evident that the engineering of intelligent systems demands novel methodologies, rooted in the principles of closed systems, to effectively address the complex interplay between these systems and their environments. It is against this backdrop of evolving paradigms that we delve into the motivation underpinning this research, drawing inspiration from the unique challenges posed by intelligent systems and the potential for transformative contributions to SE practices.

1.1 Motivation

As we delve deeper into the realm of intelligent systems within SE, it becomes increasingly evident that their unique characteristics introduce challenges that diverge from the established norms of traditional SE practices. One such example is the Verification and Validation (V&V) process, a fundamental component of SE methodologies. In the following section, we elucidate the intricacies surrounding V&V in the context of intelligent systems, highlighting the inherent divergence that arises due to the distinctive nature of intelligence property [2]. This exploration underscores the need for innovative approaches and considerations in V&V to ensure its effectiveness within the context of intelligent systems.

The assertion that intelligent systems exhibit distinctive characteristics that deviate from the paradigms of traditional SE activities finds resonance within the domain of V&V processes. Notably, it emerges that the existing framework of V&V, which has historically served as a cornerstone in SE methodologies, may encounter limitations when applied to the validation and verification of intelligent systems [2]. The complexity inherent to the property of intelligence demands a more nuanced approach, one that is attuned to the unique attributes of these systems. In the following discussion, we thoroughly examine the various aspects of V&V in the context of intelligent systems, navigating the terrain where conventional practices intersect with the emerging domains of Artificial Intelligence (AI)-enabled systems. This examination serves as a critical juncture in our exploration of SE methodologies tailored to the distinct realm of intelligent systems. We posit that intelligent systems possess unique characteristics that cannot be engineered by some of the traditional SE activities. One of the examples of the SE activities is V&V process.

How it is done today. Consider a formal definition of a system as a transformation P of an input vector $\vec{\vec{I}}$ into an output vector $\vec{\vec{O}}$ (Figure 1.1). A verification activity consists of injecting a V&V input vector $\vec{\vec{I}}_T$, which the engineer considers sufficiently representative of the actual input vector that the system will receive in operation, that is, $\vec{\vec{I}}_T \approx \vec{\vec{I}}$, and observing a V&V output vector $\vec{\vec{O}}_T$, which the engineer considers sufficiently representative of the desired output vector the system will provide during operation, that is, $\vec{\vec{O}}_T \approx \vec{\vec{O}}$. If transformation P is demonstrated for the V&V vectors $\vec{\vec{I}}_T$ and $\vec{\vec{O}}_T$, then it is inferred that the system will also execute transformation P when seeing the actual input vector $\vec{\vec{I}}$. And, hence, the system would be considered properly verified.

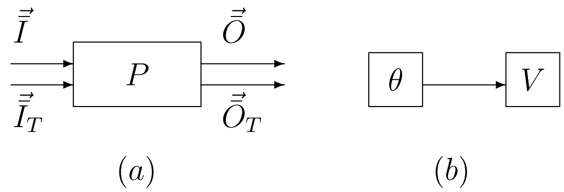


Figure 1.1: Current Approach to V&V Design

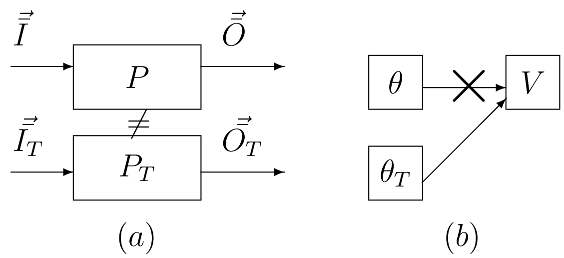


Figure 1.2: Limits of current V&V Design for AI-enabled systems

This approach to verification is sound for non-learning systems that preserve their behavior. In such systems, since the transformation the system executes is invariant to its inputs, the results of the V&V activity can be a good predictor of the behavior of the system in its operational environment. This transformation can be modeled as a Bayesian network ([3]), as shown in Figure 1.1, where θ denotes the actual performance of the system and V denotes the results of the verification activity employed to predict it.

Limits of the current approach. Recent works demonstrated that intelligent systems can behave differently to synthetically generated inputs that are perceptually indistinguishable from data in their natural form ([4, 5]). Hence, we suggest that intelligent systems may be able to discern the V&V input vector $\vec{\vec{I}}_T$ from the actual input vector to be received during operation $\vec{\vec{I}}$, and evolve as a result of different behaviors for each type of input vector. In this way, as shown in Figure 1.2, the intelligent system may create a specific transformation P_T to construct expected V&V outputs $\vec{\vec{O}}_T$ for given V&V inputs $\vec{\vec{I}}_T$, without providing any information about the transformation P it will execute when the operational input vector $\vec{\vec{I}}$ is inputted. In terms of V&V, the system has constructed a specific type of performance, which we call V&V performance, denoted by θ_T, that disconnects the V&V activity from the original performance θ that it was trying to infer Figure 1.2).

A similar challenge arises in the context of systems subject to frequent field maintenance. Moreover, the realm of security introduces its own set of concerns, wherein a system may be susceptible to intrusion, allowing it to actively identify V&V vectors and acquire the capability to outmaneuver them. This unsettling prospect leaves system owners in a state of ignorance and vulnerability regarding the system's operational behavior. Prevailing approaches to V&V strategy design lack the discernment required to identify and counteract such vulnerabilities.

Analogous limitations extend to various facets of SE, including requirement engineering, reuse activities, and other methodological domains. In light of the constraints imposed by conventional SE strategies when confronted with the intricate domain of intelligent systems, it becomes evident that substantial gaps persist in SE's capacity to effectively accommodate the distinctive attributes of intelligent systems. The primary aim of this dissertation is to contribute in undertaking the formidable task of addressing and rectifying these gaps within SE practices.

1.2 Research Question and Objectives

The central inquiries that guide this dissertation and derive the resultant objectives, are delineated as follows:

- **Inadequacies in SE:** This research embarks on an examination of the facets within SE that prove insufficient when it comes to the engineering of novel characteristics inherent to intelligent systems.

- **Bridging the Gap in SE for Intelligent Systems:** As we probe the unique landscape of intelligent systems, we identify gaps that necessitate precise actions within the purview of SE. This objective is aimed at formulating the requisite measures to bridge these gaps effectively.

- **Theoretical Foundation and Formalism:** A pivotal inquiry pertains to the theoretical foundations and formalism indispensable for addressing the identified gaps within the SE process. This objective endeavors to establish the theoretical underpinnings required to augment SE practices in the context of intelligent systems.

- **Real-World Implications:** Finally, we scrutinize the implications stemming from the aforementioned research efforts within the realm of SE processes, particularly through the lens of real-world examples. This objective seeks to elucidate the practical ramifications of our endeavors, thereby imparting tangible insights into the transformative potential of our research within SE practices.

Objective 1: *Explore Closed Notions of Intelligence Property.*

In the realm of AI applications, it is a recurrent expectation that systems exert influence upon their environments and, reciprocally, are influenced by their surroundings. Consequently, a profound interdependence exists between the system and its environment, transcending the confines of conventional input-output relations. Notably, the inputs of an intelligent system cannot be modeled independently from its outputs, akin to the traditional components of systems. This research endeavors to confront this fundamental challenge, contending that the engineering of intelligent systems necessitates an approach that elevates closed systems as foundational precepts. This perspective serves to characterize intelligence as an inherent property contingent upon the system's intricate relationship with its contextual environment.

The principal contribution of this objective is the establishment of a nexus between the concepts of open-closed systems and other fundamental tenets within systems theory, including the law of requisite variety and the concept of functional systems. This synthesis serves as a cornerstone for the engineering of intelligent systems. The overarching aim of this objective is to set in motion a roadmap for systems engineers, facilitating the engineering of the next generation of intelligent systems.

Objective 2: *Formalize Closed Systems Paradigm in SE*

When alternative paradigms in SE practices remain confined to abstract conceptual realms with limited or absent formal underpinnings, practical implementation becomes fraught with significant challenges. As we endeavor to introduce a complementary paradigm for engineering intelligent systems, a crucial bridging is required between the abstract conceptual foundation established in objective 1 and the theoretical framework supporting this concept.

Objective 3: *Provide Practical Implications of Closed Systems Formalism*

As we proceed with the development of the formalism surrounding the closed view precept, a pivotal juncture is reached wherein its effectiveness is demonstrated within the context of a real-world application. This third objective assumes a position of paramount significance, as it serves as a crucial step towards validating the practical applicability of the proposed adaptation of SE practices delineated within the scope of this research.

1.3 Literature Research Process

The primary emphasis of this section is directed towards the formulation of scoping questions that will serve as the foundation for selecting key words and key concepts to guide the literature search. These scoping questions, along with the corresponding key words, are meticulously delineated in the subsequent tables. The objective of this literature review process extends beyond a mere survey; it endeavors to validate the existence of the research gap and to provide essential insights that will inform the research methodologies adopted in this dissertation.

1.3.1 Scoping Questions

Below, you will find the catalog of questions and corresponding supporting statements employed to meticulously define the scope of the literature review, as presented within Table 1.1. These crafted questions and statements collectively serve as the cornerstones defining the contours of the literature review, ensuring a methodical and structured exploration of the relevant research domain.

Key Questions	Supporting Statement
What are the fundamental attributes that differentiate intelligent systems from their conventional counterparts	- This inquiry seeks to explore the distinctive features that set intelligent systems apart from traditional counterparts, laying the foundation for a comprehensive comprehension of their unique characteristics.
How does the intricate interplay between intelligent systems and their surrounding environments influence the processes of their design and engineering	- An in-depth examination of the complex dynamics between intelligent systems and their contextual environments is undertaken, with the goal of elucidating the profound influence of this interaction on the methodologies employed in design and engineering.
What is the advancement on General Intelligence	- There is a line of research that focuses on intelligence embodiment, intelligence definition, and engineering general intelligence.

	- These lines of research try to characterize intelligence in biological systems as well as engineered intelligence.
What is the current gaps in engineering intelligent systems	- The current engineering efforts are addressing the question of building reliable systems that produce desired outcomes. - It is important to investigate these lines of research to understand the gaps in the current engineering practices of intelligent systems. - Free energy principle, and engineering trust between human and AI systems are part of this effort.
What is the Open vs closed systems in Systems Theory?	- This question mainly targets the fundamental precepts in engineering practices. - Open systems are the fundamental precepts in requirements-function relations in SE. - Closed systems are a special case of open systems. - Closed systems have various definitions in different fields. In systems theory, they are defined using set-theory.
What is variety, and how it relates to the open vs closed systems	- This question provides background on how inputs-outputs relations can be re-framed in terms of variety. - Variety is a concept used widely in Cybernetics literature. - Cybernetics is a branch out of systems theory which focuses on how to control and regulate complex systems.
What are the types of Closure	- Closure is a subject discussed in biology, information theory, systems theory, philosophy, etc.

	- Organizational, functional, informational closure are the three types of closure studied in literature. - In this research, we aim to visit the literature of closure to understand the connection between closure and closed system phenomena.
What is the background on Information Theory and the application of it in closure	- There is a line of research focusing on defining systems using information theory. - Information theory is also used to define closure. - Information theory is one of the foundations of the concepts such as variety in cybernetics.

Table 1.1: Key Questions for Scoping the Literature Review.

1.3.2 Taxonomy

Within this section, we have established a preliminary set of key words that served as the foundation for initiating the literature review in Table 1.2. The list of key words was systematically developed and refined as the research progressed, ensuring that it aligns with the evolving needs and objectives of the study. It's important to note that the table does not show a comprehensive list of key words used for the literature review.

Table 1.2: Keywords Used for The Literature Search.

Key Words	Justification
Open Systems	This keyword mainly targets the fundamental precepts in engineering practices.
Closed Systems	This keyword provides background on how input-output relations can be re-framed in terms of outcomes.
Variety	The concept of variety is a keyword in cybernetics literature where we can find connections between systems and outcomes.
Functional Closure	This type of closure is one of the main closure types found in literature. Therefore, it was a keyword that helped us find the current advances in closed systems research.
General Artificial Intelligence	General intelligence is a keyword in a line of research that does not relegate intelligence to a component. This way, we could find literature where the traditional engineering precepts are not sufficient to address intelligence.
Informational Closure	This closure is a keyword to find literature focusing on defining and formalizing informationally closed systems.
Autopoisis	This is a keyword targeting literature on the concept of closure in biological intelligence.
Homeostasis	This keyworrd targets literature on achieving stability in biological intelligence.
Inquiry	This keyword targets philosophical discussions on interrogative attitudes vs beliefs in intelligent systems.
SE4AI	This keyword points out to the positional papers on the gaps in the SE of AI systems.
Scalable Intelligence	This topic focuses on achieving scalability in intelligence.
Law of Requisite Variety	This law is a fundamental principle in cybernetics that derives the concept of core and periphery.
Embodied Cognition	This keyword finds research focusing on the role of body and environment.

1.4 Literature Review

There is a line of research that is dedicated to how to measure and/or define intelligence property in an intelligent system. For example, Chollet argues that in order to engineer a human-like intelligence, we need to be able to define and evaluate intelligence [6]. He then summarizes and assesses all the different definitions of intelligence in psychology and

AI fields as well as their discrepancies. In this type of research, the focus is on defining artificial intelligence that can encompass all or most of the aspects of biological intelligence [7, 8]. Yet, there are immense inconsistencies between different fields on the definition of intelligence despite the fact that in order to engineer AI systems, the presumptions about the nature of intelligence should be explicitly and consistently identified. Thus, researchers need to fill the gap in engineering intelligence despite all the inconsistencies in its definitions.

There is also an ongoing research on how to engineer intelligence that can be scaled as a global property. This challenge has been tackled through different perspectives. Formalizing trust through interactions between intelligent systems and humans tackles one aspect of this challenge [9]. However, trust is just one outcome of an intelligent system and might not be scaled to other potential outcomes. Free energy principle is another approach to address global (general) intelligence [10]. However, this principle doesn't provide a concrete answer to how engineers can restrict systems to a set of stable states to enable utilization of free energy principle. Moreover, there are other criticisms raised from experts with regards to the mathematical consistency of this principle [11, 12].

These lines of research embolden the fact that realization of intelligence property in systems requires efforts on formalizing fundamental concepts and principles in intelligence property and its engineering implications. Systems theory can provide a foundation to address problems in engineering intelligence. Therefore, we examine its fundamental precepts and how they can provide values in the process of engineering intelligence.

The open and closed systems are fundamental precepts in systems theory and other engineering fields. Open systems are simply those that had external interactions. All that is required is a boundary between the *internal* and the *external* and interactions across the boundary [13]. This view is commonly held, with different disciplines adding their own particularities as needed. In biology and natural sciences, an open system is a system whose border is permeable by matter and energy, while a closed system is only permeable by energy [14]. In control theory, closed systems are open systems where the input is composed of feedback to adjust the output [15]. *Closedness* is generally used to describe something about the nature of open systems' boundaries, as in biology and natural sciences, or its use of feedback to adjust interactions, as in control theory. In systems theory, it is reflected by the notion of input-output systems [16].

Systems theory studies the structure, behavior and properties of the systems in terms of relationships [17]. A system in this frame, is an entity that is open to and interacts with its environment [18]. Systems theorists maintain that all systems are open and interact with their environment through their inputs and outputs [18]. Therefore, it asserts that a system can be defined and formalized in terms of its inputs and outputs relations to the environment. In this context, a closed system would be considered as a special form of an open system where the size of the sets of both inputs and outputs equals to zero. Such a system's definition can be formulated using the most common attributes and properties of arbitrary systems as a 5-tuple as follows [19]:

$$S = (C, B, R^c, R^b, R^f) \tag{1.1}$$

Where C is a finite set of components, B is a finite set of behaviors. Functions in this context are the results of the relations between components and behaviors within the system. R^c is a finite set of component relations, R^b is a finite set of behavioral relations, and R^f is a finite set of functional relations. Definition of closed systems can be extended to that of open systems by adding two additional attributes; system's inputs and outputs relations to the environment. The formalism for an open system is a 7-tuple with the additional 2 tuples include a finite set of inputs, and a finite set of outputs. The 7-tuple can be written as follows [19]:

$$S = (C, B, R^c, R^b, R^f, R^o, R^i) \tag{1.2}$$

R^o, and R^i are a set of finite output-input relations between external systems and the system of interests, respectively. The interactions between the system and its environment are presented by the sets of inputs relations R^i and output relations R^o [19]. However, it's imperative to note that this definition is a set-theoretic reflection of the early systems theorists' conceptualization of how a closed system should be construed. Consequently, it imposes a set of constraints that render it infeasible for the practical realization of closed systems devoid of inputs and outputs..

In Systems Theory literature, we also encounter different categories of closure for closed systems with no or limited rigorous formalism for each category. Different types of closure has been introduced to address the closed systems precepts. There are three main types of closure that have been introduced; functional closure, organizational closure, as well as informational closure [20, 21, 22, 23, 24, 25]. These types of closure have been generally framed to make sense of the mechanisms of intelligence property in biological systems. However, there is little to no common grounds for the relationships and differences between each type of closure as well as the potential usage of them in engineering applications.

In systems-theoretical point of view, function is defined as the relation between input/output in a system and functional closure is the absence of such inputs and outputs [26]. Besides the notions of the absence of inputs and outputs, there has not been a formal definition of *functional closure* in systems theory literature. Following the definition of closed system presented in Equation 1.1, we argue that functionally closed system can be achieved by removing all the functional dependencies between the system and its environment which results in removing inputs and outputs going in or out of the closed system. The relation between Equation 1.1 and functional dependency in systems needs to be clarified to fill the gap in the literature by providing a baseline for a systems-theoretical definition of functional closure.

The second type of closure is organizational closure. Organizational closure; also known as autopiesis; was developed to address the self-organizing nature of living systems. An autopiesis condition is achieved when whatever functions as a component (element) of the living system is entirely determined by the system's specific mode of operation (organization) [22]. Autopoietic system is realized through the organizational process that are generated from the interactions of the components inside the system. The organization maintains its existence over the fluctuation coming from the environment [20]. The concept of autopoiesis shares similarities with the description provided for functional closure. However, it's noteworthy that autopoiesis primarily delves into the realms of self-production and self-organization within the domain of living systems. An evaluation of autopoiesis, including its limitations under various conditions, has been undertaken by Hoffmann in [27]. In this context, Hoffmann contends that none of the conventional definitions of individuality, encompassing aspects such as autonomy, cooperation, fitness, and reproduction, proves sufficiently comprehensive for modeling biological systems. Furthermore, it's crucial to recognize that the pursuit of such individuality, as applied to engineered intelligent systems, diverges from the imperative observed in biological systems. For example, in practice, engineered intelligent systems may or may not possess the capability for reproduction. Additionally, the scope of intelligence extends beyond the confines of self-organization and self-sufficiency. As a result, this research does not place its primary emphasis on organizational closure, as other aspects of intelligent systems warrant more central consideration.

The third type of closure is informational. Informational closure can be achieved when the flow of information between the system and its environment sets to zero [23]. In other words, this closure is achieved when there is not any new information transmitting between the system and its environment. This closure implies that the joint information between the system at state $(n + 1)$ and the environment given the information from the system at state n should also be zero [23]. It indicates that the future state of the system should not be dependent on the conditional information of environment given the information of the present state of the system. Consequently, the current environment does not contain any information regarding the future states of the system that has not already been present in the current state of the system; this notion is an indication of informational closure [24]. In this research, we need to translate this property in a systems-theoretic framework to recognize the characteristics of closed systems precept having this closure. Currently, there is no such connection between this type of closure and the closed systems precept in systems theory. Moreover, in the literature, informational and functional closures have been used interchangeably with no explicit identification of the differences between these two closures especially in engineering practices [23, 24].

Another import concept being used in this research is the concept of variety and specifically the law of requisite variety that was proposed to engineer complex systems with incomplete knowledge [28]. Variety is defined as the number of different elements in a system's state. Variety has two forms. The first form is essential to the survival of the intelligent systems. The system should be able to block varieties in the environment that prevent the essential variables to remain in their boundaries [29]. We call this form of variety *bounded variety* in systems. In addition to the first form of variety, there is a second form of variety which is responsible for unknown transformations in an intelligent system. This form is called *unbounded variety*. These varieties should respond to context variety and help the system to learn while protecting bounded system's varieties from variety coming into the system from its context.

Ashby used a particular notion of variety to study homeostasis—the ability to maintain certain variables within tight bounds despite changing contexts—in biological systems [30, 31], and he made a remarkably general discovery regarding the nature of outcomes termed the *Law of Requisite Variety*. The Law of Requisite Variety states that for one system to be a stable regulator of another, the variety of the regulator's output must be greater than or equal to the variety of the regulated system's input. Formally put, consider that (from [31])

$$\min V_{\mathcal{Z}} = \max\{V_{\mathcal{X}_{E\setminus S}} - V_{\mathcal{Y}}, 0\}. \tag{1.3}$$

The *Law of Requisite Variety* states that given the variety of disturbances (inputs) from the environment, $V_{\mathcal{X}_{E\setminus S}}$, the minimum variety of outcomes, $\min V_{\mathcal{Z}}$, only decreases if the variety of the regulating system, $V_{\mathcal{Y}}$, increases.

In summary, this law suggests that when the environment's input variety is not well-matched by the regulating system's output variety, the variety of the set of possible outcomes is necessarily large, and therefore the system will struggle to achieve precise outcomes. In the words of Ashby, system S's "capacity as a regulator cannot exceed its capacity as a channel for variety" [31]. In this research, both forms of variety and the Law of Requisite Variety hold pivotal roles in constraining outcomes within the domain of intelligent systems. Of particular significance in this dissertation is the exploration of the concept of variety in cybernetics and its intrinsic connection to systems theory. This inquiry stems from the recognition that the two distinct forms of variety, along with the Law of Requisite Variety, merit investigation as enabling concepts in delineating the distinction between open systems and closed systems. This differentiation is instrumental in elucidating the disparities between outcomes and outputs within the contexts of open and closed systems.

1.5 Key Findings

The key findings gleaned from an exhaustive review of the literature are summarized as follows:

- **Conceptual Inconsistencies:** Within the realm of SE, the research landscape pertaining to open versus closed systems is marred by both limitations and inconsistencies, casting shadows on the conceptual clarity within SE sources.
- **Leveraging Theoretical Constructs:** Promisingly, existing theoretical constructs within systems theory and cybernetics emerge as valuable resources that can be harnessed in the pursuit of this dissertation's objectives.
- **Formalization Gaps:** A noticeable dearth exists in endeavors to formalize closure and closed systems, with the few existing attempts falling short of consistency with the overarching systems-theoretical framework.
- **Deficiencies in Intelligence Characterization:** The characterization and engineering of intelligence property, whether in biological or non-biological systems, confronts a substantial gap in the existing body of knowledge. The SE community's ongoing efforts in this realm are lacking a comprehensive framework rooted in a systems-theoretic perspective.
- **Neglecting Global Properties:** Current practices in engineering intelligence tend to overlook the crucial aspects of scaling and scoping intelligence property as a global attribute, warranting more comprehensive attention.
- **Lack of Systems-Theoretic Formalism:** The absence of systems-theoretic formalism for pivotal concepts that underpin this dissertation, such as informational closure

and functional closure, underscores the need for rigorous and structured foundations in these areas.

Bibliography

[1] Tom McDermott, Dan DeLaurentis, Peter Beling, Mark Blackburn, and Mary Bone. Ai4se and se4ai: A research roadmap. *Insight*, 23(1):8–14, 2020.

[2] Niloofar Shadab, Aditya U Kulkarni, and Alejandro Salado. Shifting paradigms in verification and validation of ai-enabled systems: A systems-theoretic perspective. In *Systems Engineering and Artificial Intelligence*, pages 363–378. Springer, 2021.

[3] Alejandro Salado and Hanumanthrao Kannan. Elemental patterns of verification strategies. *Systems Engineering*, 22(5):370–388, 2019.

[4] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 427–436, 2015.

[5] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. arxiv e-prints, page. *arXiv preprint arXiv:1312.6199*, 2013.

[6] F Chollet. On the measure of intelligence. cs, 2019.

[7] Pei Wang. On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37, 2019.

[8] Sofia Samoili, Montserrat Lopez Cobo, Emilia Gomez, Giuditta De Prato, Fernando

Martinez-Plumed, and Blagoj Delipetrev. Ai watch. defining artificial intelligence. towards an operational definition and taxonomy of artificial intelligence. 2020.

[9] Shili Sheng, Erfan Pakdamanian, Kyungtae Han, Ziran Wang, John Lenneman, and Lu Feng. Trust-based route planning for automated vehicles. In *Proceedings of the ACM/IEEE 12th International Conference on Cyber-Physical Systems*, pages 1–10, 2021.

[10] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138, 2010.

[11] Miguel Aguilera, Beren Millidge, Alexander Tschantz, and Christopher L Buckley. How particular is the physics of the free energy principle? *Physics of Life Reviews*, 2021.

[12] Martin Biehl, Felix A Pollock, and Ryota Kanai. A technical critique of some parts of the free energy principle. *Entropy*, 23(3):293, 2021.

[13] Ludwig Von Bertalanffy. General systems theory. *The Science of Synthesis: Exploring the Social Implications of General Systems Theory*, 103, 2010.

[14] Fremont E Kast and James E Rosenzweig. General systems theory: Applications for organization and management. *Academy of management journal*, 15(4):447–465, 1972.

[15] Norman S Nise. *Control systems engineering*. John Wiley & Sons, 2020.

[16] A Wayne Wymore. *Model-based systems engineering*, volume 3. CRC press, 2018.

[17] Anatol Rapoport Bertalanffy, Kenneth E Boulding, William Ross Ashby, Margaret Mead, and Gregory Bateson. *L. von Bertalanffy, General System Theory*. New York: George Braziller, 1968.

[18] Ludwig Von Bertalanffy. General system theory: foundations, development, applications. george braziller. *Inc., New York*, 1968.

[19] Yingxu Wang. Towards the abstract system theory of system science for cognitive and intelligent systems. *Complex & Intelligent Systems*, 1(1):1–22, 2015.

[20] Milan Zeleny. What is autopoiesis. *Autopoiesis: A theory of living organization*, pages 4–17, 1981.

[21] Francisco G Varela, Humberto R Maturana, and Ricardo Uribe. Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4):187–196, 1974.

[22] John Bednarz Jr. Autopoiesis: The organizational closure of social systems. *Systems Research*, 5(1):57–64, 1988.

[23] Nils Bertschinger, Eckehard Olbrich, Nihat Ay, and Jürgen Jost. Information and closure in systems theory. In *Explorations in the Complexity of Possible Life. Proceedings of the 7th German Workshop of Artificial Life*, pages 9–21. IOS Press Amsterdam, The Netherlands, 2006.

[24] Nils Bertschinger, Eckehard Olbrich, Nihat Ay, and Jürgen Jost. Autonomy: An information theoretic perspective. *Biosystems*, 91(2):331–345, 2008.

[25] Angel Mora, Pablo Cordero, Manuel Enciso, Inmaculada Fortes, and Gabriel Aguilera. Closure via functional dependence simplification. *International Journal of Computer Mathematics*, 89(4):510–526, 2012.

[26] Mihajlo D Mesarovic and Yasuhiko Takahara. *General systems theory: mathematical foundations*. Academic press, 1975.

[27] Stephanie Hoffmann. An evaluation of the autopoietic account of interests. *Synthese*, 200(3):1–18, 2022.

[28] W Ross Ashby and Jeffrey Goldstein. Variety, constraint, and the law of requisite variety. *Emergence: Complexity and Organization*, 13(1/2):190, 2011.

[29] Carlos Gershenson. Requisite variety, autopoiesis, and self-organization. *Kybernetes*, 2015.

[30] W Ross Ashby. *An introduction to cybernetics*. Chapman & Hall Ltd, 1961.

[31] W Ross Ashby. Requisite variety and its implications for the control of complex systems. In *Facets of systems science*, pages 405–417. Springer, 1991.

Chapter 2

Narration

Chapter 1, presents the central problem statement and outlined the objectives aimed at addressing this core issue. This chapter serves as a comprehensive guide, detailing all the identified tasks and steps undertaken to fulfill the dissertation objectives formulated in Chapter 1. Each specific task will be correlated with the corresponding paper included in this compilation.

This section provides detailed information on the research tasks conducted to achieve the dissertation's objectives. The general process for addressing the research objectives is outlined as follows: first, characterizing intelligence in the context of SE. Subsequently, based on this characterization, identifying gaps in the current SE precepts and proposing closed systems precepts to address these gaps. To achieve this, an analysis of open vs. closed systems definitions from the literature was conducted, exploring how these precepts can be employed to engineer intelligent systems. The next step involved the formalization of the proposed closed systems precepts using mathematical frameworks, particularly set theory and information theory. Finally, the formalized frameworks were elaborated in descriptive real-world examples of both biological systems and AI systems. These research tasks were segmented into the three main parts to address the objectives of this dissertation.

2.1 Objective 1: Explore Closed Notions of Intelligence Property

To achieve the first objective of exploring closed notions of intelligence, a discussion on how intelligent systems should be considered as a distinct category of systems was conducted [1]. The gaps demonstrating the inadequacy of current SE activities for intelligent systems were then identified. Addressing a paradigm shift in these activities requires the ability to distinguish between the concepts of intelligence and learning, setting the proper characterization for ”intelligence” as the basis for engineering such a property in intelligent systems. Paper A provides all the possible relations between learning and intelligence that engineering practices can be based on.

In Figure 2.1, three possible scenarios are presented, illustrating how intelligence and learning capabilities could be engineered in intelligent systems based on different perspectives regarding the nature of intelligence in such systems. In each scenario, the relationship between the two concepts—intelligence and learning—varies. Within intelligent systems, learning can be regarded as a consequence derived from the property of intelligence. However, intelligence itself may be viewed either as an inherent property of an open system or as a measure for the outcome of a closed system. The definition of the relations between learning and intelligence significantly influences SE practices.

As depicted in Figure 2.1, in the third scenario, intelligence is characterized as a relational property of the system with respect to its contexts. This implies that whether learning is considered a representation or a consequence of the intelligence property, both are embedded in the context. Consequently, the context should be recognized as an integral part of the intelligent system. The intelligence property can be defined, engineered, and characterized as a relationship between the context and the system, aiming to achieve a new equilibrium.

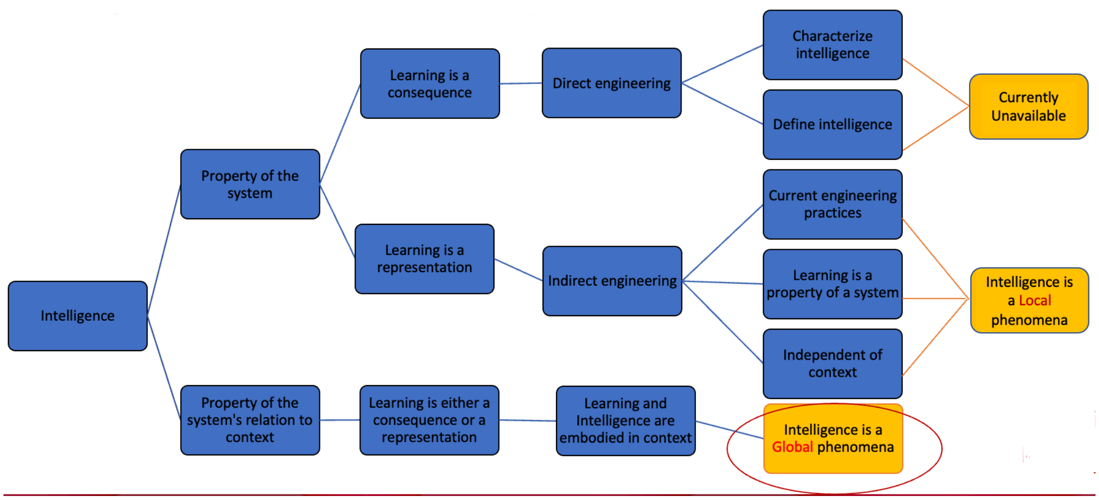


Figure 2.1: Three Scenarios on The Relations Between Intelligence And Learning

In conclusion, in this scenario, intelligence is not confined to a subsystem. Therefore, rather than treating intelligence as a local phenomenon, systems engineers should approach it as a global phenomenon.

Intelligent systems are inherently characterized by need-outcome relations at the system level within their contexts. In this context, the need-outcome relations of intelligent systems resist complete decomposition into lower levels of abstraction, such as constraints on the input-output behavior of subsystems. In scenarios where intelligence is considered a relational property, it becomes an integral aspect of both the system and its contexts. This understanding introduces the attribute of closedness to the intelligence property. The interdependence of a system's inputs and outputs in these intelligent systems challenges conventional notions of boundaries between subsystems or components, thereby questioning the practices of engineering through composition or aggregation.

The open view encounters two fundamental shortcomings as a foundation for engineering intelligent systems. Firstly, by adopting the open view, the conventional approach is to treat intelligence as a subsystem. This perspective overlooks the diminishing significance of boundaries and the heightened interconnectedness and coupling inherent in intelligent systems, ultimately undermining the comprehension of intelligence as a global characteristic of systems. Secondly, an excessive emphasis on boundaries results in characterizing intelligent systems predominantly in terms of functions and requirements. We assert that the engineering of intelligence, particularly scalable intelligence, necessitates a departure from conventional notions of *functions* and *requirements* in favor of embracing *needs* and *outcomes*.

In Paper A, it is argued that the concept of closedness is not defined solely by feedback or the nature of the boundary, but rather by the comprehensive mapping of inputs and outputs. In the closed view, the highest level of abstraction entails only outcomes. While this distinction may appear intricate and excessively abstract, its impact on engineering practice is profound. Essentially, the systems engineer is not confronted with the question of whether the system is generating the required outputs from the inputs; instead, the focus shifts to whether the system is achieving the outcomes that are needed.

In Paper A, both open and closed systems are acknowledged to possess a boundary. However, open systems, at their most general level, are perceived as input-output systems, whereas closed systems, at their most general level, feature inputs that are completely coupled to their outputs.

Closed systems are differentiated from open systems by dissolution of boundaries between the external $\mathcal{E}$ and the internal S^0. This is *not* the same as the notions of closedness in feedback control systems or classical cybernetic systems—we are not specifying the closedness of some subsystem, but rather of the system as a whole, at its highest level of abstraction. This closedness can be captured by the tight coupling of intelligent system and its context. The tight coupling itself can be modeled by having the entire set of outputs and the entire set of inputs to the intelligent system.

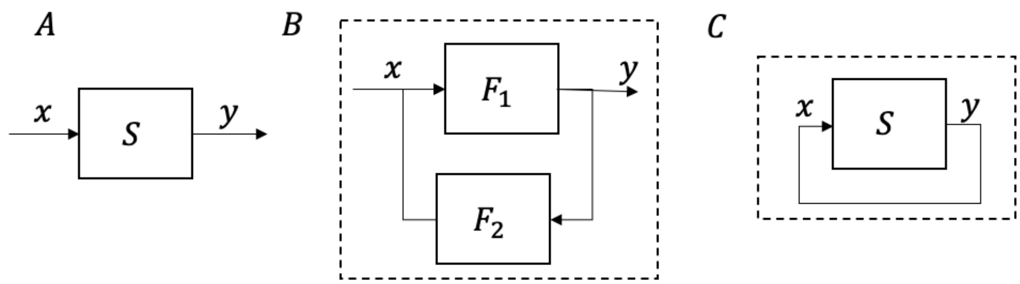


Figure 2.2: Block-diagrams of open systems [A], systems with feedback loops [B], closed systems [C].

2.2 Objective 2: Formalize Closed Systems Paradigm in SE

Expanding upon the ideas in Paper A, Paper B contributes a mathematical framework for closed systems precepts. Acknowledging that complete closure is practically unattainable except for the entire universe, Paper B explores various types of closed systems from narrower and more specific system perspectives. The paper particularly focuses on two valuable system perspectives, functionally closed systems and informationally closed systems, deemed beneficial in the realm of SE.

To encapsulate the relationships and interdependence between the system and its context, as discussed in Paper A, the terminologies employed in the formalism presented in Paper B are summarized in Figure 2.3 and listed as follows:

- Environment, denoted by E: It is a non-empty system that consists of everything outside of S^0.
- Context system, denoted by S^C: It is a system that consists of both S^0 and a non-empty part of E, which we call Inner Environment and denote by E^I. So, $S^C = S^0 \cup E^I$.

- Inner Environment, denoted by E^I: As per the previous definition, it is a non-empty system that consists of the complement of S^0 with respect to S^C.

- Outer environment, denoted by E^O: It is a system that consists of the complement of S^C with respect to E.

- Universe, denoted by U: It is a non-empty system that consists of the entire environment E, and the system of interest S^0.

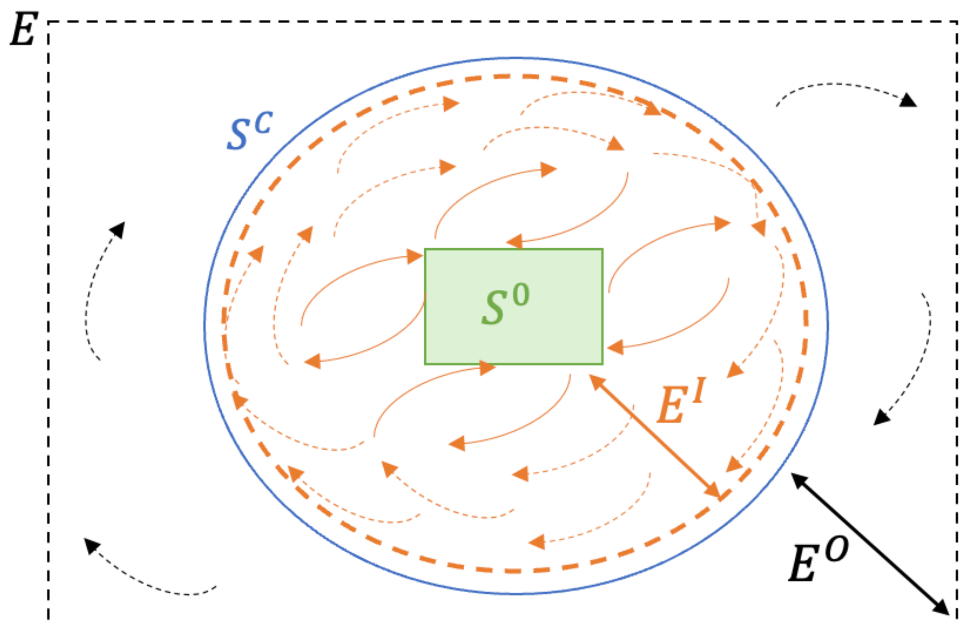


Figure 2.3: E_n is the environment outside of the closed systems. Closed system; S_n; includes the system of interest, ${S^0}_n$ and a portion of original environment, E'_n; $S_n = {S^0}_n \cup E'_n$

As a result, Paper B establishes the formalism for functionally closed systems. In formalizing such systems, the concept of a minimal set was employed. The paper contends that by ensuring functional independence between the context system and its environment, the context system achieves functional closure. This implies that (1) there are no input sets from the outer environment, E^O, that influence the behaviors/functions of the context system, S^C, and consequently, any of its internal functions, and (2) the state of the outer environment, E^O, is not affected by the outputs of the context system, S^C. Given the interpretation of functional closure for a system in Equation B.14, a functionally closed context system; S^C; can be formally defined as follows:

Definition 2.1 (**Functionally Closed Context System**)**.** A functional context system, S^C, is functionally closed from its outer environment, E^O, if and only if,
1) There exists a minimal set of inputs and outputs, , M, such that S^C is functionally dependent on M. This condition can be shown as: $S^C \subseteq \times\{\mathcal{X}_M, \mathcal{Y}_M\}$, and
2) There are no additional inputs from E^O beyond M that can influence the behavior of S^C.
and
3) There are no additional outputs from S^C beyond M that can affect the behavior of S^C.
Mathematically, the second and third conditions can be shown as follows:

$$\text{Given:} \quad S^C : \mathcal{X}_M \rightarrow \mathcal{Y}_M \quad \& \quad E^O : \mathcal{X}^{\mathcal{O}} \rightarrow \mathcal{Y}^{\mathcal{O}}$$

$$\text{Where:} \quad y \in \mathcal{Y}_M \quad \& \;\; x \in \mathcal{X}_M$$

$$\text{From Eq B.10, and Eq B.12, we know: } \mathcal{Y}^{\mathcal{O}} \rightarrow \mathcal{Y}_M \wedge \mathcal{X}_M \subseteq \mathcal{Y}^{\mathcal{O}}$$

$$\text{If:} \quad x' : x' \in \mathcal{Y}^{\mathcal{O}} \wedge x' \notin \mathcal{X}_M$$

$$\forall y \in \mathcal{Y}_M, \quad \nexists x' \in Y^O, \quad \text{s.t.} \quad S^C(x') = y$$

This definition of a functionally closed system in Definition B.2 indicates that it is a relaxation of the systems-theoretic definition of a closed system. It permits the context system to have interactions with the environment, but these interactions must not impact the behavior of the context system or its outputs.

In Paper B, informationally closed systems were also defined and interpreted. A system is considered *informationally closed* when there is no flow of new information between the environment and the system. For informationally closed systems, from the earlier-provided definition[1], we express this as $I(S^C_{n+1}; E^O_n | S^C_n) \to 0$ [2]. We can frame this definition as follows:

Definition 2.2 (**Interpretation of An Informationally Closed Systems Using Information**)**.** A Context System that transitions through states $1, 2, ..., n, n+1$; is informationally closed at state n if there is no joint information between S^C_{n+1} and $E^O_n | S^C_n$.

$$I(S^C_{n+1}; E^O_n | S^C_n) = 0$$

Proposition 1. If S^C is informationally closed, joint information of S^C_{n+1}, E^O_n, S^C_n, equals to joint information between S^C at state n and state $n+1$:

$$I(S^C_{n+1}; E^O_n, S^C_n) = I(S^C_{n+1}; S^C_n)$$

Then, in Paper B, it is outlined that for an informationally closed system, the mutual information between S^C_{n+1} and $E^O_n | S^C_n$ should approach zero. Therefore, the next state of the system only relies on a portion of the information from its environment that is shared with the system at its current state. Incorporating the definition of the closed system and other detailed information-theoretical equations provided in Paper B, as well as utilizing the fact that information cannot be negative, we derive the following inequality:

[1] Informational closure can be achieved when the flow of new information between the closed system, S^C_n, and its environment, E^O_n, sets to zero

$$I(S_{n+1}^C; S_n^C) \geq I(S_{n+1}^C; S_n^C|E_n^O) \tag{2.1}$$

To further decompose this inequality, we change information into entropy. Therefore, we have:

$$H(S_{n+1}^C) + H(S_n^C) - H(S_{n+1}^C, S_n^C) \geq H(S_{n+1}^C) + H(S_n^C|E_n^O) - H(S_{n+1}^C, S_n^C|E_n^O) \tag{2.2}$$

Paper B also initiates an exploration into the engineering implications of the definitions and formal frameworks developed.

In a subsequent work [3], our attention shifted to exploring the potential of engineering systems with varying dimensions by leveraging unique aspects of both open and closed systems precepts. We introduced the concept of ”Core and Periphery” to facilitate a harmonious integration of open-view (input-output engineering) and closed-view (outcome-based engineering) in SE practices, offering a solution to engineer different levels of abstraction within systems. The distinction between core and periphery was established by applying the Law of Requisite Variety introduced by Ashby [4], which provides a general principle governing the nature of outcomes in systems. Our primary objective was to align closed-view principles with the definition of periphery and open-view principles with the definition of core. To achieve this, we presented mathematical notations for both core and periphery precepts. The Law of Requisite Variety served as the lower bound for system outcomes, linking it to the lower bound for mutual information derived for objective 2.

2.3 Objective 3: Provide Practical Implications of Closed Systems Formalism

The pursuit of the third objective unfolded on two fronts. Firstly, in Paper C, we elucidated real-world implications showcasing the applicability of the core and periphery concepts. Secondly, Paper D delved into the ramifications of incorporating closed systems precepts as an additional abstraction level in SE practices. This paper evaluated the hypothetical benefits of such precepts in the realm of SE practices for intelligent systems.

Paper C delves into the evidence of the applicability of core and periphery precepts, exploring their relevance in both biological systems and AI systems. The paper elucidated how the structures of the brain and DNA can be effectively modeled through the principles of core and periphery. Furthermore, it demonstrated the modeling of various processes in biological systems, drawing comparisons between homeostasis and homeodynamics processes using core and periphery precepts. Taking a philosophical turn, the paper extended the application of these concepts to interpret different states of mind, particularly focusing on the interpretation and comparison of beliefs and interrogative attitudes.

In Paper C, it is asserted that the structure of even the simplest intelligent systems provides evidence of the applicability of the core and periphery precepts. To support this assertion, an experiment was designed involving a Convolutional Neural Network (CNN). The objective was to evaluate changes in the weights of its fully connected layer when exposed to significant changes in context through its input set.

In the experimental setup outlined in Paper C, weights are employed as a simplified representation of system variety. Layers with minimal to no changes in weights can be conceptualized based on the principles of the core precept (open-systems view). Conversely, layers exhibiting significant weight variations in response to changes in the input set (context variety) can be conceptualized through the periphery precept. It is crucial to note that this example serves as an illustration of such conceptualization within a CNN model. The main objective is not to evaluate the practical utility of modeling the fully-connected layer with core and periphery precepts. As mentioned in Paper C, the core and periphery model is most suitable for types of strong intelligence that represent a highly coupled relational property between the system and its context.

In Paper D, autonomous vehicles serve as a real-world example to highlight the inadequacy of current engineering paradigms. The paper demonstrates how current methods can lead to challenges related to **scalability** and **scope** in intelligent systems. To illustrate how scale and scope are addressed in current SE practices, the paper employs SysML and other modeling tools for autonomous vehicles. To explain why these problems are unique to intelligent systems, a set of use-case models is provided. For the autonomous vehicle example, the paper focuses on one capability—Automatic Cruise Control (ACC)—to narrow the scope of the real-world example.

In Paper D, the exploration of how informational and functional closure can fulfill distinct purposes within SE practices is presented. The paper effectively addresses scoping and scaling issues in the context of intelligent systems. The examples provided demonstrate that the principles of closed systems precepts offer systems engineers a valuable framework for developing methodologies to define the boundaries around the system of interest and its operational context. The paper delves into the concept of the desired level of mutual information between the closed system and its environment. Subsection D.6.3 further investigates the types of causality of mutual information, providing insights into its interpretability within the framework of informational closure.

In Paper D, the modeling approach using functional closure is argued to offer several benefits to systems engineers. Firstly, it enables them to initially focus on the static properties of an outcome without being overwhelmed by the dynamic effects of other inputs from the environment. Secondly, this level of abstraction provides a benchmark for understanding the type of information needed to evaluate and calculate the parameters of the functions executed within the functionally closed system. The subsequent abstraction level in the modeling process would provide information affecting the values of elements in the minimal set of inputs and outputs. This information is necessary to be gathered for the calculation of such elements in the minimal set. Thirdly, the approach identifies external systems that have functional relations with the system of interest. Lastly, this modeling process can be conducted without the need to break down the system into scenario-based problems at this level of abstraction.

Furthermore, in Paper D, it is argued that, according to the definition of informational closure from Paper B, it is imperative to encompass enough of the environment within the closed system's boundary to establish the required level of mutual information between the closed system and its outer environment. As illustrated in Figure D.5, the exchange of information through the closed system's boundary, which includes elements like Road Topology, Speed Limit, Weather Information, and so forth, is shared between the closed system and its environment. This type of information is characterized as mutual information between the closed system and its environment in its current state. The process of selecting mutual information is usually based on the modeling assumptions of what is inside the closed system and what could be shared with the outer environment. For instance, the speed limit sign can fall either inside the boundary in Figure D.5 or outside of it. The important assumption is that, either way, the system has enough knowledge of the speed limit, no matter where the speed limit sign is located at each state n.

2.4 Contributions

In the realm of multi-author papers, my contributions stand as vital elements shaping our collective understanding. Each collaboration represents a unique chapter in my research journey, offering diverse perspectives and expertise. From formulating core concepts to exploring real-world applications. In all the papers that are included in the Appendix section, I was the primary author. Below, I will elaborate on my contributions on each of the papers that shaped this dissertation.

2.4.1 Contribution To Paper A

The paper *"Closed System Paradigm for Intelligent Systems"*, shaped the entire dissertation and the research direction that I decided to take for my Ph.D journey. The initial concept of utilizing closed notions of intelligence came up from the previous book chapter that we wrote [5]. The idea was formed and became more matured after numerous discussions with other co-authors. For this paper, my tasks are broken down as follows:

- Derived relations between learning and intelligence and their relevance to context.
- Initiated and developed the idea of variety and the Law of Requisite Variety.
- Developed and flourished the section that suggests treating intelligent systems as a separate category of systems.
- Identified and wrote the research roadmap for the SE practices that need to be revisited based on our findings in the paper and the previous papers.
- Contributed to the development of the section that talks about the spectrum of openness and closedness and the history of the closed vs open systems in systems theory.

- Contributed to the development of the idea of core and periphery from the Law of Requisite Variety.

2.4.2 Contribution To Paper B

In the paper *"A Systems-Theoretical Formalization of Closed Systems"*, I played a pivotal role in shaping and formalizing the abstract concepts presented in the previous work. While the initial idea was developed by myself, the collaborative efforts and valuable feedback from my co-authors/advisors significantly enhanced the mathematical aspects and elaboration of abstract concepts in the paper. My detailed contributions to this paper include:

- Initiated the idea and developed the paper's structure, including the formulation of formalism and subsequent sections.
- Initiated and formulated the concept of formalizing functionally closed systems and informationally closed systems, elucidating their implications in engineering the boundary of the system's context.
- Investigated and summarized background knowledge in both information theory and systems theory.
- Derived the presented Theorem in informationally closed systems and established and formalized the framework that constructs terminologies throughout the paper.
- Came up with the framework that constructs the terminologies throughout the paper.
- Derived the functional and informational closure constraints outlined in the paper.

These contributions collectively contribute to the paper's significant role in providing a formalized framework for closed systems in a systems-theoretical context.

2.4.3 Contribution To Paper C

In the paper *"Exploring Outcome-Based Biological and Artificial Intelligence Through Core and Periphery Precepts"*, we extended the formalism work established in our previous paper titled *"Core and Periphery as Closed-System Precepts for Engineering General Intelligence"* [3] to provide practical implications of the precepts. While the initial idea for demonstrating these practical implications was a collaborative effort with my co-authors, particularly Dr. Peter Beling and Dr. Tyler Cody, I served as the sole author of the paper with feedback from co-authors to ensure consistency and fluency. A notable addition to the paper, introducing the philosophical aspect using beliefs and attitudes, was suggested by one of my committee members, Dr. Daniel Hoek. My detailed contributions to this paper are outlined below:

- Designed, developed, and conducted the experiment with the ResNet-50 model.
- Structured and authored most of the paper.
- Applied the idea of homeostasis vs. homeodynamics.
- Developed and formalized core-dominant vs. periphery-dominant systems.

These contributions collectively contribute to demonstrating the practical implications of core and periphery precepts in both biological and artificial intelligence, offering insights into the behavior of systems with core-dominant and periphery-dominant characteristics.

2.4.4 Contribution To Paper D

In the paper *"Transition From Scenario-Based to Outcome-Based Engineering in SE4AI"*, we presented concrete examples from the real world to illustrate the challenges faced by SE practices in addressing issues of scope and scalability when modeling intelligent systems.

This paper served as a follow-on from Paper B [6]. The initial idea and development of examples showcasing the utilization of closed systems precepts to writing the structure of the paper were spearheaded by me. While I took the lead, my co-authors, especially Dr. Alejandro Salado, contributed to shaping the final structure of the paper by offering valuable feedback and suggestions to enhance the presentation of SE practices in both the existing scenario-based paradigm and the proposed outcome-based engineering approach leveraging closed systems precepts. My detailed contributions to this paper are outlined below:

- Identified, structured, and described the scope and scaling problems in SE4AI.
- Developed the SysML models to illustrate examples of SE limitations in modeling outcome-based systems through scenario-based paradigms.
- Defined, elaborated, and mathematically demonstrated the applications of both functionally closed systems and informationally closed systems in modeling a simple AI system.
- Defined and elaborated on the importance of causality when utilizing informational closure.
- Structured and authored the entire paper.

Bibliography

[1] Niloofar Shadab, Tyler Cody, Alejandro Salado, and Peter Beling. Closed systems paradigm for intelligent systems. In *2022 IEEE International Systems Conference (SysCon)*, pages 1–8. IEEE, 2022.

[2] Nils Bertschinger, Eckehard Olbrich, Nihat Ay, and Jürgen Jost. Autonomy: An information theoretic perspective. *Biosystems*, 91(2):331–345, 2008.

[3] Tyler Cody, Niloofar Shadab, Alejandro Salado, and Peter Beling. Core and periphery as closed-system precepts for engineering general intelligence. In *International Conference on Artificial General Intelligence*, pages 209–219. Springer, 2022.

[4] W Ross Ashby. *An introduction to cybernetics*. Chapman & Hall Ltd, 1961.

[5] Niloofar Shadab, Aditya U Kulkarni, and Alejandro Salado. Shifting paradigms in verification and validation of ai-enabled systems: A systems-theoretic perspective. In *Systems Engineering and Artificial Intelligence*, pages 363–378. Springer, 2021.

[6] Niloofar Shadab, Tyler Cody, Alejandro Salado, and Peter A. Beling. A systems-theoretical formalization of closed systems. 2023.

Chapter 3

Conclusions

3.1 Summery

In this dissertation, we have diligently pursued the achievement of three primary objectives through a comprehensive series of tasks:

To fulfill the first objective, we undertook a comparative analysis between open and closed systems perspectives, highlighting the inherent limitations of the open system view in representing intelligence as a relational property. Subsequently, we advocated for the adoption of a closed system perspective, thus establishing intelligence as a relational attribute encompassing the system and its contextual environment. This objective culminated in the formulation of an intelligence property framework predicated on need-outcome relationships and the principles of closed systems.

For the second objective, we embarked upon an extensive exploration of functional and informational closure types within closed systems, presenting a unified systems-theoretic formalism to elucidate these concepts. We proposed the practical implementation of the closed systems doctrine within the domain of SE, delineating frameworks that demarcate the boundaries of closed systems and their environment, while diligently accounting for stakeholder requirements and system design constraints. Central to this endeavor was our argument that grounding this approach in the core and periphery concept empowers the design and engineering of intelligent systems across multiple levels of abstraction. Notably, these levels may vary in terms of informational closure, some being informationally closed while others retain informational and functional openness. This perspective posits a method-agnostic solution to significant challenges encountered when engineering intelligence across multiple system levels.

For the third objective, we conducted empirical investigations employing ResNet-50, a CNN model, trained with CFAR-10 and CFAR-100 datasets. This rigorous empirical analysis aimed to substantiate the applicability of the core and periphery concept. Furthermore, we transitioned from this foundational exploration to a real-world exemplification, where we formally applied the closed systems doctrine. In doing so, we addressed the limitations of current SE processes concerning intelligent systems, particularly in the context of scoping and scaling intelligence. We harnessed the concepts of functional closure and informational closure as pivotal tools for delineating precise boundaries around intelligent systems, effectively extending their physical bounds to facilitate the resolution of diverse aspects associated with the scoping and scaling of intelligence. The exemplification through core and periphery and closed systems precepts presented herein offers a preliminary insight into the practical applicability of the concepts expounded within this dissertation.

3.1.1 Contributions

In the pursuit of addressing gaps in SE4AI, this dissertation has made significant contributions to the field by unraveling novel insights and advancing existing knowledge. The study navigates through the complexities of engineering global intelligence property through the lens of need-outcome relations, shedding light on key findings that not only answer the posed research questions but also extend the boundaries of understanding in SE practices, particularly, SE4AI. This research not only establishes itself as a unique contribution to the methodological toolkit but also introduces fresh perspectives to theoretical frameworks. This section outlines the substantial contributions of the dissertation, emphasizing its impact on both theoretical discourse and practical applications within the broader [your field] community.

- This study puts forth compelling arguments advocating a paradigm shift in the realm of SE4AI to accommodate intelligent systems effectively.
- It contributes significantly to SE theory by offering formal definitions for closed systems concepts and elucidating variety across different abstraction levels of a system.
- This research serves as a catalyst for systems engineers, empowering them to construct methodologies capable of elucidating intelligent outcomes in complex systems.
- By proposing a formal framework, this work not only facilitates the direct engineering of outcome-based problems but also lays the groundwork for scaling and scoping intelligence as a global property.
- The research introduces a novel layer of abstraction in the SE of intelligent systems, promising a streamlined process that potentially reduces the number of use-case models in the SE workflow.

3.2 Limitations

The primary limitation in this dissertation was the absence of a testing or operational environment to observe real-world intelligent systems and assess how the application of the various precepts introduced here could enhance the modeling of the relational nature of intelligence between the system and its environment.

Consequently, we relied on descriptive examples and confined ourselves to working with pre-existing datasets and deep learning algorithms.

Another constraint involved the lack of formalism in numerous concepts related to systems theory and systems engineering. Progressing through this dissertation required defining and formalizing several concepts. While this dissertation contributed to the formalization of some concepts, it is acknowledged that numerous other concepts await formalization and exploration in future research endeavors.

The following list outlines the substantial limitations of this dissertation:

- There is a notable absence of empirical evidence illustrating the practical application of the proposed framework in real-world intelligent systems. This gap not only highlights the necessity for empirical validation but also underscores the need for a new breed of measurement techniques tailored to assess the effectiveness and performance of such frameworks.

- The absence of a concrete framework regarding what specific aspects should be observed and measured in empirical evidence poses a critical obstacle.

- There is a deficiency in the formalism associated with certain concepts within systems theory and SE.

- The abstract level of the formalism provided in this research, currently falls short of suitability for practical engineering applications.

- The research acknowledges that existing AI systems often fall below the expected threshold of intelligence expression. Consequently, the true applicability of the proposed framework remains untested and awaits recognition.

3.3 Future Work

As we conclude this dissertation, it is important to acknowledge the avenues for future work that emerge from the groundwork laid herein. The exploration of concepts such as functional closure, informational closure, and the core and periphery model has opened doors to new perspectives in SE4AI. However, the journey is far from complete, and several promising directions for future research beckon. In the realm of practical applications, empirical studies can shed light on the real-world impact of these introduced precepts on SE practices, particularly in the context of intelligent systems. Furthermore, a deeper dive into formalizing additional concepts promises to enrich the theoretical foundation of our understanding. As we look forward, the need for empirical evaluations, formal elaborations, and comparisons with traditional engineering methods becomes evident. The dynamics of boundaries between highly coupled systems, as well as the human-centric aspects of incorporating these precepts in engineering workflows, represent compelling domains for continued exploration. The journey continues, and the future promises exciting opportunities to refine, expand, and apply the insights gleaned from this dissertation. In the following subsections, we will briefly discuss such avenues for future work in more details.

3.3.1 Investigation of Practical Values

Future work should focus on substantiating the practical value of the proposed concepts. Firstly, it is essential to empirically assess the capability to isolate system functions through core and periphery disambiguation, particularly in systems demonstrating general intelligence. A detailed comparison with traditional SE methods is also warranted. Lastly, the determination of core and periphery involves defining boundaries between systems, necessitating further research on the dynamics of boundaries, especially in the context of highly coupled systems. [1].

Moreover, to assess the practical value of implementing the introduced precepts in SE practices, such as Model-Based Systems Engineering (MBSE) or document-based SE processes, empirical research involving human subjects is crucial. This research should focus on investigating the enhancement or efficiency achieved through the application of these precepts in the day-to-day activities of systems engineers, particularly when dealing with intelligent systems. Potential areas of investigation may include context definition and engineering for intelligent systems, use case construction and decomposition, as well as decision-making processes based on outcome-based engineering.

3.3.2 Formalism of SE Concepts

The field of SE4AI has seen limited research despite numerous calls for exploration. While this dissertation has made contributions by formalizing concepts like functional closure, informational closure, core and periphery, there remains a need for the formalization of other crucial concepts such as *capability* and *property* within systems, particularly in the context of intelligent systems. Establishing a concrete formalism for these concepts could empower systems engineers to assess the realization of intelligent capabilities and quantify relational properties at a system-level concerning the environment.

Furthermore, there is a need for a comprehensive and formal elaboration of core and periphery, supported by mathematical theorems and corollaries.

3.3.3 Revisit of the SE Practices

Following our work on utilizing openness and closedness framework in engineering intelligent systems at different levels of abstractions with regards to the spectrum of openness versus closedness, the paradigm shift resulted from the works of this dissertation leads to the need for a revisit in several SE activities that are dependent on such formulation. Here, we aim to shed light on the required modifications of these types of activities based on the intelligent features of a closed system.

- **Needs vs Requirements:** Based on this viewpoint, the first traditional SE concept that needs to be evaluated in using for intelligent systems is the concept of *needs and requirements*. Requirements are used directly to engineer functionality of the systems. However, outcomes cannot be engineered directly from requirements or functions. Meaning that engineers should be able to directly engineer outcomes from needs space. The outcomes emerge from the relations between system's needs and context varieties. Addressing the question of how to engineer and/or bound such outcomes could be one of the main tasks of systems engineers

- **Verification and Validation:** Following the shortcomings of the requirement-functionality aspect of engineering intelligent systems, we could encounter similar problem with verification and validation (V&V) activities. V&V activities are fundamentally developed based on the relationships between requirements and the functions and performance of the systems while holding the assumption of behavior preservation of systems [2]. Even though, validation activities focus on the stakeholder's needs and outcomes, it relies on the open systems practices to satisfy those needs. We believe the nature of this satisfaction cannot be achieved only by open system techniques. If for engineering intelligent systems, we require to achieve outcomes directly from needs, the current approach will not necessarily achieve the outcomes needed of an intelligent system. And to the extent intelligent systems intertwine with their contexts, behavior will not be preserved. Recreating this coupling in testing environments is not straightforward [3]. Verification strategies are considered bottom-up activities that can be traced by the networks of parent-child requirements. Most of the current capabilities of the systems can be achieved and verified by this approach of requirement engineering and V&V activities. There are a few other capabilities such as sustainability that cannot be directly verified using deductive logic in the set of verification activities sets. Intelligent capabilities, however, could be different than those capabilities such as sustainability as those capabilities can be easily bounded with regards to the project's constraints. Limits to learning, fidelity of V&V environments, evolution of performance during operation, learning transferability, could be considered as the reasons to change the paradigms of V&V activities to suit with the new types of SE.

- **Reusability:**The concept of reusability for intelligent systems needs a serious revisit as we no longer work with systems that change exogenously in the presence of different context/environment. How we can assess the effect of these changes when we want to use these systems in different contexts, requires systems engineers to develop new Interface Control Documents (ICDs), new protocols for change, as well as new ways to translate systems outcomes from one context to another one [4]. Additionally, the practice of translating functions and requirements between contexts needs to be extended to also translate needs and outcomes.

- **Product Line in Manufacturing** In a product line, some units get tested and some do not. Each system can be exposed to a specific, pre-determined set of missions after the manufacturing process. For intelligent systems, every mission can influence the intrinsic behaviors of the system. As a result, the behavior of each system can diverge.

Yet, based on current practice, some of these systems skip some tests and evaluations during the manufacturing process on the assumption of equivalent (isomorphic) behavior and performance [5]. Currently, these settings in a product line cannot suffice to identify the intrinsic changes in intelligent behaviors. This dissertation could provide a roadmap on how principles of core and periphery can be utilized to enable mass production of core while having periphery parts to be engineered in relations with their unique environment. The closed systems precepts combined with the concept of core and periphery can be investigated to provide a better solution to overcome shortcomings of manufacturing intelligent systems in a product line.

- **Systems Sustainment:** Sustainment is another concept in SE of intelligent systems that needs to be revisited. For intelligent systems, we need need to answer questions such as: 'What does maintainability look like for AI?'; 'What does refreshment mean for learning capability of the system'; 'Is there a need for resetting data?'; 'Does AI yield any type of degradation that we do not understand yet?'. These questions could be addressed through closed systems engineering.

- **Systems Management:** Systems management of intelligence can be another research roadmap that requires close attention from the SE community. Questions such as: 'Can new formulations of intelligent systems better assist in top-down managing/designing/-operating intelligent systems?', 'Are we really content with Artificial intelligence as a bottom-up phenomena that expands outward from sub-systems?' could be raised in SE practices. Answering these questions seems critical to understand the system management of intelligent systems through closed-system vs open-system view.

Bibliography

[1] Tyler Cody, Niloofar Shadab, Alejandro Salado, and Peter Beling. Core and periphery as closed-system precepts for engineering general intelligence. In *International Conference on Artificial General Intelligence*, pages 209–219. Springer, 2022.

[2] Jie M Zhang, Mark Harman, Lei Ma, and Yang Liu. Machine learning testing: Survey, landscapes and horizons. *IEEE Transactions on Software Engineering*, 2020.

[3] Niloofar Shadab, Aditya U Kulkarni, and Alejandro Salado. Shifting paradigms in verification and validation of ai-enabled systems: A systems-theoretic perspective. In *Systems Engineering and Artificial Intelligence*, pages 363–378. Springer, 2021.

[4] Niloofar Shadab and Alejandro Salado. Towards an interface description template for ai-enabled systems. *arXiv preprint arXiv:2007.07250*, 2020.

[5] Niloofar Shadaba, Tyler Codya, Peter Belinga, and Alejandro Saladob. Product herding for intelligent systems. 2023.

Appendix A

Closed System Paradigm for Intelligent Systems

Niloofar Shadab, Tyler Cody, Alejandro Salado, Peter Beling

SysCon 2022 - 16th Annual IEEE International Systems Conference, Proceedings

DOI: 10.1109/syscon53536.2022.9773829

Abstract— Intelligent systems ought to be distinguished as a special type of system. While some adopt this view informally, in practice, systems engineering methods for intelligent systems are still centered around traditional systems engineering notions of engineering by aggregation of components. We posit that this traditional approach follows from holding a notion of open systems as the fundamental precept, and that engineering intelligent systems, in contrast, requires an approach that holds notions of closed systems as fundamental precepts. We take a systems-theoretic approach to defining closed system phenomena and their relation to engineering intelligence. We propose the concept of variety; particularly the law of requisite variety to enable closed view in engineering. We discuss how open and closed view approaches to engineering intelligent systems address variety differently, as well as the implications of this difference on engineering practice.

A.1 Introduction

Distinguishing intelligent systems as a special type of system is valuable if it informs different principles for applying systems engineering (SE) practices. Historically, a new category of systems emerges every time that we need to recognize new system's properties [1]. We provide a brief overview of different categories of systems and why we believe that intelligent systems should be treated as a separate category of systems.

- **Traditional Engineered Systems:** These systems are engineered such that they have only exogenous evolution. Traditionally, systems were discipline-centric [1]. Expertise in one domain was sufficient to develop such systems. They were referred to as mechanical systems, electrical systems, etc. Later, more complex systems that integrated traditionally separated engineering systems became relevant. Traditional engineering disciplines were insufficient to develop those systems successfully [2]. As a result, SE was born primarily as an integrative discipline of other engineering disciplines and the term engineered systems was coined.

- **Socio-technical systems:** These systems recognize the interactions between humans and technology. As the scope of systems expanded, people became an active part of the system and traditional SE techniques were not sufficient to address the human aspects of engineering such systems [3]. Command and control, a common paradigm in traditional systems, becomes unpredictable due to the agency of humans. Mechanisms based on influence and persuasion must be introduced instead, with the support of techniques such as agent-based modeling, human-machine integration, etc., were developed to broaden the scope of related SE efforts [4, 5]. Therefore, the category of socio-technical systems was born out of this expanded scope.

- **System of systems:** These systems can be defined as a collection of systems that interoperate to achieve a common goal [6]. As the scope of systems continued to expand, systems engineering's focus turned to large scale systems consisting of interacting constituent systems—corresponding to notions of aggregates of systems that have independent governance, autonomy, or management, but that can work together to provide unforeseen capabilities [7]. This independence makes some traditional SE activities inapplicable or significantly different from how it is applied in traditional systems. For example, there may be a lack of authority to impose requirements on some constituent systems. Similarly, it may be impossible to access verification information from some other constituent systems, so verifying the whole system of systems may not even be a feasible consideration. Consequently, a larger category of systems called System of Systems (SOS) was recognized in SE [7, 8].

- **Intelligent Systems:** Intelligent systems here are specified as engineered systems that have endogenous evolution of behavior, and/or function over their life-cycle [9]. We posit that engineered intelligent systems [10] also require different techniques from those traditionally used in SE. If this is the case, then it makes sense to distinguish intelligent systems from those that are not intelligent. To do so, we need to elucidate the notion of intelligence in engineered systems. In fact, there is an ongoing research on what intelligence means and how it exhibits its properties in engineered systems [11]. Each definition could lead the discussion of engineering intelligence in a different direction. While there is not a definitive description for what intelligence is, we propose a framework for engineering intelligent systems that exhibit certain behaviors and structure. Our framework builds upon the concepts of open and closed systems, where open systems are those that exchange information, matter, and/or energy through their boundaries and closed systems are those that do not exchange information, matter, and/or energy through their boundaries [12].

> Following the traditional SE paradigm, intelligence has been relegated to sub-systems (i.e., AI algorithms). This allows for a perpetuation of open-system-type (input-output) techniques to the domain of intelligent systems [13]. Similar to how domain expertise in mechanical or electrical engineering was sufficient for designing mechanical or electrical systems until complexity arrived [14], treating intelligence as an addressable component will likely be insufficient for a scalable intelligent property. Current practice has a hyper-focus on the components of intelligence, rather than intelligence of the system as a whole [15].

In this paper, we provide an argument on how open-system view in SE practices fails to capture the unique features of intelligent property. As most of its activities lie in the open-system paradigm, SE lacks the capability to engineer and characterize intelligence as a global property[1]. Therefore, a new paradigm is needed to address intelligence and untangle its characteristics. We aim to institute a closed-system approach to characterize intelligence as a property of system's relation to its context by emboldening the shift from the concepts behind the requirement-functions relationships in SE and aligning them with the presented closed systems engineering view.

The paper is structured as follows. First, the relationship between intelligence and learning is discussed. Next, open and closed systems are introduced. Then, with these foundations, we discuss the need and role of taking a closed systems view in engineering intelligent systems. In particular, we explore how the Law of Requisite Variety can be used to develop precepts for engineering with a closed systems view. Finally, we remark on the need to revisit existing best practices in SE before concluding.

[1] i.e., in terms of the closed relationship between system and context, not the system in isolation

A.2 Intelligence vs Learning in SE

In the introduction section, we discussed about how intelligent systems could be considered as a different category of systems in SE activities. To address this change, we need to be able to distinguish the concepts of intelligence and learning, and set the proper relations for these two concepts as the basis for intelligent engineered systems. We provide three possible scenarios where intelligence and learning capabilities could be engineered in intelligent systems depending on what is considered as the nature of intelligence in such systems. In each scenario, the relationship between the two concepts of intelligence, and learning is different. In intelligent systems, we can consider learning as a consequence derived from intelligence property. Intelligence, however, could be either considered as an inherent property of an open system or a measure for the outcome of a closed system. Depending on our definition of relations between learning and intelligence, SE practices could vary significantly. The three possible situations deriving from learning-intelligence relationships are as followed.

- **Situation 1:** In this situation, learning itself is a **representation** of intelligence. The current practices of AI-centric systems fall into this category. Learning happens from patterns identified in training data. This training data can be gathered either off-line or online through experiences of actions in the real world. But there is no separate property as intelligence in these systems. In this case, traditional SE practices could be sufficient for reverse engineering of intelligence from learning as they only need to follow input-output paradigm for their activities. However, in this situation, both learning and intelligence would be considered a property of the system. As a result, they can be engineered and performed independent of their context. This means that intelligence and learning both are considered as a local phenomena. Yet, many of the bottlenecks in engineering practices of intelligent systems– such as “catastrophic inference” or "drift of concepts" [16, 17]–occur as a consequence of considering intelligence as being independent of its context. For example, in autonomous cars, traditional SE practices may be suitable for well-structured, highway contexts, but are lacking when it comes to unstructured contexts or those that have moral, ethical or legal burdens [18].

- **Situation 2:** In a more complicated situation, learning is a **consequence** of intelligence. In other words, intelligence is a property of the system that enables learning capability. However, in this situation, intelligence is a property that needs to be engineered directly to provide an appropriate context for the system to achieve learning as a desired capability. This approach requires a solid and universal definition of intelligence. However; the definition of intelligence is different from one field to another [19]. Hence, there is no solid approach currently available to characterize, engineer, or measure intelligence as a property of a system. In their efforts to ensure desired outcomes, systems engineers are therefore limited to engineering and testing concrete examples of the desired learning capability. In this situation, systems engineers characterize intelligence gradually over the life-cycle of the system by observing learning capabilities and associating these observations with the characterization of intelligence property in systems. This leads systems engineers vulnerable to potential risks coupled with the limited knowledge of the full characterization of intelligence over the system's life-cycle.

- **Situation 3:** In this case, intelligence is a relation property of the system with regards to its contexts; it means that regardless of considering learning as a representation or a consequence of intelligence property, they both are embodied in the context. Hence, context should be distinguished as a part of the intelligent system. The intelligence property can be defined, engineered and characterize as a relation between context and the system to reach a new equilibrium. In conclusion, in this situation, intelligence is no longer relegated to a subsystem. Therefore, rather than treating intelligence as a local phenomena, systems engineers should treat it as a global phenomena.

Following Situation 3, intelligence cannot be characterized using the current SE approaches. To distinguish context and system as one, we need to adopt a closed view. Due to the closedness of intelligence as a global property, characterization of intelligence cannot be accomplished through only decomposition, feedback, and integration relationships inside the system components. We could consider it as a global property that could be a measurement for a system's success in different contexts [19].

In traditional systems, some measures are employed to form confidence in how and if the system can achieve its outcomes. Outcomes in the context of traditional systems are defined as achieving stakeholder's needs. Therefore, the fundamental principle to engineer realization of outcomes in a traditional system is requirement-function relations. However, we cannot exert this view in Situation 3, as outcomes should be derived directly through the intelligence property–but the intelligence property is not local to a system; it is dependent on context.

Consequently, direct engineering of system functions in a manner that is isolated from context is insufficient to achieve needed outcomes from an intelligent system. That is, context should be a part of the requirement-function relations. While there is necessarily a portion of needs that can be derived into functions, e.g., sensing and acting, we posit that in intelligent systems, there is necessarily a portion that cannot be.

The main consequence of selecting Situation 3 is that, in contrast to traditional systems (and other aforementioned system categories), which are well characterized by input-output relations between subsystems, intelligent systems are highly subject to system's level need-outcome relations. That is, intelligent system's need-outcome relations cannot be totally decomposed into lower levels of abstraction, e.g., into restrictions on the input-output behavior of subsystems. In Situation 3, intelligence belongs to the system and its contexts. This interpretation brings the closedness attribute to the intelligence property. The dependence of system's inputs on outputs in these systems challenges notions of boundaries between subsystems or components, and thereby practices of engineering by composition or aggregation.

Here, we tackled different contexts that the SE community can deal with intelligent systems. In this paper, we concentrate on Situation 3, and propose closed systems approaches to overcome the hassle of engineering intelligence property in systems. Thus, in the next section, we provide a comprehensive background on open vs closed systems theory. Then we define our notion of closed systems using Mesarovic's System's Theory.

A.3 Closed System View Over Open System

In this paper, we establish the position that the practice of engineering intelligent systems ought to anchor its precepts to *open* and *closed* views of systems. And, moreover, that scaling intelligence relies on the latter more than the former. Here, we introduce these two principal views and their ties to SE practice.

The open and closed views are termed in reference to open and closed systems. Common notions of open systems are held across disciplines. To early general systems theorists and biological systems theorists, open systems were simply those that had external interactions. All that was required was a boundary between the *internal* and the *external* and interactions across the boundary [12]. This simple view can capture richness in the exchange of matter, energy, and information between a system and its environment [20]. It is commonly held, with different disciplines adding their own particularities as needed. In SE, it is reflected by the notion of input-output systems [21].

The open system view of boundaries and interactions, when taken to the extreme, focuses exclusively on the system within the boundary. This *internal* system is taken to be the system and the *external* system is taken to be everything else. Systems analysis proceeds by the decomposition of the internal system into components, and systems engineering proceeds by the aggregation and composition of those input-output component systems into the internal system [21, 22]. Although the abstract notion of open systems leaves room for consideration of the external system in tandem with the internal, in SE practice, a focus on the input-output nature of open systems leads to an unintended fixation on the parts instead of the whole, and a modus operandi of, 'If the parts work, and the interfaces between them work, then the whole will work.' This treatment will not hold for intelligent systems.

As we will explore, intelligent subsystems cannot be appropriately decomposed and treated as separate components—they are a part, perhaps, but a part that is closely coupled with, if not inseparable from, the whole. This means a drastic shift from the traditional, long-standing SE practice of composition and aggregation; however, as we will argue, it may well be necessary to scale intelligence. To make our case, we adopt a particular notion of closed systems.

Notions of closed systems are less consistent across disciplines than notions of open systems. In biology and natural sciences, an open system is a system whose border is permeable by matter and energy, while a closed system is only permeable by energy [23]. In control theory, closed systems are open systems where the input is composed with feedback to adjust the output [24]. In the main, *closedness* is generally used to describe something about the nature of open systems' boundaries, as in biology and natural sciences, or its use of feedback to adjust interactions, as in control theory.

Our notion of closedness is not characterized by feedback *per se*, or by the nature of the boundary, but rather by the non-existence of inputs and outputs. That is, under the closed view, at the highest level of abstraction there are only outcomes. This distinction may seem laborious and overly abstract, but the difference in engineering practice is profound. In essence, the systems engineer is not faced with the question of, 'Is the system producing the *outputs* we require from the *inputs*?', but rather, 'Is the system achieving the *outcomes* we *need*?'

We consider both open and closed systems to have a boundary, but consider open systems, at their most general level, to be input-output systems, and closed systems, at their most general level, to have inputs which are entirely coupled to their outputs.

This is represented in Figure A.1. By our definition, feedback systems, so-called "closed" systems in control theory, are open—they are open systems with a closed subsystem. Similarly, cybernetic systems characterized by an internal feedback mechanism that cannot be decomposed into a composition of input-output systems, such as Mesarovic's goal-seeking systems [13], are also open.

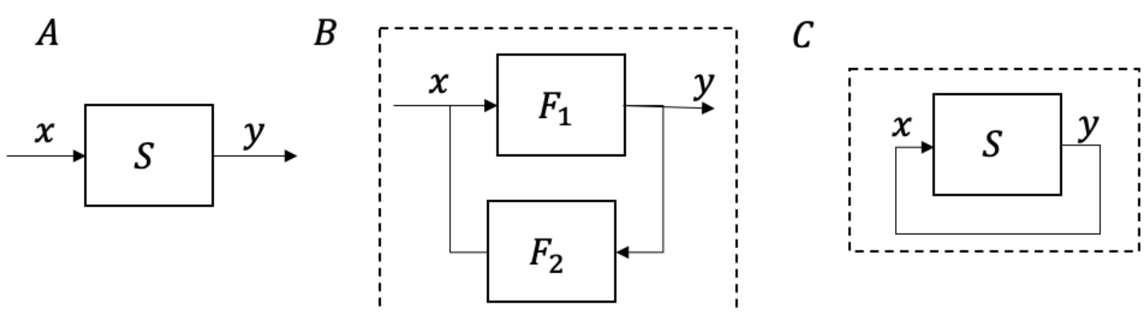


Figure A.1: Block-diagrams of open systems [A], systems with feedback loops [B], closed systems [C].

The open view of systems establishes a boundary between the external and the internal, termed the *world* $\mathcal{W}$ and the *system* S, respectively, and defines that which passes in through the boundary, termed the *inputs* $\mathcal{X}$, as a thing apart from that which passes out through the boundary, termed the *outputs* $\mathcal{Y}$. If we take the formal-minimalist worldview that a system is a relation on sets, we can formalize an open system S as a relation

$$S \subset \times\{\mathcal{X}, \mathcal{Y}\}$$

where $\mathcal{X} \cap \mathcal{Y} = \emptyset$.

The closed view of systems also considers a boundary or scope that distinguishes the external $\mathcal{W}$ from the internal S, however, in contrast, it does not necessitate that the inputs and outputs are mutually exclusive, but rather that there exists a complete influence of the entire set of outputs $\mathcal{Y}$ on the entire set of inputs $\mathcal{X}$. That is, $\exists f : \mathcal{Y} \to \mathcal{X}$. Thus, closed systems are differentiated from open systems by dissolution of boundaries between the external $\mathcal{W}$ and the internal S. This is *not* the same as the notions of closedness in feedback control systems or classical cybernetic systems—we are not specifying the closedness of some subsystem, but rather of the system as a whole, at its highest level of abstraction. This closedness can be captured by the tight coupling of intelligent system and its context. The tight coupling itself can be modeled by having the entire set of outputs and the entire set of inputs to the intelligent system.

Let us underscore this distinction formally. Consider an open system $S \subset \times\{\mathcal{X}, \mathcal{Y}\}$ where, as is common, $S : \mathcal{X} \to \mathcal{Y}$. A feedback control system uses a subsystem S' to modify inputs $\mathcal{X}$ with a control $\mathcal{Z}$ to adjust the output $\mathcal{Y}$. Put formally, S is termed a feedback control system if

$$(x, y) \in S \leftrightarrow \exists z[(x, z), (y, z) \in S']$$

where $x \in \mathcal{X}, y \in \mathcal{Y}$, and $z \in \mathcal{Z}$. In other words, so-called 'closed' feedback control systems are actually a composition of open systems which, at the highest level of abstraction, result in an open system.

Now consider a classical cybernetic system. A cybernetic system is classically described as a system with internal feedback characterized by goal-seeking. Put formally, a cybernetic system is a system S where the control mechanism for selecting $z \in \mathcal{Z}$ is specified by a set of consistency relations

$$G : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to V, E : \mathcal{X} \times \mathcal{Y} \times V \to \mathcal{Z}.$$

G is termed the goal or value relation and E is termed the seeking or search relation. Together they specify the goal-seeking nature of S, but not by decomposition [25]. That is, S' cannot be formed by composing G and E. And thus, the consistency relations in cybernetic systems form a closed subsystem—there is some closedness within S which cannot be composed away as in feedback control—however, at the highest level of abstraction, the system is still open.

The principal failings of the open view as a basis for engineering intelligent systems are two-fold. First, under the open view, the prototypical approach to engineering intelligence is to treat intelligence as a subsystem. In taking the open view, one ignores the weakening of boundaries and increased prevalence of interdependence and coupling associated with intelligent systems, thereby neglecting, or at least muddling, any understanding of intelligence as a global character of systems. Second, and moreover, the over-appreciation of boundaries leads to a characterization of intelligent systems in terms of functions and requirements. It is our position that engineering intelligence, and particularly engineering scalable intelligence, demands a shift from traditional notions of *functions* and *requirements* to notions of *needs* and *outcomes*.

A.4 Engineering Intelligent Systems

In earlier sections, we argued that the open view is a limiting paradigm for engineering intelligent systems. It places an over-emphasis on the functional nature of systems. Corresponding engineering methodology restricts itself to specification on the relationship between inputs $\mathcal{X}$ and outputs $\mathcal{Y}$. Under the closed view, in stark contrast, at the highest level of abstraction inputs $\mathcal{X}$ and outputs $\mathcal{Y}$ cannot be differentiated; there are no distinguishable inputs and outputs and therefore no notions of functional requirements about which to orient SE efforts. In this context, instead of outputs, closed systems may be characterized in terms of outcomes. These outcomes are accomplished to meet system's needs [26]. Therefore, deriving requirements from needs are replaced by working directly on the needs space.

To build a better understanding of need-outcome and requirement-function relations, we provide the following example. When we inflate a balloon, we may think that the function was 'inflating a balloon'. But it is more appropriate to say that we have a function of blowing air, and the balloon has functions of stretching and holding air. In this situation, the inflated balloon is an outcome of these functions interacting. Functional decomposition and analysis have their limits. Requirements are tied to functions, but needs are tied to outcomes. When it comes to engineering intelligence, those outcomes are not readily derived or partitioned by the typical SE approach of divide and conquer.

This balloon example has parallels in machine learning. Take neural networks for instance. We can specify neural networks as a composition of functions. They have a function of passing information from layer to layer. They require certain mathematical structure when they do it. And so on. And this functional break-down of neural networks applies to the enormous number of systems where similar neural networks are applied. Neural networks have nearly the same mechanics in every application. In neural networks, it is apparent that the interesting part is not their mechanical nature, because it is the same nearly everywhere. The requirements-function breakdown barely changes, if at all from one application to another. But the outcomes are different. Using those homogeneous functions, the neural network achieves outcomes that are not homogeneous [27]. We posit that to engineer outcomes we need a different approach, focused not on the partitioning of an intelligent system into functions, but rather on the functioning of the intelligent system as a whole.

This is exemplified by the current practice of using so-called requirements like 'the environment must be stationary to use machine learning'. This takes a functional approach to an outcome—stationarity. Not only is it an inappropriate approach, but also it muddles the true requirements-function issues by blending in the more abstract, complex, and difficult-to-deal-with issues related to needs-outcomes.

The closed phenomena stand in contrast to typical input-output phenomena, as they are not appropriately formalized as open systems. This means that traditional engineering treatments of functions and requirements, decomposition and re-composition, and combinatorial testing cannot be readily applied. For example, consider that closed-systems phenomena, such as action-perception, may have no true, independent variable—either for AI to use in scientific induction or for the engineer to use in analysis.

The examples in this section indicate the need to consider the context that the system resides in within the boundary of the closed system view. The environment can no longer be assumed outside of the system and stationary as it plays a significant role on the outcome of the system. This is in contrast with the traditional systems function-requirements relations.

In Figure A.2, we demonstrate the steps that SE community should take to transfer from open systems domain of engineering intelligent systems to the closed systems domain of SE. Our basic position is that while the open systems view may suffice for awhile, at some

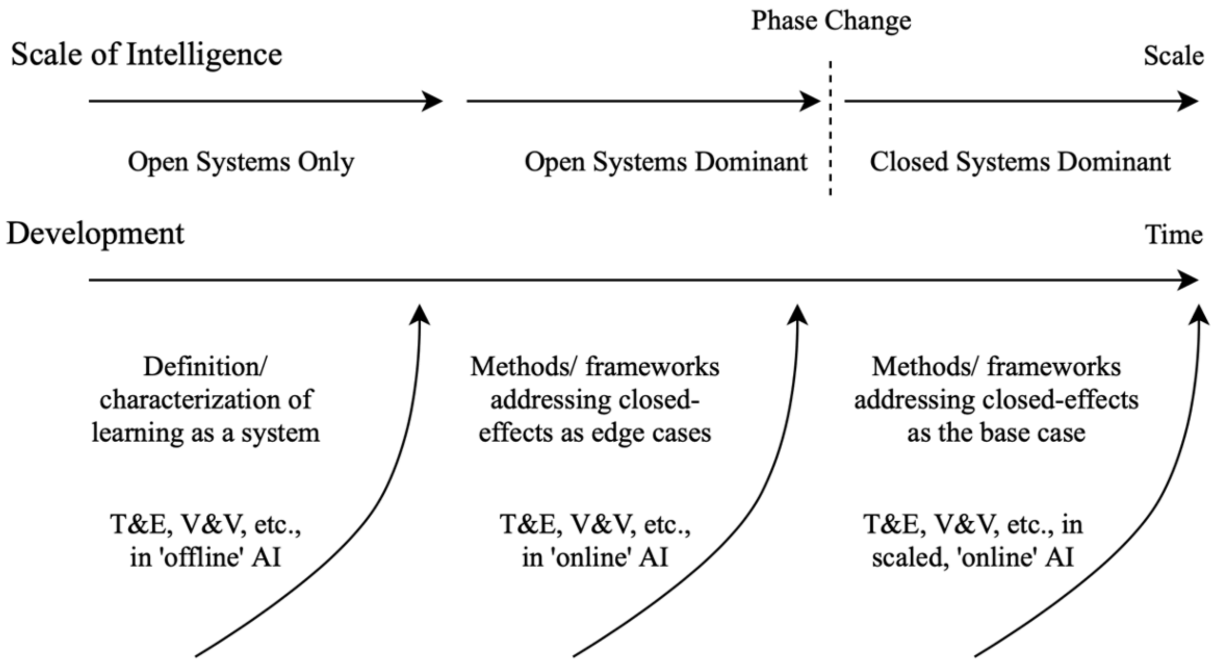


Figure A.2: Closed and Open Systems Engineering Practices Spectrum.

point, when scaling intelligence, there is a phase change where the closed systems phenomena become the dominant means of achieving needs and outcomes.

In the next section, we elaborate on how the concept of variety and the law of requisite variety could be taken advantage of to enable closed and open systems engineering of intelligent systems.

A.5 The Law of Requisite Variety

For complex systems, finding a definitive solution to predict all possible system's states could be infeasible. Therefore, it is safe to assume that an observer will not have a complete understanding of all the future states of the system [28]. This is due to the complexity of the system as well as the unpredictability of its environment. As a result of this partial knowledge, system's behaviors can be construed through the observer's goals and the context where the system operates in [29].

To engineer such systems with incomplete knowledge, the concept of variety could be employed [30]. Variety is defined as the number of different elements in a system's state. In the case of intelligent systems, one might not know of the entire set of varieties of an intelligent system while engineering it. In this situation, the Law of Requisite Variety can be utilized to help put a lower bound constraint on the possible number of varieties the system can exhibit. According to this law, variety in a system's state cannot be more than the variety of its environment stimuli that is transitioning to the system at that state [30]. However, the variety of system's outcomes should be equal or greater than the difference between the variety of the environment and the variety of the system. In traditional engineered systems, the set of system's states is static; meaning that all the components have finite set of plausible states to be in, even if some of these are unknown. In addition, the processes of transferring varieties mainly consist of one-to-one or many-to-one transformations; thus enabling the system to reduce the number of possible varieties at each state, and simplifying prediction of the set of future system states.

In contrast, in the case of intelligent systems, variety of context can enter the system and create one-to-many or many-to-many transformations in the system's variety. What makes such a system to survive in specific states is the fact that the set of variables essential to the survival will be bounded in an acceptable range [28]. Therefore, it is important for an intelligent system to be able to block varieties in the environment that prevent the essential variables to remain in their boundaries [31]. These essential variables belong to the first form of variety. We call this form of variety *bounded variety* in systems. This type of variety builds up the core features of an intelligent system. These core features should remain stable during system's operations by keeping the bounded varieties within their bounded range.

In addition to the first form of variety, intelligent systems require a second form of variety which is responsible for one-to-many or many-to-many transformations that lead to intelligence property. This form is called *unbounded variety*. These varieties should respond to context variety and help the system to learn while protecting bounded varieties from context variety. These notions suggest a hierarchical structure of system's varieties to enable both open view and closed view for engineering intelligent property.

Figure A.3 summarizes how two forms of varieties could fit in the open vs closed systems argument. As it is depicted in Figure A.3, the bounded form of variety can be engineered using the requirement-function relations (open system paradigm). The second form of variety, unbounded variety, should be engineered through closed systems engineering of outcome. Bounded varieties will be preserved in the presence of context variety by the help of different types of regulators. This process creates desired outputs from the core. On the other hand, unbounded varieties in the system will create a closed system with context varieties and in line with the core outputs, create desired system's outcomes. With that in mind, in the next section, we propose the concept of core and periphery to enable such distinction in engineering intelligent systems.

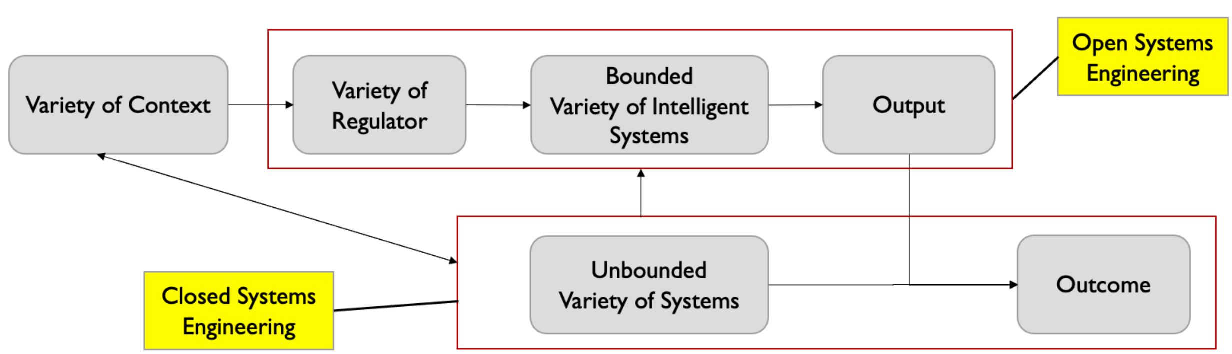


Figure A.3: Relevance of Two Types of Varieties to Closed and Open Systems Engineering

A.6 Notions of Variety, Core, and Periphery

Here, we elaborate on the concepts of core and periphery and how they fulfill the requirement of encompassing both bounded and unbounded varieties to engineer intelligence property. We use the Law of Requisite Variety to elucidate the interplay between the open and closed views in engineering intelligence.

Simply put, the Law of Requisite Variety posits that in order to control the external $\mathcal{W}$, the variety internal to S must match the variety in the external $\mathcal{W}$. It follows, then, that in engineering systems there are two principal varieties of concern: those varieties which we bound and those which we do not. Those which we bound form the system's core, its main characteristics around which variety is layered outwardly to the system's periphery.

Th core is secured from disorder. Thus, the core is naturally fragile. Many times engineering the core corresponds to establishing an open functionality and involves, for example, the derivation of systems-level requirements at lower levels of abstraction. In the core, following the open view, the focus is on *if* information flows and *how* it flows.

The periphery refers to those parts of the system where variety is left unbounded. In contrast to the core's focus on order, the periphery is concerned with the system's response to disorder. Whereas the core is dominated by the open-view, the periphery is appropriately dominated by the closed view, and thereby a consideration of needs and outcomes over functions and requirements. For example, what equilibrium between system and context needs to be maintained such that outcomes are satisfactory? In the periphery, following with the closed view, the focus is on *what* information flows and *why*.

Perhaps most importantly, strengthening the core and periphery is an imposition on the system as a whole, not just the subsystem or component where intelligence 'resides'. Thus, it is facilitated by both the strengthening of the 'intelligent-part' via better data, algorithm design, etc., and also the system more broadly. For example, the physical security of information and access to compute relate to the strength of the core. Or, more abstractly, the core can be strengthened by making sure the inputs entering the system from environments have certain orderly points of reference. For example, a self-driving system may be engineered with the assumption that it will never leave the highway. If so, then now the core not only includes access to power and compute or software, but also the system's use. In this way, the core of intelligence is distributed across the system. If the self-driving system is taken off of the highway, the core breaks, and the system fails.

A.7 Research Roadmap

Following our discussion on utilizing openness and closedness conceptualization in engineering intelligent systems as open systems or close systems, this paradigm shift leads to the need for a change in several SE activities that are dependent on this definition. Here, we aim to shed light on the required modifications of these types of activities based on the intelligent features of a closed system.

A.7.1 Needs vs Requirements:

Based on this viewpoint, the first traditional SE concept that needs to be evaluated in using for intelligent systems is the concept of *needs and requirements*. Requirements are used directly to engineer functionality of the systems. However, outcomes cannot be engineered directly from requirements or functions. Meaning that engineers should be able to directly engineer outcomes from needs space. The outcomes emerge from the relations between system's needs and context varieties. Addressing the question of how to engineer and/or bound such outcomes could be one of the main tasks of systems engineers

A.7.2 Verification and Validation:

Following the shortcomings of the requirement-functionality aspect of engineering intelligent systems, we could encounter similar problem with verification and validation (V&V) activities. V&V activities are fundamentally developed based on the relationships between requirements and the functions and performance of the systems while holding the assumption of behavior preservation of systems [32]. Even though, validation activities focus on the stakeholder's needs and outcomes, it relies on the open systems practices to satisfy those needs. We believe the nature of this satisfaction cannot be achieved only by open system techniques. If for engineering intelligent systems, we require to achieve outcomes directly from needs, the current approach will not necessarily achieve the outcomes needed of an intelligent system. And to the extent intelligent systems intertwine with their contexts, behavior will not be preserved. Recreating this coupling in testing environments is not straightforward [9]. Verification strategies are considered bottom-up activities that can be traced by the networks of parent-child requirements. Most of the current capabilities of the systems can be achieved and verified by this approach of requirement engineering and V&V activities. There are a few other capabilities such as sustainability that cannot be directly verified using deductive logic in the set of verification activities sets. Intelligent capabilities, however, could be different than those capabilities such as sustainability as those capabilities can be easily bounded with regards to the project's constraints. Limits to learning, fidelity of V&V environments, evolution of performance during operation, learning transferability, could be considered as the reasons to change the paradigms of V&V activities to suit with the new types of SE.

A.7.3 Reusability:

The concept of reusability for intelligent systems needs a serious revisit as we no longer work with systems that change exogenously in the presence of different context/environment. How we can assess the effect of these changes when we want to use these systems in different contexts, requires systems engineers to develop new Interface Control Documents (ICDs), new protocols for change, as well as new ways to translate systems outcomes from one context to another one [33]. Additionally, the practice of translating functions and requirements between contexts needs to be extended to also translate needs and outcomes.

A.7.4 Systems Sustainment:

Sustainment is another concept in SE of intelligent systems that needs to be revisited. For intelligent systems, we need need to answer questions such as: 'What does maintainability look like for AI?'; 'What does refreshment mean for learning capability of the system'; 'Is there a need for resetting data?'; 'Does AI yield any type of degradation that we do not understand yet?'. These questions could be addressed through closed systems engineering.

A.7.5 Systems Management:

Systems management of intelligence can be another research roadmap that requires close attention from the SE community. Questions such as: 'Can new formulations of intelligent systems better assist in top-down managing/designing/operating intelligent systems?', 'Are we really content with Artificial intelligence as a bottom-up phenomena that expands outward from sub-systems?' could be raised in SE practices. Answering these questions seems critical to understand the system management of intelligent systems through closed-system vs open-system view.

A.8 Conclusion

In this paper, we argued why open systems engineering for intelligent systems cannot be sufficient in various contexts. We did so by comparing open vs closed systems viewpoints and how the open system is unable to achieve intelligence as the global property. We then developed the idea of core and periphery to engineer intelligent systems to encompass both requirement-functions and need-outcomes relationships. The main contribution of this paper was to tie the concept of open-closed systems with other concepts in systems theory to form a foundation for engineering intelligent systems. These concepts included core-periphery and the law of the requisite variety.

This research aims to initiate a roadmap for systems engineers to engineer the next generation of intelligent systems. Future endeavours would include formalization of the closed systems view, identification of the risk factors of the engineering of unbounded variety, and dealing with instability of core and periphery parts in these types of system.

Bibliography

[1] Niamat Ullah Ibne Hossain, Raed M Jaradat, Michael A Hamilton, Charles B Keating, and Simon R Goerger. A historical perspective on development of systems engineering discipline: A review and analysis. *Journal of Systems Science and Systems Engineering*, 29(1):1–35, 2020.

[2] Andrew P Sage. *Systems engineering*, volume 6. John Wiley & Sons, 1992.

[3] Gordon Baxter and Ian Sommerville. Socio-technical systems: From design methods to systems engineering. *Interacting with computers*, 23(1):4–17, 2011.

[4] Koen H Van Dam, Igor Nikolic, and Zofia Lukszo. *Agent-based modelling of socio-technical systems*, volume 9. Springer Science & Business Media, 2012.

[5] M Kudret Yurtseven and Walter W Buchanan. Socio-technical system design: a general systems theory perspective. In *Proc. The International Conference on Engineering and Computer Education-ICECE'2013*, 2013.

[6] Louis S Metzger, George Rebovich Jr, Robin A Cormier, Spurgeon Jr T Norman, Deborah L Schuh, Peter A Smyton, Robert S Swarz, and Frederick C Wendt. Systems engineering guide: Collected wisdom from mitres systems engineering experts. Technical report, MITRE CORP BEDFORD MA BEDFORD United States, 2014.

[7] Charles Keating, Ralph Rogers, Resit Unal, David Dryer, Andres Sousa-Poza, Robert Safford, William Peterson, and Ghaith Rabadi. System of systems engineering. *Engineering Management Journal*, 15(3):36–45, 2003.

[8] John Boardman and Brian Sauser. System of systems-the meaning of of. In *2006 IEEE/SMC International Conference on System of Systems Engineering*, pages 6–pp. IEEE, 2006.

[9] Niloofar Shadab, Aditya U Kulkarni, and Alejandro Salado. Shifting paradigms in verification and validation of ai-enabled systems: A systems-theoretic perspective. In *Systems Engineering and Artificial Intelligence*, pages 363–378. Springer, 2021.

[10] Ryszard Tadeusiewicz. Introduction to intelligent systems. In *Intelligent systems*, pages 1–1. CRC Press, 2018.

[11] Pei Wang. On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37, 2019.

[12] Ludwig Von Bertalanffy. General systems theory. *The Science of Synthesis: Exploring the Social Implications of General Systems Theory*, 103, 2010.

[13] Mihajlo D Mesarovic and Yasuhiko Takahara. *General systems theory: mathematical foundations*. Academic press, 1975.

[14] Alexander Kossiakoff, William N Sweet, Samuel J Seymour, and Steven M Biemer. *Systems engineering principles and practice*, volume 83. John Wiley & Sons, 2011.

[15] Antonio R Damasio. Descartes' error and the future of human life. *Scientific American*, 271(4):144–144, 1994.

[16] Jie Lu, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang. Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12):2346–2363, 2018.

[17] Pauching Yap, Hippolyt Ritter, and David Barber. Addressing catastrophic forgetting in few-shot problems. In *International Conference on Machine Learning*, pages 11909–11919. PMLR, 2021.

[18] Hong Wang, Yanjun Huang, Amir Khajepour, Dongpu Cao, and Chen Lv. Ethical decision-making platform in autonomous vehicles with lexicographic optimization based model predictive controller. *IEEE transactions on vehicular technology*, 69(8):8164–8175, 2020.

[19] François Chollet. On the measure of intelligence. *arXiv preprint arXiv:1911.01547*, 2019.

[20] Cristina Mele, Jacqueline Pels, and Francesco Polese. A brief review of systems theories and their managerial applications. *Service science*, 2(1-2):126–135, 2010.

[21] A Wayne Wymore. *Model-based systems engineering*, volume 3. CRC press, 2018.

[22] A Wayne Wymore. *A mathematical theory of systems engineering: the elements*. Wiley, 1967.

[23] Fremont E Kast and James E Rosenzweig. General systems theory: Applications for organization and management. *Academy of management journal*, 15(4):447–465, 1972.

[24] Norman S Nise. *Control systems engineering*. John Wiley & Sons, 2020.

[25] Tyler Cody, Stephen Adams, and Peter A Beling. A systems theoretic perspective on transfer learning. In *2019 IEEE International Systems Conference (SysCon)*, pages 1–7. IEEE, 2019.

[26] Alejandro Salado. A systems-theoretic articulation of stakeholder needs and system requirements. *Systems Engineering*, 24(2):83–99, 2021.

[27] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. arxiv e-prints, page. *arXiv preprint arXiv:1312.6199*, 2013.

[28] W Ross Ashby. *An introduction to cybernetics*. Chapman & Hall Ltd, 1961.

[29] W Ross Ashby. Principles of the self-organizing system. In *Facets of systems science*, pages 521–536. Springer, 1991.

[30] W Ross Ashby and Jeffrey Goldstein. Variety, constraint, and the law of requisite variety. *Emergence: Complexity and Organization*, 13(1/2):190, 2011.

[31] Carlos Gershenson. Requisite variety, autopoiesis, and self-organization. *Kybernetes*, 2015.

[32] Jie M Zhang, Mark Harman, Lei Ma, and Yang Liu. Machine learning testing: Survey, landscapes and horizons. *IEEE Transactions on Software Engineering*, 2020.

[33] Niloofar Shadab and Alejandro Salado. Towards an interface description template for reusing ai-enabled systems. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2893–2900. IEEE, 2020.

Appendix B

A Systems-Theoretical Formalization of Closed Systems

Niloofar Shadab, Tyler Cody, Alejandro Salado, Peter Beling

In Revision 1, The IEEE Open Journal of Systems Engineering

Abstract— There is a lack of formalism for some key foundational concepts in systems engineering. One of the most recently acknowledged deficits is the inadequacy of systems engineering practices for engineering intelligent systems. In our previous works, we proposed that closed systems precepts could be used to accomplish a required paradigm shift for the systems engineering of intelligent systems. However, to enable such a shift, formal foundations for closed systems precepts that expand the theory of systems engineering are needed. The concept of closure is a critical concept in the formalism underlying closed systems precepts. In this paper, we provide formal, systems- and information-theoretic definitions of closure to identify and distinguish different types of closed systems. Then, we assert a mathematical framework to evaluate the subjective formation of the boundaries and constraints of such systems. Finally, we argue that engineering an intelligent system can benefit from appropriate closed and open systems paradigms on multiple levels of abstraction of the system. In the main, this framework will provide the necessary fundamentals to aid in systems engineering of intelligent systems.

B.1 Introduction

There has long been a call for a theory of Systems Engineering (SE) within the SE community with the aim of establishing SE as a standalone engineering field capable of addressing modern engineering problems [1, 2]. However, there is an existing gap in concrete formalism and distinction for some fundamental concepts within the field that has led to ambiguity in some SE practices [3]. While such formalism might not have been necessary in the past, the emergence of new kinds of complex systems such as Artificial Intelligence (AI)-enabled systems has challenged traditional SE practices [4, 5, 6, 7, 8, 9].

In our previous work, we identified potential gaps in the current SE foundations to address the unique nature of AI-enabled systems [10]. We argued that intelligence is a relational property that can be characterized and engineered as a relation between the system and its context with both learning and intelligence properties embodied in the context regardless of the nature of the relations between them [10]. In this situation, intelligence is no longer relegated to a component or the physical boundary of the system. Therefore, we posited that owing to this high coupling between AI-enabled systems and their environment, utilizing the concept of closure in SE is a potential path forward to build general engineered intelligence [10]. We proposed that closed SE practices could be employed to realize the closed notion of this relational property between intelligent systems and their context. We further examined the possibility of employing closed systems precepts in an engineering framework in our later paper [11], concluding a lack of concrete definitions and formalism in the theory and practice of SE presents a barrier to applying closed system precept in engineering applications. Currently, most of the theoretical foundations in both systems theory and the theory of SE are bounded to the open systems precepts (i.e., inputs-outputs relations) [12]. Although the concept of closed systems is being utilized in limited applications in SE, there is little to no theoretical basis for these practices, making SE activities based on closed system precepts prone to interpretation and over-abstraction. As we have identified at least one domain that can benefit from closed systems precepts (AI-enabled systems), the need for clear definitions and formalism becomes increasingly important in the field of SE. This paper revisits the concept of closure in SE, aiming to formalize, define, and evaluate this concept as the first step towards employing closed systems precepts for intelligent systems.

As mentioned earlier, closure has been vaguely applied in SE with limited underlying formal foundations [13, 14, 15]. Various types of closure have been introduced in systems theory literature, including functional closure, organizational closure, operational closure, and informational closure, among others [16, 17, 18, 19, 20, 21]. As a starting point, closure can be understood as a property of a system that makes the system closed, and a closed system is defined as one that does not exchange energy, information, or matter through its boundaries [22]. (This concept will be revisited in detail later in the paper.) However, there is little to no formal framework to describe the relationships and differences between each type of closure, and many of the closure types lack formal systems-theoretic definitions that distinguish them from the other types of closure; in fact, on many occasions, these terms are used interchangeably, which can cause confusion in the application of each type of closure [14, 15]. In this paper, we develop formal systems-theoretic definitions for two types of closure, functional and informational closure, in systems and compare them in terms of system's characteristics and the relations between systems and their environment. We utilize a mathematical definition of functional dependency, information theory, and the systems-theoretic foundations for open and closed systems to produce formalism for functional and informational closure. Then, we determine the conditions and constraints to meaningfully use each of the two types of closure in systems. Our aim is to elaborate on the relations between these types of closure to determine their applications at different levels of abstractions for systems. Throughout this paper, we will use the terms *closure* and *closedness* interchangeably.

B.2 Background in Information Theory

Before delving into the formalism, we will provide a brief introduction to the concepts and mathematics of information theory.

Information theory establishes a relation between information and uncertainty, where information is inversely proportional to uncertainty. It enables the receiver of information to make more accurate predictions than chance [23, 24]. Similar to uncertainty, information is subjective and depends on the observer's prior knowledge. Information theory uses this relationship and explains how information can resolve uncertainty about an event [25], making it one application of probability theory [26]. Information of a system depends on the observer's degree of freedom to measure the number of unknown states of the system. Entropy is a crucial concept in information theory which measures the uncertainties of a random variable [24]. In information theory, the entropy of a random variable calculates the average level of information regarding the possible outcomes of that variable [26]. Entropy connects information with probability and uncertainty of a random variable. We can interpret this connection as follows: *the lower the entropy of a system, the more information we possess about the possible future states of the system.* Entropy can be calculated as follows:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{B.1}$$

Where X is a set of random variables, and $p(x)$ is the probability of occurrence of element x in set X. Equation B.1 represents information entropy, which was introduced by Shannon to compute the information transmitted from a source to a receiver through an information channel that the receiver can identify [27]. As shown in Equation B.1, information is proportional to the logarithm of the number of unknown states of a system. The concept of information entropy is analogous to Boltzmann's definition of information, which captures the number of ways to rearrange a system [26]. Under Boltzmann's definition of information, entropy is a measure of a random process of a final arrangement of molecules and it connects information to the uncertainty of the system's structure.

Shannon's definition of information characterizes it as the communication of relationships and relates it to probability. Entropy is considered a relational property that has a subjective component; it depends on the prior and conditional information that an observer (receiver) has about a system.

Building up from Equation B.1, one can define conditional and joint entropy of two discrete variables, which can be derived as follows (a similar process can be applied to continuous random variables) [28]:

$$H(Y,X) = -\sum_{x\in X, y\in Y} p(x,y)\log p(x,y) \tag{B.2}$$

$$H(Y|X) = -\sum_{x\in X, y\in Y} p(x,y)\log p(x|y) \tag{B.3}$$

Joint entropy indicates the amount of information needed to determine the value of two discrete variables. Conditional information depicts the amount of additional information needed to determine the value of a random variable given the knowledge of the value of the other random variable. If we expand Equations B.2 and B.3 using simple algebra and the chain rule, we can achieve the following relation between joint entropy and the conditional entropy:

$$
\begin{aligned}
H(Y|X) &= H(Y,X) - H(X) \\
H(X|Y) &= H(Y,X) - H(Y)
\end{aligned}
\tag{B.4}
$$

In probability theory, we have $p(x, y) = p(y, x)$. As a result, we can deduce from Equation B.2 that the joint entropy of X and Y is the same as the joint entropy of Y and X. Using $H(Y, X) = H(X, Y)$, and simple algebra from Equation B.4, we can derive the following relation:

$$
H(Y|X) = H(X|Y) + H(Y) - H(X) \tag{B.5}
$$

B.2.1 Mutual Information

The concept of *Mutual Information* is also relevant for this paper. Mutual information captures the dependency between two random variables X and Y. It represents the amount of uncertainty that is common to both X and Y. Therefore, by observing one random variable, the uncertainty that is mutual with the other random variable will be resolved. In other words, mutual information determines the amount of information one can get from random variable X by observing the other random variable Y. This information is jointly distributed according to the joint probability of X and Y. The formula for mutual information is given as follows [28]:

$$
I(X;Y) = \sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{B.6}
$$

Based on the definition, mutual information is always non-negative[1] [29]. Using Equation

[1] This property can be proved mathematically using Jensen's inequality and relative entropy. For more information please check [29]

B.3 and simple algebra, Equation B.6 can be written in terms of entropy as follows:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned} \tag{B.7}$$

Conditional mutual information can be also defined when we have three random variables and have joint distribution $p(x, y, z)$. It is a measurement of how much uncertainty is shared between X, and Y but not with Z. It can be defined as follows [30]:

$$I(X;Y|Z) = -\sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} \tag{B.8}$$

Equation B.8 also can be written as follows:

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|YZ) \\ &= H(XZ) + H(YZ) - H(XYZ) - H(Z) \end{aligned} \tag{B.9}$$

B.3 Background on Closed vs Open Systems In Systems Theory

In this section, we provide a brief introduction to the foundations of closed and open systems in systems theory to establish the background required for the formalism of closed systems precepts.

In systems theory, open and closed systems precepts are foundational precepts. Early general systems theorists and biological systems theorists defined open systems as those that have external interactions, with a boundary between the internal and the external, allowing interactions across the boundary [31]. This definition captures richness in the exchange of matter, energy, and information between a system and its environment [32]. A system with no external interactions is referred to as a closed system. Using a modeling framework that studies the structure, behavior, and properties of the systems in terms of relationships [22], a closed system description can become as a special form of an open system: one whose input and output sizes are assumed to equal zero [33].

In biology and natural sciences, an open system is a system whose border is permeable by matter and energy, while a closed system is only permeable by energy [34]. In control theory, closed systems are open systems where the input is composed of feedback to adjust the output [35]. Generally, *closedness* is primarily used to describe the nature of open systems' boundaries, as in biology and natural sciences, or the use of feedback to adjust interactions, as in control theory. In systems theory, as discussed, it is reflected by the absence of any input-output relations in systems. The definition of a closed system is limited to the abstract notions of the absence of inputs and outputs. However, some attempts have been made to describe open and closed systems using a set-representation of main relations on the components, behaviors, functions, inputs and outputs of the system [36]. An open system in this context can be described as a 7-tuple $S = (C, B, R^c, R^b, R^f, R^o, R^i)$, where C is a finite set of components, B is a finite set of behaviors, R^c is a finite set of component relations, R^b is a finite set of behavioral relations, and R^f is a finite set of functional relations, where a functional relation is defined between components and behaviors within the system. R^o and R^i are sets of finite output and input relations between external systems and the system of interests, respectively. Consequently, a closed system is a special case of an open system that does not include R^o, R^i. Therefore, it is represented as 5-tuple $S = (C, B, R^c, R^b, R^f)$ [36]. In this paper, however, we aim to be consistent with a systems-theoretical framework for defining systems which is based on the relations on system's inputs-outputs.

In summary, the general systems-theoretical definition of a closed system in systems theory literature refers to a system that has no inputs or outputs. This definition can be formally shown in Definition B.1:

Definition B.1 (**Systems-theoretical Closed System**)**.** A system that has no input set, $\mathcal{X}$, and output set, $\mathcal{Y}$, from/to its environment.

$$\mathcal{X} = \mathcal{Y} = \emptyset$$

From an objective perspective, it is arguable that only the entire universe might satisfy the condition in Definition B.1. In fact, closed systems were deemed by early general systems theorists as nonexistent [31]. However, closedness can still be used as a relaxation to support engineering work [37]. Specifically, a modeler could choose to ignore the existence of inputs and outputs, thereby assuming a closed system. This approximation has proven to be useful in several engineering applications, such as thermodynamics. In this paper, we posit that functional closure and informational closure are two potential paths to enable such relaxation of Definition B.1 by formulating a direction on how to ignore inputs and outputs set to model a closed system. These paths provide a framework to define closed systems in specific contexts, enabling modelers and systems engineers to make simplifying assumptions and develop accurate models.

With this background on systems theory, we will provide systems-theoretic definitions of the terminologies that we use for our formalism in Section B.4.

B.4 Elaboration of Terminologies:

Noting that system boundaries are subjective and fluid, with no restriction as to the relationships between the elements that may fit within the boundaries [38], we begin by defining a system of interest denoted by S^0. S^0 is a system that will be engineered for a specific purpose. With respect to an S^0, we define the following systems:

- Environment, denoted by E: It is a non-empty system that consists of everything outside of S^0.
- Context system, denoted by S^C: It is a system that consists of both S^0 and a non-empty part of E, which we call Inner Environment and denote by E^I. So, $S^C = S^0 \cup E^I$.
- Inner Environment, denoted by E^I: As per the previous definition, it is a non-empty system that consists of the complement of S^0 with respect to S^C.
- Outer environment, denoted by E^O: It is a system that consists of the complement of S^C with respect to E.
- Universe, denoted by U: It is a non-empty system that consists of the entire environment E, and the system of interest S^0.

The terminology used throughout this paper is depicted in Figure B.1, where the boundary of S^C is represented by a blue circle and the curved orange arrows depict the coupling between S^0 and E^I. In our definition, we partition the environment into two sets, $E = E^O \cup E^I$ wher $E^O \cap E^I = \emptyset$.

To maintain consistency with systems theory, we will use set-theoretic formalism to define system introduced in this paper. A set representation of the systems is based on the relation of the set of systems objects given the objects indices. Mathematically, we can represent

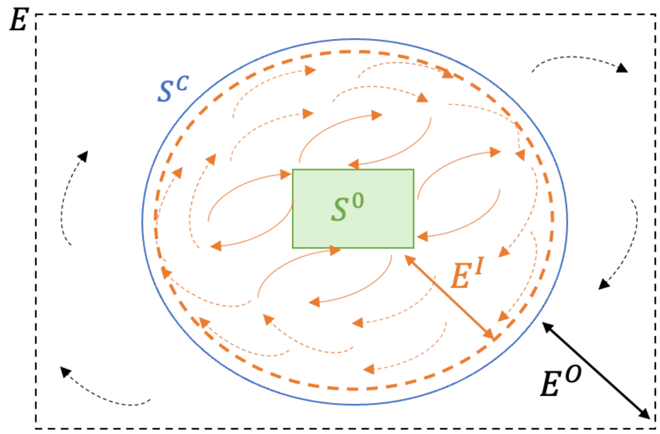


Figure B.1: Interactions and relations between systems in a closed system setting are depicted. E^O is the environment outside of the Context system. The context system, S^C, includes the system of interest, S^0, and a portion of the overall environment, the inner environment, E^I

this set as follows: $S \subseteq \times\{V_i | i \in I\}$ where V_i denotes the i^{th} element of the set. In this context, let S be an input-output system $S \subseteq \times\{\mathcal{X}, \mathcal{Y}\}$ and $\bar{S}$ the component sets of set S, i.e., $\{\mathcal{X}, \mathcal{Y}\}$. Consequently, we can define system of interest, Context system, and Universe, S^0, S^C, U, as follows:

$$
\begin{aligned}
S^0 &: S^0 \subseteq \times\{\mathcal{X}^0, \mathcal{Y}^0\} \\
U &: \mathcal{X}^u = \mathcal{Y}^u = \emptyset \\
S^C &: S^C \subseteq \times\{\mathcal{X}, \mathcal{Y}\}
\end{aligned}
\tag{B.10}
$$

Having defined S^0, S^C, U, we can formulate the environment with respect to these three defined systems. As it is shown in Figure B.1, the entire environment consists of everything outside S^0, therefore, we can define E as follows:

$$
\begin{gathered}
E : E \subseteq \times\{\mathcal{X}^E, \mathcal{Y}^E\} \\
\text{Where, } \bar{E} = \bar{U} \setminus \bar{S}^0
\end{gathered} \tag{B.11}
$$

Now that we formulated the entire environment, E, in Equation B.11, Inner environment, E^I, and Outer environment, E^O, will be presented as a subset of E:

$$
\begin{gathered}
E^I : E^I \subseteq \times\{\bar{E} \cap \bar{S}^C\} \\
E^0 : E^0 \subseteq \times\{\bar{E} \setminus \bar{S}^C\}
\end{gathered} \tag{B.12}
$$

Having defined different systems' boundaries, we also provide mathematical definitions for the system's terminologies that we will be using in the formalism process of closed systems. These terminologies are defined in a systems-theoretical framework. The list below contains these terminologies and their mathematical representation/definitions:

- **Functional System**: S is functional system if S executes a function, i.e, with every input $x \in \mathcal{X}$, there is associated a single output: $S : \mathcal{X} \rightarrow \mathcal{Y}$

  In line with abstract systems theory [39], we assume all functional systems to be surjective, unless otherwise specified. Also, note that, by definition, every open system will be a functional system.

- **Functional Dependency**: Functional dependency is defined based on the relationship between a dependent set Y and an independent set X through a function $f(X)$ [40]. We can express this relationship as $X \rightarrow Y$, meaning that one set (X) can accurately determine the value of the other set (Y). In a surjective function, there is always one dependent set or variable and one independent set or variable, making functional dependency a sufficient condition for a function to be surjective. Therefore, functional dependency can be also defined as:

$$\forall y \in Y,\ \exists x \in X \quad \text{s.t.} \quad S(x) = y \tag{B.13}$$

- **Set of System's Functions**: F is a set of system's function if, F is defined as $F \subseteq \times\{F_i | i \in I\}$ such that $F : \mathcal{X} \to \mathcal{Y}$, i.e., F is a mapping between system's inputs and outputs [41].

- **System's Behavior**: System's behavior can be described in reference to the system's response to any stimulus from its environment, $x \in \mathcal{X}$ [42]. For any stimulus $x \in \mathcal{X}$, and its associated decision problem; $\Delta(x)$, there is a solution $d(x) \in D$, where D is the decision set. Therefore, system's behavior is defined as a mapping of system's decision set D to the system's outputs $\mathcal{Y}$. This mapping can be shown as: $Q : D \to \mathcal{Y}$.

- **System's States**: State of the system represents the history of the system's behavior. The state at any point in time can be represented as an equivalence class generated by the equivalence relation defined on the system's past behaviors [41].

To investigate how and under what conditions the context system may be conceived or relaxed as a closed system, we explore two potential paths: functional closure and informational closure. We examine the requirements and implications of each approach and discuss their applicability to real-world SE problems. Ultimately, our goal is to provide insights and guidance for system engineers seeking to model closed systems and understand the implications of such models.

B.5 Functional Closure

Given the terminologies introduced and formalized in Section B.4, this paper now explores the concept of functional closure. We provide a formal definition for a functionally closed system. With the functional closure concept, instead of having a systems-theoretical closed systems, we have a relaxation of a closed system that can be considered closed only from the functional point of view.

Based on the systems-theoretic definition of functional system, and the condition of functional dependency, a system can be considered to be functionally closed with respect to its environment when the system's outputs are independent of the inputs received from the environment and the behavior of the environment is independent of the inputs it receives from the system. Mathematically, this can be represented as $\times\{\mathcal{X}, \mathcal{Y}\} = \emptyset$, indicating that there is no mapping between the system's inputs and outputs. This condition is consistent with the definition of a systems-theoretic closed system, which also requires $\mathcal{X} = \mathcal{Y} = \emptyset$, indicating no mapping between inputs and outputs. Therefore, it can be concluded that a systems-theoretic closed system, which has an empty set of inputs and outputs, is also a functionally closed system, as it implies an empty set of mappings between inputs and outputs, indicating functional independence of the system from its environment.

However, this definition may not have any practical or formal systems-theoretical implications. In systems theory, systems are defined as relations on their inputs and outputs. Therefore, the absence of any mappings between inputs and outputs in functionally closed systems prevents us from defining such systems in a systems-theoretical formulation, as expressing relations of an empty set has no meaning. To overcome this complication, we can define a condition where a system is conceived as functionally closed if its set of functions can be defined by a minimal set of inputs and outputs. By minimal set, we suggest that any changes in the inputs set will not change the function of the system, i.e., the system's outputs set. This allows us to define a functionally closed system in a practical and meaningful way, as it captures the concept of functional independence from inputs while avoiding the issues associated with expressing relations on an empty set. We formally demonstrate functional closure using the minimal set constraint as follows:

$$
\begin{aligned}
&\text{Given a functional system, } S : X \to Y \\
&\text{where} \quad \forall y \in Y,\ \exists x \in X, \quad \text{s.t.} \quad S(x) = y \\
&\text{Given a functional system, } S' : X' \to Y \text{ where } X \subset X' \\
&\text{If} \quad \forall y \in Y,\ \nexists x' \in X', \quad \text{s.t.} \quad S'(x') = y \ \wedge \ x' \notin X \\
&\Rightarrow S' \text{ is functionally the same as S} \\
&\Rightarrow S \text{ is functionally independent from additional inputs } x' \\
&\Rightarrow S \text{ is functionally closed from } S'
\end{aligned}
\tag{B.14}
$$

Based on Functional Closure formulation in Equation B.14, we deduce that the outputs set of system S is not dependant on any additional inputs beyond the minimal inputs set from System S'. Therefore, we can outline the conclusion from Equation B.14 as follows:

S is said to be functionally closed from S' if and only if:

- There exists a minimal set M of inputs and outputs, where $M \subseteq \times\{\mathcal{X}, \mathcal{Y}\}$, such that S is functionally dependent only on M, i.e., $S \subseteq \times\{\mathcal{X}_M, \mathcal{Y}_M\}$

- There are no additional inputs beyond M that can influence the behavior of the system, S.

- The state of S' is not affected by the outputs of S.

When these conditions are present, functional independence of the system from inputs set and outputs set can be relaxed as $\mathcal{X} = \mathcal{Y} = \emptyset$. This relaxation is due to the independence of the system's functions from changes in inputs set, $\mathcal{X}$, and hence, existence of a constant set of outputs, $\mathcal{Y}$. A functionally closed system can be represented via the set of system's functions (which already captures the minimal set of input-output relation). This representation allows us to ignore $\mathcal{X}$ and $\mathcal{Y}$ in the representation of such system. Therefore, a functionally closed system can be a relaxation of the systems-theoretic closed system (Definition B.1). To prove that functional closure is a relaxation of the systems-theoretic closure, we will have:

$$\begin{aligned}
&F \subseteq \times\{F_i | i \in I\} \quad \text{s.t} \quad F : \mathcal{X}_m \to \mathcal{Y}_m \\
&\text{Where:} \quad \mathcal{X}_m \subseteq M \quad \mathcal{Y}_m \subseteq M \\
&\text{If:} \quad \exists X_E, \quad \exists Y_E, \quad \text{s.t} \quad \mathcal{X}_m \subseteq X_E, \quad \mathcal{Y}_m \subseteq Y_E : \\
&\ \text{We have: } F : X_E \to \mathcal{Y}_m \\
&\text{If: S is defined as } S \subseteq \times\{F_i | i \in I\} \\
&\text{Let: } \mathcal{X} = X_E \setminus M \text{ and } \mathcal{Y} = Y_E \setminus M \\
&\text{We have: } S \not\subseteq \times\{\mathcal{X}, \mathcal{Y}\}
\end{aligned} \tag{B.15}$$

Where F is the set of system's functions. When the conditions stated above are present, functional independence of the system from inputs set and outputs set can be relaxed as $\mathcal{X} = \mathcal{Y} = \emptyset$, which is the same condition for a systems-theoretic closed system. Therefore, a functionally closed system can be considered a relaxation of the systems-theoretic closed system because it is a less restrictive condition that still allows for some interaction with the environment.

Extending Equation B.14 into the paper's terminologies, we posit that by having functional independence between the context system and its environment, the context system achieves functional closure. It means that (1) there is no inputs set from the outer environment, E^O, that influences the behaviors/functions of the context system, S^C, and therefore any of its internal functions, and (2) the state of the outer environment, E^O, is not influenced by the outputs of the context system, S^C. Given the interpretation of functional closure for a system in Equation B.14, a functionally closed context system; S^C; can be formally defined as follows:

Definition B.2 (Functionally Closed Context System). A functional context system, S^C, is functionally closed from its outer environment, E^O, if and only if,
1) There exists a minimal set of inputs and outputs, , M, such that S^C is functionally dependent on M. This condition can be shown as: $S^C \subseteq \times\{\mathcal{X}_M, \mathcal{Y}_M\}$, and
2) There are no additional inputs from E^O beyond M that can influence the behavior of S^C.
and
3) There are no additional outputs from S^C beyond M that can affect the behavior of S^C.
Mathematically, the second and third conditions can be shown as follows:

$$\text{Given:} \quad S^C : \mathcal{X}_M \to \mathcal{Y}_M \quad \& \quad E^O : \mathcal{X}^{\mathcal{O}} \to \mathcal{Y}^{\mathcal{O}}$$

$$\text{Where:} \quad y \in \mathcal{Y}_M \quad \& \;\; x \in \mathcal{X}_M$$

$$\text{From Eq B.10, and Eq B.12, we know: } \mathcal{Y}^{\mathcal{O}} \to \mathcal{Y}_M \wedge \mathcal{X}_M \subseteq \mathcal{Y}^{\mathcal{O}}$$

$$\text{If:} \quad x' : x' \in \mathcal{Y}^{\mathcal{O}} \wedge x' \notin \mathcal{X}_M$$

$$\forall y \in \mathcal{Y}_M, \quad \nexists x' \in \mathcal{Y}^{\mathcal{O}}, \quad \text{s.t.} \quad S^C(x') = y$$

This definition of a functionally closed system in Definition B.2 shows that it is a relaxation of the systems-theoretic definition of a closed system. It allows for the context system to have interactions with the environment, but these interactions must not affect the behavior of the context system or its outputs.

It should be noted that functional closure derived from Definition B.2 is fundamentally different from a closed control system where inputs and outputs are coupled together through feedback loops. The thorough comparison between these two types of systems is discussed in our previous work [10].

B.5.1 Functional Closure Constraint For Engineering Purposes

Figure B.2 shows the translation of system of interest's relations with its environment through functional closure. In the top diagram, we have $Item3 = f(Item2, Item5)$. In the bottom diagram, we have $Item3 = f(Item2)$. Assume that system of interest's behavior is independent of the outer environment (or that its effect is negligible). Then, we can state: $Item3 = f(Item2, Item5) = f(Item2)$. Functional closure allows us to remove the arrow $Item5$. The same story can be constructed from the perspective of $Item6$, assuming that the SOI causes negligible impact on the outer environment.

From the perspective of system of interest, the context is functionally closed if also for the inner environment, we can ignore the impact of $Item1$ in $Item2 = f(Item3, Item1)$. Therefore, we have: $Item2 = f(Item3, Item1) = f(Item3)$. The same story would be constructed for $Item4$.

Therefore, to have functional closure between S^C and E^O, we need to attest that there is

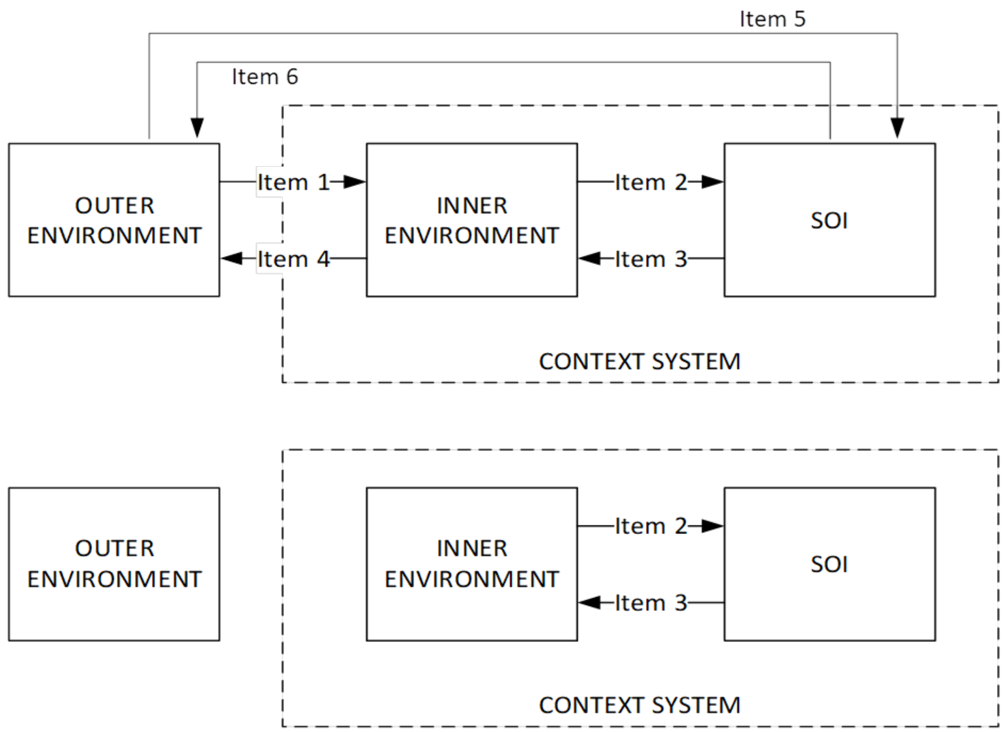


Figure B.2: Top Diagram captures all inputs and outputs between S^0, E^I and E^O. Bottom diagram shows functional closure framing of the top diagram.

functional independence between these two systems. The absence of functional dependencies between S^C and E^O implies that both S^0, E^I have functional independence from E^O. Since S^0, and E^I are functionally dependent on each other (within closed system's boundary), if E^I is independent from E^O, S^0 will also be independent from E^O and vice versa.

In a scenario where E^I and E^O can be relaxed as functionally independent, S^C is functionally closed from E^O. Relaxation of functional closure may suffice in some engineering purposes. Some SE practices such as mission engineering are based on the assumption of functional independence of E^I and E^O and functional closure of S^C. However, an absence of mapping between inputs and outputs for the two partitions of the environment is not a realistic condition in many of the engineering applications.

We demonstrated that for S^C to be engineered as a functionally closed system, it either requires to encompass all of E, which is the universal system or to deal with an environment that can be divided into two functionally independent parts which is not a valid assumption for the majority of the complex environments. In functionally closed systems, we ought to create a system that is built upon its minimal set of inputs and outputs and will not change or functionally evolve over time. Therefore, a functionally closed system can be suitable for some categories of SE practices and cumbersome for the rest. The modeling benefits of this relaxation would be realized if the system is characterized at an accurate level of abstraction. We, next, examine informational closure as another path to have a relaxation of closure for engineering systems.

B.6 Information Closure

In this section, similar to Section B.5, we will elicit formalism of informational closure definitions. Then we will examine if compared to functional closure, informational closure is a better solution for a relaxation of the systems-theoretical definition of a closed system.

Informationally closed system can be achieved when there is no *new* information exchanged between the context system S^C and its environment E^O [19]. This definition is different from functional closure where any type of information (new or expected) cannot cross the closed system's boundary. The interpretation of this definition is that mutual information between the closed system at state n, S^C_n, and its environment, E^O_n, needs to be enough to ignore any new information transmission between these two systems. This closure implies that the joint information between the closed system at state $n+1$, S^C_{n+1}, and the environment, E^O_n, given the information from the closed system at state n, S^C_n, should also be zero [19]. It indicates that the future state of an informationally closed system should not be dependent on the conditional information of its environment given the information of the present state of the system. Consequently, the current environment does not contain any information regarding the future states of the system that has not already been present in the current state of the system; this notion is an indication of informational closure [20].

According to Definition B.2, a functionally closed system is not affected by any changes in its inputs set. In contrast, in an informationally closed system, there could be changes in the inputs and outputs sets to and from the closed system, S^C, that can be interpreted in the form of mutual information between the closed system; S^C; and its environment; E^O. An informationally closed system considers a boundary or scope that distinguishes the external environment, E^O, from the internal system, S^C, however, in contrast to functionally closed systems, it does not necessitate that the mapping between inputs and outputs are bounded to a minimal set, but rather that the changes in the inputs set; $\mathcal{X}$; will influence the set of outputs $\mathcal{Y}$ by changing/updating the content of mutual information between the current states of the closed system and outer environment. In functionally closed systems, on the other hand, we don't consider any influence on input-output sets crossing between the closed system and environment beyond their minimal sets. To translate this property of informationally closed systems, the systems-theoretic framework that is based on the mapping relations between $\mathcal{X}$ and $\mathcal{Y}$ should be extended such that it can take into account the dynamic nature of information transmission at each state in the closed system. In contrast to functionally closed system that closure is a static property of the system, informationally closed systems are defined at system's states. Closure is a transitional property that is bounded by the previous states of the system and outer environment. Therefore, to define this type of closed system, we utilize information theory.

As described earlier in this paper, we consider a system *informationally closed* when there is no flow of new information between the environment and the system. For informationally closed systems, from the definition that we provided earlier[2] , we have $I(S^C_{n+1}; E^O_n|S^C_n) \to 0$ [20]. We can frame this definition as follows:

Definition B.3 (**Interpretation of An Informationally Closed Systems Using Information**)**.** A Context System that transitions through states $1, 2, ..., n, n+1$; is informationally closed at state n if there is no joint information between S^C_{n+1} and $E^O_n|S^C_n$.

$$I(S^C_{n+1}; E^O_n|S^C_n) = 0$$

Proposition 2. If S^C is informationally closed, joint information of S^C_{n+1}, E^O_n, S^C_n, equals to joint information between S^C at state n and state $n+1$:

$$I(S^C_{n+1}; E^O_n, S^C_n) = I(S^C_{n+1}; S^C_n)$$

Proof.

$$I(X;Y) = H(X) - H(X|Y) \implies$$

$$I(S^C_{n+1}; E^O_n, S^C_n) = H(S^C_{n+1}) - H(S^C_{n+1}|S^C_n, E^O_n)$$

Thus; from Equation B.7 and Equation B.9, and substituting X, Y, Z with S^C_{n+1}, E^O_n, S^C_n, we have:

$$I(S^C_{n+1}; E^O_n, S^C_n) =$$

$$I(S^C_{n+1};, S^C_n) + I(S^C_{n+1}; E^O_n|S^C_n) \to$$

[2] Informational closure can be achieved when the flow of new information between the closed system, S^C_n, and its environment, E^O_n sets to zero

$$I(S_{n+1}^C; E_n^O, S_n^C) = I(S_{n+1}^C;, S_n^C)$$

$\square$

To rigorously compare informational closure with functional closure, we will provide an informational interpretation of functional closure. Interpreting functional closure using the concept of joint information follows the set-theoretic formalism of having the mapping between the environment and the closed system, $\times\{\mathcal{X}, \mathcal{Y}\}$, as an empty set beyond their minimal set. The absence of any mappings between outputs set and changes in inputs set indicates that the closed system was built upon the minimal information from the environment. No information from the environment would change the output of the system. If the information from outer environment enters the system at state n and the output of the system is produced at the next state of the system, $n+1$, functional closure can be relaxed as if there would be no information transition between the next state of the system and the current state of outer environment. This absence of information transition from the environment to the closed systems reflects in the next state of the system such that there would be no mutual information between the next state of the system and the current state of the environment. This reflection suggests that all the information in the next state of the system will be provided by the current state of the system. Using this information-theoretic framing, functional closure for the context system; S^C; can be defined as:

Definition B.4 (Interpretation of A Functionally Closed System Using Information). A Context System that transitions through states $1, 2, ..., n, n+1$; is functionally closed if there is no mutual information between E_n^O and S_{n+1}^C:

$$I(S_{n+1}^C; E_n^O) = 0$$

Proposition 3. If S^C is functionally closed, all the information existing in S^C_{n+1} comes from S^C_n. Therefore, we have:

$$I(S^C_{n+1}; S^C_n) = H(S^C_{n+1})$$

Proof. From Equation B.7, the joint information of S^C_{n+1} and S^C_n in a functionally closed system would be:

$$I(S^C_{n+1}; S^C_n) = H(S^C_{n+1}) - H(S^C_{n+1}|S^C_n)$$

As the entire information of S^C_{n+1} comes from S^C_n, $H(S^C_{n+1}|S^C_n)$ will be zero.

$$H(S^C_{n+1}|S^C_n) = 0$$

Thus;

$$I(S^C_{n+1}; S^C_n) = H(S^C_{n+1})$$

□

In a functionally closed system, the inputs set from E^O at state n would not affect the outputs set of S^C at state $n + 1$. Having no effects of inputs change on the outputs set, we argue that there is also no joint information between the system at state n, S^C_n, and environment at state n, E^O_n. Definition B.4 could therefore be extended such that it includes the fact that there will be no input from the environment, E^O_n, with new information to the system at state n and no output from the system with new information to the environment at state n. This relation emphasizes on the static nature of functional closure.

Informational closure, on the other hand, adds a dynamic nature to functional closure where the system can be considered closed relative to its set of states and the outer environment set of states. Informational closure implies that an informationally closed system is not derived from the constraints on inputs set. It is rather derived from the ability to predict or expect such an inputs set from its environment, E_n^O. This implication again is related to the relationship between the concepts of information and uncertainty in systems which means the more the mutual information, the less the uncertainty related to inputs set. Intelligent systems in particular can be engineered utilizing such closure as these systems are expected to have the capability of prediction of inputs from the environment. In line with this implication, later in the paper, we will argue how informational closure can be employed as an engineering constraint to build intelligent systems.

In informational closure, there is causal dependency between the environment; E_n^O; and the closed system; S_n^C. Another major distinction between the informationally and functionally closed systems roots in the difference in assumptions about the causal (or functional) dependencies between the system and its environment. In Definition B.2, it is shown that there is no functional (causal) dependencies between the functionally closed system and its environment beyond its minimal set. In contrast, there are inputs and outputs to and from an informationally closed system which can be defined as mutual information. This input-output relation; mutual information; can be built upon various types of causal relationships between the two systems. Mutual information can have causality originated from either the closed system or its environment. The causal dependency analysis for this type of closure answers the question of "where the information originates from". However, in this paper, we investigate *"the very existence of mutual information transmitting through the closed system's boundary"*. Thus, although, understanding causal relations is necessary for engineering applications, the causal dependency of this mutual information is not the focus of this paper, and it could be considered as a potential future work.

B.6.1 Informational Closure constraint for Engineering Purposes

Figure B.3 shows the translation of system of interest's relations with its environment through informational closure. In the top diagram, we have the same systems as in Figure B.2. In the bottom diagram, we have an additional $Item7$ which is the mutual information between the informationally closed system and its environment E^O. This mutual information is a subset of all the inputs-outputs to/from S^C and E^O. The impact of the rest of the inputs-outputs can be ignored in S^C.

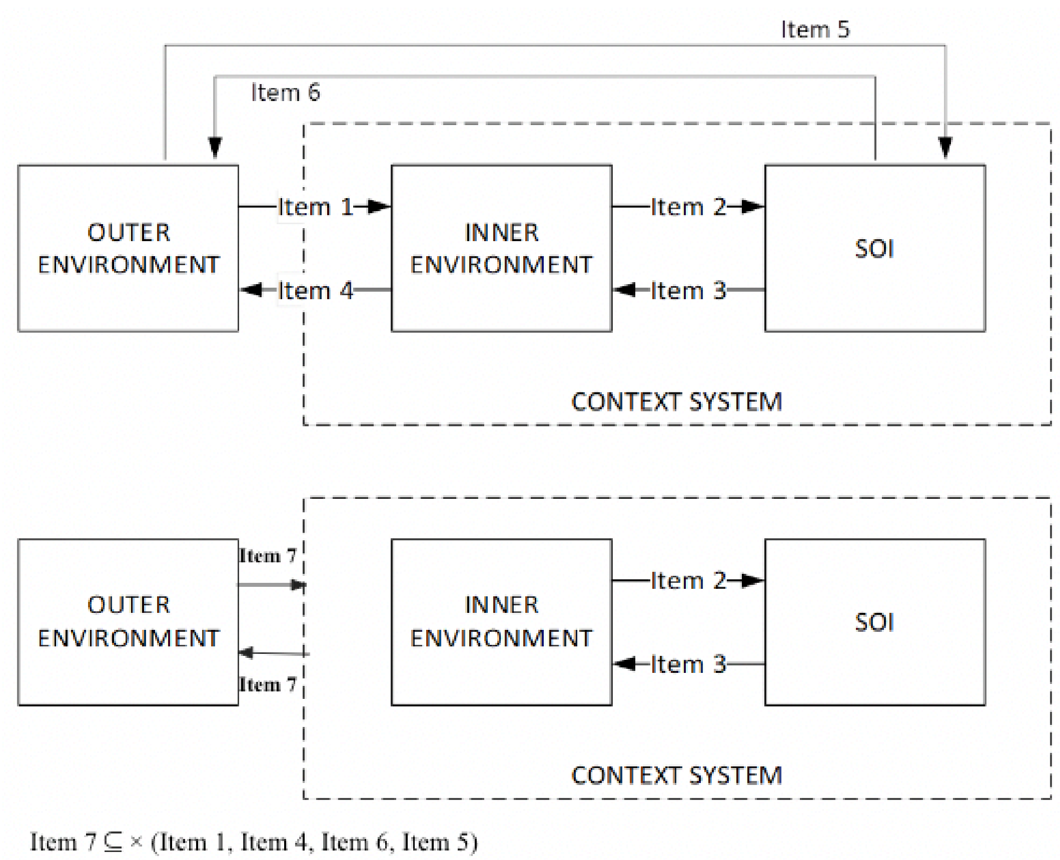


Figure B.3: Top Diagram captures all inputs and outputs between S^0, E^I and E^O. Bottom diagram shows informational closure framing of the top diagram.

So far, we provided a foundation of why mutual information between the informationally closed system and its environment should stay high enough to maintain the condition of informational closure. As a result, the next step is to define "high enough" level of mutual information to enable realization of such a closed system in engineering applications. For this purpose, we need to find the condition for having at least a minimum level of required mutual information $I(S_n^C; E_n^O)$ between the context system; S_n^C; and Outer Environment; E_n^O; to realize an informationally closed system at state $n+1$.

To derive such a constraint for mutual information in the context of informational closure, we will employ the entropy and information formulas while observing the bounded definition of informational closure. To do so, we assume that the informationally closed system is the sole sender and receiver of the information at different time steps as there will be no new information coming in and out of the closed system's boundaries. Accordingly, we consider the system at state n; S_n^C; as the sender of the information, and the system at state $n+1$; as the receiver of such information. From Equation B.7, we can write:

$$I(S_{n+1}^C; E_n^O, S_n^C) = H(S_{n+1}^C) - H(S_{n+1}^C|S_n^C, E_n^O) \tag{B.16}$$

If we use Equation B.8 and substitute the last entropy in Equation B.16, we can rewrite Equation B.16 as follows:

$$I(S_{n+1}^C; E_n^O, S_n^C) = I(S_{n+1}^C; S_n^C) + H(S_{n+1}^C|S_n^C) + I(S_{n+1}^C; E_n^O|S_n^C) - H(S_{n+1}^C|S_n^C) \tag{B.17}$$

We can now define the mutual information between these systems and their environment E_n^O using Equation B.17 and chain rule as follows [43]:

$$I(S^C_{n+1}; E^O_n, S^C_n) = I(S^C_{n+1}; S^C_n) + I(S^C_{n+1}; E^O_n|S^C_n) = I(S^C_{n+1}; E^O_n) + I(S^C_{n+1}; S^C_n|E^O_n) \quad \text{(B.18)}$$

Equation B.18 is derived using the chain rule, as well as conditional and mutual information formulas. For informationally closed systems, from Definition B.3, we have $I(S^C_{n+1}; E^O_n|S^C_n) = 0$. This condition depicts that for an informationally closed system, the amount of mutual information between S^C_{n+1} and $E^O_n|S^C_n$ should become zero. Therefore, the next state of the system only relies on a portion of the information from its environment that is shared with the system at its current state. Incorporating the definition of the closed system into Equation B.18, and utilizing the fact that information cannot be negative, we have the following inequality:

$$I(S^C_{n+1}; S^C_n) \geq I(S^C_{n+1}; S^C_n|E^O_n) \quad \text{(B.19)}$$

To further decompose this inequality, we change information into entropy using Equation B.7. We replace $I(S^C_{n+1}; S^C_n)$ with $H(S^C_{n+1}) + H(S^C_n) - H(S^C_{n+1}, S^C_n)$. We also replace $I(S^C_{n+1}; S^C_n|E^O_n)$ with $H(S^C_{n+1}) + H(S^C_n|E^O_n) - H(S^C_{n+1}, S^C_n|E^O_n)$ and incorporate them into Equation B.19. Therefore, we have:

$$H(S^C_{n+1}) + H(S^C_n) - H(S^C_{n+1}, S^C_n) \geq H(S^C_{n+1}) + H(S^C_n|E^O_n) - H(S^C_{n+1}, S^C_n|E^O_n) \quad \text{(B.20)}$$

Based on the chain rule, we can provide relationships between mutual and conditional entropy for S^C_n and E^O_n. In other words, we have $H(E^O_n) + H(S^C_n|E^O_n) = H(S^C_n) + H(E^O_n|S^C_n) =$

$H(S_n^C, E_n^O)$. The relation between mutual and conditional entropy is similar to that of probability. We can achieve mutual entropy of two systems using the summation of the entropy of one system and the entropy of the second system given the first system. Using simple algebra, if $H(S_n^C)$ is replaced with $H(E_n^O) + H(S_n^C|E_n^O) - H(E_n^O|S_n^C)$ and substitute it into Equation B.20, we have:

$$H(E_n^O) - H(E_n^O|S_n^C) \geq H(S_{n+1}^C, S_n^C) - H(S_{n+1}^C, S_n^C|E_n^O) \tag{B.21}$$

Equation B.21 is the entropy representation of Equation B.19. Now to simplify Equation B.21, we utilize Equation B.4 where we have $H(E_n^O|S_n^C) = H(S_n^C, E_n^O) - H(S_n^C)$. Replacing $H(E_n^O|S_n^C)$ with $H(S_n^C, E_n^O) - H(S_n^C)$, and substituting $H(S_n^C) + H(E_n^O) - H(S_n^C, E_n^O)$ with $I(S_n^C; E_n^O)$ in Equation B.21, we will get the following inequality for minimum mutual information between S_n^C and E_n^O.

Theorem B.5 (**Inequality for mutual information in closure**).

$$I(S_n^C; E_n^O) \geq H(S_{n+1}^C, S_n^C) - H(S_{n+1}^C, S_n^C|E_n^O)$$

Theorem B.5 provides the relation for the level of mutual information being presented in the boundary of an informationally closed system. To maintain closure at state $n + 1$, the output of the closed system to the environment at state n (i.e., the amount of information transmitted from the system to its environment) should follow Theorem B.5. This inequality shows that the capacity of the channel between environment E_n^O and S_n^C should be more than a specific minimum. This lower bound is also dependent on the entropy of the next state of the system; S_{n+1}^C. Theorem B.5 has an edge case scenario where mutual information of S_n^C and E_n^O becomes zero while being informationally closed. The zero mutual information in Theorem B.5 indicates the two entropy $H(S_{n+1}^C, S_n^C)$ and $H(S_{n+1}^C, S_n^C|E_n^O)$ are equal. This equality means that the system does not send or get any relevant information to or from the current state of the environment; thus the system becomes functionally closed.

The inequality in Theorem B.5 ensures that the information transmitted from the system at state n to the system at state $n+1$; $I(S_n^C; S_{n+1}^C)$; would be maximal to confirm the closure of information in the closed system boundary. The level of mutual information between the current state of the closed system and its environment will be determined by the difference between the mutual entropy of the system at state n and at the next state $n+1$ as well as the mutual information between the system at the two states $n+1$ and n given E_n^O.

This dependency on mutual entropy of S_{n+1}^C and S_n^C indicates that in order to enable informationally closed systems engineering, one needs to capture the space of unknown states for the system in the future given the information of its current states, and current environment.

Therefore, informational closure in SE is tightly interdependent with the ability to identify incomplete information and the states of the closed system in the future. In other words, by employing informational closure in SE, we don't need to have complete information of the system of interest's future states; ${S^0}_{n+1}$. There will be no need to provide a capability to predict 100% of the the system of interest's states in advance. Rather, it relies on identification of what will be unknown in the next state of the system $H(S_{n+1}^C)$ given the current level of information of S_n^C and $S_n^C|E_n^O$. The fact that complete predictability is not a required feature for SE of our system of interest makes this methodology suitable for engineering intelligent systems. Intelligent systems may encounter new and/or unpredicted conditions and they are expected to learn, survive, and meet their goals when encountering such conditions.

The closedness of information allows systems engineers to bound intelligent systems' environment and create a closed intelligence within the closed system's boundaries that emerges from the high coupling between the system of interest and part of its environment, E^I. Using informational closure, systems engineers decide what part of the environment needs to be coupled and modeled with the systems of interest [10].

Systems engineers require to recognize and interpret what part of information should be engineered as mutual information between the closed system and its environment. Thus far, we concluded that the level of mutual information between the system and its environment should be present at a certain minimum level for realization of a closed system. However, mutual information, due to its subjective nature, should be determined as an important design choice in different engineering applications. Therefore; to understand the influence of design constraints on the level of mutual information, its upper bound can be captured by specifying a cost function. This cost function can be derived by other properties of the system of interest such as safety, cost, etc. As a result, systems engineers need to do trade-off analysis between $H(S^C_{n+1}, S^C_n)$ and $H(S^C_{n+1}, S^C_n|E^O_n)$ to achieve the inequality in Theorem B.5. We can maximize $I(S^C_{n+1}, S^C_n)$ up to a point or maximize $H(S^C_{n+1}, S^C_n|E^O_n)$ up to a threshold δ where:

$$H(S^C_{n+1}, S^C_n) - H(S^C_{n+1}, S^C_n|E^O_n) \leq I(S^C_n; E^O_n) \leq \delta \tag{B.22}$$

Decreasing $H(S^C_{n+1}, S^C_n)$, systems engineers create a core-dominant system while decreasing $H(S^C_{n+1}, S^C_n|E^O_n)$, they create a periphery-dominant systems [11]. We argue that δ would be a design parameter that would depend on several factors. Mutual information between the closed system and its environment helps improve the level of predictability of environment inputs to the system. This factor is a critical measure to determine the upper bound of the mutual information between the closed system and its environment. increasing mutual information between the closed system and its environment has a direct relation with increasing the cost of design and engineering of the system of interest as it adds more complexity and needs more resources to provide this communication channel between the environment and the system. As a result, cost can be considered as constraint to the upper bound of mutual information.

B.7 Conclusion

In this paper, we asserted formalism for two types of closure; functional, and informational; and compared them with each other. We derived constraints to achieve each type of closure in systems. We showed that functional closure is a special form of informational closure. We also argued that intelligent systems can be designed and engineered at multiple levels of abstraction. Some levels are informationally closed; while others are informationally and functionally open, etc. We provided foundation on how the open and closed systems concept can be utilized to engineer different levels of abstractions for such systems. Finally, we elaborated on how we can exert constraints for the level of mutual information required between the next state of the system and the current states of the system and its environment to enable informational closure.

Bibliography

[1] George A Hazelrigg and Donald G Saari. Toward a theory of systems engineering. *Journal of Mechanical Design*, 144(1), 2022.

[2] JE Kasser and DK Hitchins. Unifying systems engineering: Seven principles for systems engineered solution systems, proceedings of the 20th international symposium of the incose. 2011.

[3] Alejandro Salado. A systems-theoretic articulation of stakeholder needs and system requirements. *Systems Engineering*, 24(2):83–99, 2021.

[4] David D Walden and Robert E McKillip. Intelligent systems: a systems engineering approach. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3):871–882, 1991.

[5] Russell C Eberhart and James S Albus. Systems engineering for artificial intelligence. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 32(4):435–440, 2002.

[6] Thomas A McDermott, Mark R Blackburn, and Peter A Beling. Artificial intelligence and future of systems engineering. In *Systems Engineering and Artificial Intelligence*, pages 47–59. Springer, 2021.

[7] James Llinas, Hesham Fouad, and Ranjeev Mittu. Systems engineering for artificial intelligence-based systems: A review in time. *Systems Engineering and Artificial Intelligence*, pages 93–113, 2021.

[8] Niloofar Shadab, Aditya U Kulkarni, and Alejandro Salado. Shifting paradigms in verification and validation of ai-enabled systems: A systems-theoretic perspective. In *Systems Engineering and Artificial Intelligence*, pages 363–378. Springer, 2021.

[9] Ernesto Exposito, Juan R Rabu nal, and Rafael S Barbado. Systems engineering for artificial intelligence and cognitive systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(6):827–834, 2016.

[10] Niloofar Shadab, Tyler Cody, Alejandro Salado, and Peter Beling. Closed systems paradigm for intelligent systems. In *2022 IEEE International Systems Conference (SysCon)*, pages 1–8. IEEE, 2022.

[11] Tyler Cody, Niloofar Shadab, Alejandro Salado, and Peter Beling. Core and periphery as closed-system precepts for engineering general intelligence. In *International Conference on Artificial General Intelligence*, pages 209–219. Springer, 2022.

[12] Defense Acquisition Guidebook DAG and Workplace Ethic. *Introduction to systems engineering*. Citeseer, 2000.

[13] David D Walden and John B Goodenough. Mission engineering: A framework for the systems engineering of complex systems. *Systems Engineering*, 7(1):1–18, 2004.

[14] John B Goodenough and Richard E Schatzki. The concept of closure in systems engineering. *Systems Engineering*, 5(2):89–106, 2002.

[15] John B Goodenough and David D Walden. A closure-based approach to systems engineering. *Systems Engineering*, 8(1):1–17, 2005.

[16] Milan Zeleny. What is autopoiesis. *Autopoiesis: A theory of living organization*, pages 4–17, 1981.

[17] Francisco G Varela, Humberto R Maturana, and Ricardo Uribe. Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4):187–196, 1974.

[18] John Bednarz Jr. Autopoiesis: The organizational closure of social systems. *Systems Research*, 5(1):57–64, 1988.

[19] Nils Bertschinger, Eckehard Olbrich, Nihat Ay, and Jürgen Jost. Information and closure in systems theory. In *Explorations in the Complexity of Possible Life. Proceedings of*

the 7th German Workshop of Artificial Life, pages 9–21. IOS Press Amsterdam, The Netherlands, 2006.

[20] Nils Bertschinger, Eckehard Olbrich, Nihat Ay, and Jürgen Jost. Autonomy: An information theoretic perspective. *Biosystems*, 91(2):331–345, 2008.

[21] Angel Mora, Pablo Cordero, Manuel Enciso, Inmaculada Fortes, and Gabriel Aguilera. Closure via functional dependence simplification. *International Journal of Computer Mathematics*, 89(4):510–526, 2012.

[22] Anatol Rapoport Bertalanffy, Kenneth E Boulding, William Ross Ashby, Margaret Mead, and Gregory Bateson. *L. von Bertalanffy, General System Theory.* New York: George Braziller, 1968.

[23] Robert K Logan. What is information?: Why is it relativistic and what is its relationship to materiality, meaning and organization. *Information*, 3(1):68–91, 2012.

[24] Christoph Adami. What is information? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2063):20150230, 2016.

[25] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[26] Hans Christian Von Baeyer. *Information: The new language of science.* Harvard University Press, 2004.

[27] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55, 2001.

[28] Thomas M Cover, Joy A Thomas, et al. Entropy, relative entropy and mutual information. *Elements of information theory*, 2(1):12–13, 1991.

[29] Erik G Learned-Miller. Entropy and mutual information. *Department of Computer Science, University of Massachusetts, Amherst*, page 4, 2013.

[30] Mai Vu. Entropy and mutual information. *Retrieved March*, 2, 2014.

[31] Ludwig Von Bertalanffy. General systems theory. *The Science of Synthesis: Exploring the Social Implications of General Systems Theory*, 103, 2010.

[32] Cristina Mele, Jacqueline Pels, and Francesco Polese. A brief review of systems theories and their managerial applications. *Service science*, 2(1-2):126–135, 2010.

[33] A Wayne Wymore. *Model-based systems engineering*, volume 3. CRC press, 2018.

[34] Fremont E Kast and James E Rosenzweig. General systems theory: Applications for organization and management. *Academy of management journal*, 15(4):447–465, 1972.

[35] Norman S Nise. *Control systems engineering*. John Wiley & Sons, 2020.

[36] Yingxu Wang. Towards the abstract system theory of system science for cognitive and intelligent systems. *Complex & Intelligent Systems*, 1(1):1–22, 2015.

[37] Jeff A Estefan et al. Survey of model-based systems engineering (mbse) methodologies. *Incose MBSE Focus Group*, 25(8):1–12, 2007.

[38] Alejandro Salado and Aditya U Kulkarni. An assessment of the adequacy of common definitions of the concept of system. In *INCOSE International Symposium*, volume 31, pages 510–521. Wiley Online Library, 2021.

[39] Mihajlo D Mesarovic and Yasuhiko Takahara. Complex systems. *Abstract Systems Theory*, pages 330–414, 2006.

[40] Robert Laurini and Derek Thompson. *Fundamentals of spatial information systems*, volume 37. Academic press, 1992.

[41] Mihajlo D Mesarovic and Yasuhiko Takahara. *General systems theory: mathematical foundations*. Academic press, 1975.

[42] MIHAJLO D MESAROVIC. Conceptual basis for a mathematical theory of general systems. *Kybernetes*, 1(1):35–40, 1972.

[43] David Krakauer, Nils Bertschinger, Eckehard Olbrich, Jessica C Flack, and Nihat Ay. The information theory of individuality. *Theory in Biosciences*, 139(2):209–223, 2020.

Appendix C

Exploring Outcome-Based Biological and Artificial Intelligence Through Core and Periphery Precepts

Niloofar Shadab, Tyler Cody, Alejandro Salado, Peter Beling

Ready to be Submit to Journal of Systems Research and Behavioral Science

Abstract— Engineering methodologies predominantly revolve around established principles of decomposition and recomposition. These principles involve the partitioning of inputs and outputs at the component level, ensuring that component-specific properties remain intact upon composition. Our prior research contends that the engineering of general intelligence necessitates a fresh set of overarching systems principles. As a result, we introduced a novel set of principles, denoted as the ”core and periphery”. These principles have been explored in-depth within the context of abstract systems theory and the Law of Requisite Variety. In this paper, we assert that these abstract concepts hold practical significance. To support this claim, we present empirical evidence demonstrating the applicability of these principles in both biological and artificial intelligence systems.

C.1 Introduction

Early cyberneticists, most notably Ashby, harnessed the concept of ”variety” to model phenomena within complex systems. In his studies, Ashby specifically employed the concept of variety to explore homeostasis, which refers to a system's capacity to maintain specific variables within narrow bounds, even in the face of changing contexts, focusing on biological systems [1, 2, 3].

Variety, as understood here and throughout, signifies the count of distinct elements. In this context, *system* variety pertains to the number of unique elements within a system, while *context* variety characterizes the number of unique elements present within the system's context [4]. Typically, variety is defined concerning states, but it also extends to encompass inputs, denoted by $\mathcal{X}$, and outputs, denoted by $\mathcal{Y}$.

The main result of Ashby's research is a principle termed the *Law of Requisite Variety* [3]. It states that for a system to be stable, the system variety must be greater than or equal to the variety in the context it regulates. Formally put, given a system S and context C,

$$S \ is \ stable \rightarrow Variety(S) \geq Variety(C). \tag{C.1}$$

Ashby presented a second, less-referenced, but similarly fundamental result. Let an outcome be an appearance of a particular element or co-occurrence of elements in the context. The variety of possible outcomes is lower-bounded by the difference between system variety and context variety. In other words, a system's (relative) variety limits the precision with which it can regulate phenomena in the context. Formally put, for a system S, context C, and outcomes O [5],

$$Variety(O) \geq Variety(C) - Variety(S). \tag{C.2}$$

Naturally, $Variety(O) \geq 0$, so in the case where system variety is greater than context variety the lower-bound is 0.

In summary, Condition C.1 suggests stability requires a system's variety to be able to scale to match the variety of a system's contexts. Condition C.2 suggests that when the context variety is not well-matched by system variety, the set of possible outcomes is necessarily large and the system will struggle to achieve precise regulation of its context.

The Law of Requisite Variety found immediate and long-lasting use in organization management [6, 7, 8]. Recently, biological theoreticians have used requisite variety to posit how the mind emerges from the brain in a theory termed practopoiesis [9], and have suggested that deep learning methods lack the variety required to scale to human-level intelligence [10]. Though not explicitly connected, the importance of variety is echoed in learning theoretic notions of capacity and local learning [11] as well as in emerging notions of free energy minimization and active inference [12].

We assert that the Law of Requisite Variety can be effectively harnessed to structure Systems Engineering (SE) practices, distinguishing between open and closed paradigms, particularly within the context of intelligent systems. To facilitate the practical application of this concept, we have introduced the "core and periphery" precepts. This framework enables the division of systems into core and periphery components, allowing them to be methodically engineered in accordance with the principles of open-view and closed-view SE, as deemed appropriate. Subsequently, we present compelling evidence that underscores the relevance of the core and periphery precepts in modeling the behaviors and structures of intelligent systems.

The paper's structure is organized as follows: In the following section, we will provide a concise recap of our previous work concerning the formalization of the core and periphery precepts, elucidating its derivation within the context of the Law of Requisite Variety. Following that, we will offer a brief overview of how intelligent systems can benefit from the core and periphery framework, emphasizing the distinct characteristics of their structure, functionality, and behavior. Finally, we reinforce our argument with real-world examples that illustrate how these systems can be effectively represented as core and periphery components.

C.2 Background

In this section, we offer a comprehensive overview of the formal definitions of key concepts, as derived from the Law of Requisite Variety in our prior work, employing the perspective of general systems theory [13]. The general systems theory's lens facilitates the application of the concept of variety in SE practices.

In the domain of general systems theory, systems are typically characterized as relations among sets. General systems theory predominantly deals with the fundamental principles governing relationships within sets. These principles can span across categories, topologies, algebraic structures, and more. However, it is important to note that the study of set theory, on its own, can provide valuable insights into the fundamental nature of these specific concerns. As mentioned earlier, Ashby introduced a particular concept of ”variety” to investigate homeostasis in biological systems [1, 14]. In our previous work, we provided a set-theoretic redefinition of the concept of ”Law of Requisite Variety”.

Consider two systems S and S_E where $S : \mathcal{X} \to \mathcal{Y}$ and $S_E : \mathcal{X}_E \to \mathcal{Y}_E$. Without loss of generality term S the system and S_E the context. Suppose S is acting as a regulator of S_E. Let $\mathcal{X}_{E\setminus S} = \mathcal{X}_E \setminus \mathcal{Y}$ where $\setminus$ denotes set difference. In other words, inputs to the context $\mathcal{X}_E = \mathcal{X}_{E\setminus S} \cup \mathcal{Y}$. Consider a set of outcomes $\mathcal{Z}$ with support over $\mathcal{X}_{E\setminus S} \times \mathcal{Y}$, i.e., $\mathcal{X}_{E\setminus S} \times \mathcal{Y} \to \mathcal{Z}$. Alternatively, variety can be expressed in terms of information complexity, a concept that Ashby quantified using the formula $log_2(n)$, where n corresponds to the count of unique elements [1]. Let V_A be termed *variety* and be the Shannon entropy of a finite set A, i.e.,

$$V_A = -\sum_{i}^{|A|} p_i \log_2 p_i \tag{C.3}$$

Where $|A|$ denotes the cardinality of A and p_i the probability of the i^{th} element of A. Variety describes the number of unique elements in a system. The "Law of Requisite Variety" stipulates that in order for one system to effectively serve as a stable regulator for another, the variety present in the regulator's output must be either greater than or at least equal to the variety in the input of the system being regulated. In a more explicit formulation, this law posits that V_Y (referring to the regulator's variety) must surpass or be on par with $V_{X_{E\setminus S}}$ (the variety associated with the context's input) for the achievement of precise, well-defined outcomes. This law implies that when the variety in the system's output fails to match with the variety present in the context's input, it results in a broader spectrum of possible outcomes, thereby making the attainment of precise results more challenging. In the words of Ashby, this law essentially conveys that a system's "capacity as a regulator cannot exceed its capacity as a channel for variety" [14]. Formally put, consider that (from [14])

$$\min V_Z = \max\{V_{\mathcal{X}_{E\setminus S}} - V_{\mathcal{Y}}, 0\}. \tag{C.4}$$

To help formalizing core and periphery concept, the Law of Requisite Variety can be redefined as follows.

Definition C.1 (Law of Requisite Variety)**.** The *Law of Requisite Variety* states that given $V_{\mathcal{X}_{E\setminus S}}$, the minimum variety of outcomes $\min V_Z$ only decreases if V_Y increases.

Only if $V_Y \geq V_{X_{E\setminus S}}$, is it information theoretically possible to determine outcomes $\mathcal{Z}$, i.e.,

$\min V_Z = 0$.

Ashby linked a system's survival to the bounding of varieties [1]. Bounded varieties refer to system varieties that remain constant, while unbounded varieties pertain to those that change. This distinction allows us to identify core components of a system with bounded varieties, and the peripheral components with unbounded varieties. In the following, we will delve into how the Law of Requisite Variety serves as the foundation for establishing the core and periphery precepts.

To define core and periphery, let S be a system $S \subset \times\{\mathcal{X}, \mathcal{Y}\}$ and let $\overline{S}$ denote the component sets of S, i.e., $\{\mathcal{X}, \mathcal{Y}\}$. Let $\mathcal{X}^t$ denote the input structure at time t, and so forth. Bounded and unbounded varieties are distinguished by measuring the variety of a system's residual change over time. Let R denote this residual change, i.e.,

$$R_{\overline{S}}^{t,t'} = \{\mathcal{X}^{t'} \setminus \mathcal{X}^{t}, \mathcal{Y}^{t'} \setminus \mathcal{Y}^{t}\} \tag{C.5}$$

$R_{\overline{S}}^{t,t'}$ gives the residual change in system structure between time t and t'. The core and periphery are defined as follows.

Definition C.2 (Core and Periphery)**.** Consider a system S at time t and at a later time t'. The *core* of S from t to t' is

$$\mathcal{C}_{\overline{S}}^{t,t'} = \overline{S} \setminus R_{\overline{S}}^{t,t'} \tag{C.6}$$

The *periphery* of S from t to t' is

$$\mathcal{P}_{\overline{S}}^{t,t'} = R_{\overline{S}}^{t,t'}. \tag{C.7}$$

The core are those elements of S's component sets that are identical at times t and t', and the periphery are those elements that are not. As a result, the core and periphery precepts can be utilized to model relative balance of variety in system and context as well as to understand how system addresses variety in the context [13].

In summery, the Law of Requisite Variety serves as a fundamental principle underpinning our core-periphery precepts. In the next section, we will transition from the formalization of these concepts to explore the unique features of intelligent systems that make them particularly amenable to core-periphery modeling. We will delve into how the core and periphery precepts are applicable in modeling the structure, behavior, and organization of these complex systems.

C.3 Intelligent Systems

The assessment of relative balance of variety in system and its context, as well as the system's adaptation to context variety, can be longitudinally observed to comprehend the learning or evolution of the system-context relationship in terms of core and periphery. With this perspective, we propose that core and periphery constitute pertinent principles for the engineering of intelligent systems, particularly when intelligence is defined as an attribute intrinsic to the interaction between the system and its context.

C.3.1 Open and Closed System Phenomena

Open systems and closed systems precepts are important concepts in the SE practices [15]. Systems can be engineered with different spectrum of openness and closedness. On one extreme, as shown in C.1-A, an open system S exists in isolation, with no notion of context other than as everything outside the boundary. Loosening this, there exists a scoped context C around S, and, further, there are other open systems in C that receive S's outputs $\mathcal{Y}$ or produce S's inputs $\mathcal{X}$, as shown in Figure C.1-B. The context is prepared for closure when

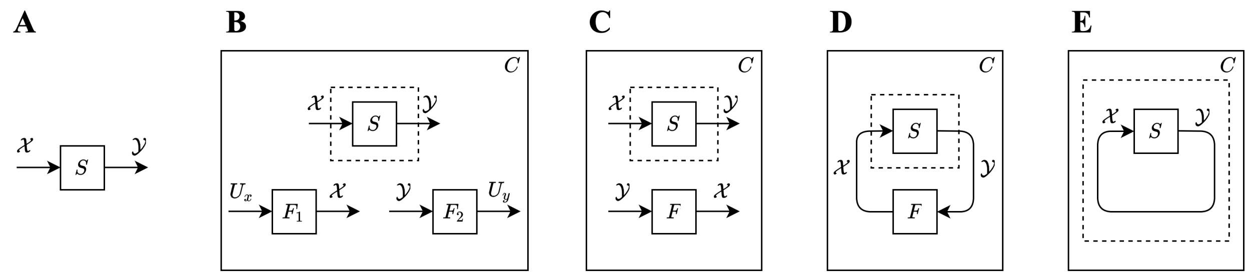


Figure C.1: A depiction of the spectrum between open systems (A) and closed systems (E).

S is able to discover how other systems combine to produce $\mathcal{X}$ from $\mathcal{Y}$, as shown in Figure C.1-C. S can use this relation to achieve closure, as shown in Figure C.1-D. The coupling becomes stronger, and ultimately reaches the point where the system producing S's inputs is integrated with S itself. As shown in Figure C.1-E, the scope of S expands to the limits of C, and S becomes a closed system—the other extreme.

Closed systems S have inputs $\mathcal{X}$ that are influenced by their outputs $\mathcal{Y}$ by way of context C. Intelligent systems S are expected to influence their context C. And since the closure of S is mediated by C, S is expected to influence the nature of their closure with C. In other words, intelligent systems have influence over their coupling with context. This suggests that even if intelligence can be relegated to a subsystem at conception, the boundaries between the intelligent system and those under its influence face dissolution as the intelligent system and its context intertwine. This is depicted by the waves of influence in Figure **??**-A, first shown in blue when influencing the greater system, then in red when influencing context directly. Under the closed view, in some ways, the system is just a vessel for intelligent subsystems to gain influence over context.

The open-system view may suffice in certain contexts, but as algorithms and their applications become increasingly intricate, the closed-system view might become indispensable as the closed effects of the context and system will rise.

Putting open-view and closed-view dogma aside, consider now the distribution of open and closed system phenomena across the core and periphery. The core is given by system varieties that are invariant to context. So, the core is decoupled from context. It follows then that the core corresponds to open system phenomena. By the same logic, the periphery corresponds to the part of the system coupled with context, i.e., to closed system phenomena. Importantly, while a distributed core may be formed by boundary-dissolving closed system phenomena, once it forms, its invariance to changes in context characterize it by open system phenomena.

Adaptation to changes in context involves interplay between the core and periphery precepts, and thus interplay between open-view and closed-view modeling of the system. To see this, consider that the periphery, by way of its coupling, is the part of the intelligent system that initially responds to changes in context. Put differently, by definition, the core is invariant to changes in context, so new information must enter by way of the periphery. This leads to the first principle.

Principle 1. The adaptation of intelligent systems to changes in context is principally a closed system phenomena (from the periphery) resulting in open system phenomena (in the core).

Principle 1 suggests that in order to adapt the core to new information, new varieties from the periphery must be identified and progressively bounded in a process that involves both closed and open system phenomena.

C.3.2 Scaling Intelligence

The core and periphery precepts could be utilized to explain the scale of intelligent property in systems. Addressing the scalability of intelligence is perhaps the strongest motivation for orienting the engineering of intelligent systems towards the concepts of core and periphery.

Presently, the notion of scale in Artificial Intelligence (AI) and Machine Learning (ML) systems primarily centers around parameters like the number of queries per second, the number of user interactions per second, or data processing volume. In essence, scalability is often measured by a system's ability to handle increased input and output complexity. However, this perspective on scaling is rather limited.

A more comprehensive approach to measuring intelligence's scalability considers the system's variety. Expanding the number of input and output elements results in the scaling of variety. Nonetheless, it's crucial to recognize that scaling inputs and outputs represents just one special case of scaling system variety, and it is not the most generalized one. Focusing solely on scaling inputs and outputs aligns with an open-system perspective. A focus on inputs and outputs corresponds to scaling bounded variety, but, at some point, scaling intelligence requires engineering intelligent systems described predominantly by unbounded variety [16]. In other words, when confronted with the challenge of scaling intelligence, engineering intelligent systems that primarily adhere to closed system phenomena becomes essential [16].

To illustrate this, imagine a scenario where the context's variety grows indefinitely with increased scaling. Eventually, by logical necessity, the system's variety, which matches its bounded varieties with unbounded ones in the context, cannot maintain stability or attain precise outcomes. Consequently, the onus of scaling intelligence gravitates toward the system's unbounded varieties – its periphery – and, in turn, toward closed system phenomena. This insight leads to the second fundamental principle:

Principle 2. At some point, scaling the variety of an intelligent system relies predominantly on closed system phenomena.

Yet, one might wonder why it is crucial to scale variety. Equations C.1 and C.2 make it evident that scalability is a prerequisite for both stability and achieving precise outcomes. The strong couplings between systems and their contexts essentially prevent us from establishing a clear boundary between system inputs and outputs when the coupling is potent. The concept of closure can be used in modeling process in the extreme case of strong coupling, where the system's boundary dissolves with the context [16]. This emphasizes the importance of adopting precepts that aren't contingent on steadfast system boundaries. Therefore, given the close relationship between coupling and scale, the core and periphery concepts prove their relevance as essential precepts for engineers when scalability in intelligence is the primary objective.

C.4 Evidence of Core-Periphery Precepts in Real-World

In this section, we embark on a journey to explore the compelling evidence of core and periphery precepts manifesting in real-world systems. Our goal is to unveil and meticulously examine the persuasive evidence supporting the feasibility of modeling structural, behavioral, or organizational aspects within these intricate real-world systems through the lens of core-periphery concepts.

C.4.1 DNA and Neurons

The DNA structure encapsulates key features of an intelligent system; however, its slow adaptation to changes unfolds over generations, influencing the fundamental characteristics of the intelligent system [17]. The alterations in DNA structure, essential for defining the system's traits, undergo a gradual process that spans multiple generations. In contrast, the intricate coordination and communication among neurons in the brain enable a broad spectrum of tasks, ranging from simple reflexes to complex cognitive functions like problem-solving and decision-making [18].

Neurons serve as both information processors and signal transmitters, facilitating the brain and nervous system's effective response to the environment and the execution of various tasks. This structural arrangement allows intelligent systems to maintain core characteristics, such as the number of legs or skull structure, while also providing flexibility for evolution in response to environmental demands. The brain's ability to increase neuronal connections contributes to the system's variety, enabling adaptations to the environment. Over time, these adaptive changes, driven by the brain, influence the DNA to transition towards a more stable core by addressing emerging needs arising from contextual adaptation [4].

The concept of "survival of the fittest" is rooted in the idea that genetic variations leading to better adaptation to the environment increase the likelihood of survival and subsequent inheritance in subsequent generations [19].

The brain's capacity for self-reorganization, characterized by the formation of new neural connections, is an example of periphery. This capability enables adaptation through learning, recovery from injury, and adjustments to changes in sensory input. In contrast, DNA represents a tightly bounded variety. The nervous system, acting as the periphery, exhibits a lesser degree of constraint. This conceptualization gives rise to the core and periphery framework in intelligent systems—featuring a core set of tightly bounded varieties enveloped by layers of progressively less constrained varieties extending into the system's periphery, where variety is unbounded. The core and periphery dynamics, in terms of survival, involve preventing disorder from penetrating the periphery deeply enough to reach the core. The approach to model the core and periphery diverges significantly based on whether one adopts an open or closed view.

C.4.2 Homeostasis vs Homeodynamic

In the study of biological systems, two fundamental concepts are often employed to describe stability and equilibrium: homeostasis and homeodynamics [20, 21]. These concepts are essential for understanding how living organisms maintain stability in different ways.

Homeostasis is a fundamental physiological and biological concept that refers to the body's ability to maintain a stable and balanced internal environment, despite external changes and fluctuations [22]. It involves the regulation of various factors such as temperature, pH levels, blood pressure, and the concentration of nutrients and gases within the body [23]. The key principle of homeostasis is to keep these internal conditions within a narrow and optimal range, which is essential for the proper functioning of cells, tissues, and organs [21]. When external factors or internal processes disrupt this balance, the body employs various mechanisms to restore equilibrium. For example, if body temperature rises due to external heat, mechanisms like sweating and dilation of blood vessels help cool the body down. Conversely, if body temperature drops in a cold environment, shivering and constriction of blood vessels help generate and conserve heat.

On the other hand, homeodynamic is a term used to describe a state of dynamic equilibrium or balance within a biological system [24]. Homeodynamic points out to the biological systems' ability to dynamically self-organise at bifurcation points of their behaviour where they lose stability [20]. For example, the immune system is highly dynamic. It can mount a rapid response to pathogens, adapt to new threats, and return to a resting state once the threat is eliminated. This dynamic equilibrium allows the immune system to protect the body without causing excessive inflammation. Another example of dynamic equilibrium is metabolism. Metabolism rate can have dynamic equilibrium that actively adjust with the body needs and activities.

Homeostasis can be viewed as the result of spatio-temporal chaotic dynamics, as discussed by Ikegami [25]. In this context, the homeodynamic variables within biological systems exhibit temporal changes, while the parameters governing homeostasis remain constant over time [25]. Drawing a parallel to the core and periphery framework, we can equate homeostasis process aligned with the core, while homeodynamics process finds its place in the the periphery model within these biological systems. This alignment strongly indicates the presence of a specific organizational structure within biological systems that lends itself to modeling through the core and periphery principles.

C.4.3 Interrogative Attitude vs Belief

This paper delves into a philosophical discourse concerning the nature of inquiry and its role on transforming interrogative attitudes into beliefs and vice versa. Within this philosophical context, we elucidate how the cognitive dynamics of an intelligent mind operate and, notably, how these facets correlate with the previously introduced conceptual framework of core and periphery within this paper.

Suspension of judgment; a state of mind that one withholds judgement [26]; emerges as an integral facet closely intertwined with the process of inquiry [27]. It is imperative to recognize that the act of suspending judgment is a deliberate step taken in the pursuit of genuine inquiry. Inquiry, in this context, is defined as an interrogative attitude, representing a specific cognitive state of mind. Within the realm of interrogative attitudes lie various psychological states and processes, encompassing curiosity, wonder, contemplation, deliberation, and more [27]. These attitudes can manifest as either static states or dynamic processes. Consequently, inquiry emerges as a manifestation of a goal-oriented cognitive state that necessitates corresponding actions to foster and conduct the inquiry. The essence of inquiry lies in the act of asking questions, which serves as a goal-directed endeavor aimed at acquiring specific epistemic information.

A crucial distinction emerges between belief and interrogative attitude. Belief signifies the possession of a complete answer to a given question, denoted as Q_1, at a specific time, t_1 [27]. In contrast, an interrogative attitude represents a distinct cognitive state of mind wherein one actively seeks a comprehensive answer to a different question, labeled as Q_2, at a subsequent time, t_2. Notably, a belief held at time t_1 can transition into an interrogative attitude at a subsequent time, t_2, when a suspension of judgment occurs and additional information becomes available [27]. Conversely, the initiation of a suspension of judgment marks the point at which the transformation begins. For instance, a detective aiming to identify the criminal in a case often starts with an initial suspect and poses goal-oriented questions while exploring additional evidence. Throughout this process, the detective may exhibit various attitudes such as questioning, curiosity, and wonder. However, as the investigation progresses and sufficient inquiries are addressed, a definitive belief about the guilt of a particular individual can be established.

In this context, it is essential to recognize that interrogative attitudes undergo more frequent fluctuations, and these shifts in attitude through inquiries inherently culminate in a transformative process that ultimately influences changes in one's beliefs. Beliefs, conversely, tend to exhibit a higher degree of stability. This stability persists unless the ongoing inquiry successfully yields a comprehensive answer to the posed question.

The distinction between interrogative attitude and belief further underscores their fundamental disparities. Interrogative ttitude, by nature, adheres to a closely defined and goal-oriented paradigm. It operates as an active and dynamic cognitive endeavor, inherently focused on a specific objective. In contrast, belief adopts a more open-view conceptual framework. It operates akin to a function, mandating the attainment of a complete answer, represented as outputs, to a given question, denoted as inputs [28]. This distinction underscores the fundamental dichotomy between interrogative attitude and belief.

Essentially, the interrelationship between interrogative attitude and belief encapsulates a dynamic interplay, wherein the evolving attitude often acts as the harbinger of alterations within one's belief system. This dynamic can be aptly modeled or conceptualized through the framework of the core and periphery precepts. Within this context, beliefs are a set of states that can be modeled as bounded varieties encompassed by the core. This core represents the central and more stable component of the belief system. Conversely, interrogative attitudes assume the role of the periphery of the belief system. Periphery represents the dynamic and ever-changing aspect of one's cognitive framework within this system.

C.5 Evidence of Core-Periphery Precepts in Engineered Intelligent Systems

In the previous section, we presented supporting evidence for the applicability of the core and periphery precepts in modeling biological systems. In this section, we delve into analogous evidence within engineered intelligent systems.

We assert that the structure of even the simplest intelligent systems exhibits evidence of the applicability of the core and periphery precepts. To substantiate this claim, we devised an experiment involving a Convolutional Neural Network (CNN). The aim was to assess alterations in the weights of its fully connected layer when confronted with significant changes in context through its inputs set.

The experiment was structured as follows: Utilizing the ResNet-50 model, an existing CNN model from Tensorflow [29], we conducted training on the CIFAR-10 dataset. This dataset

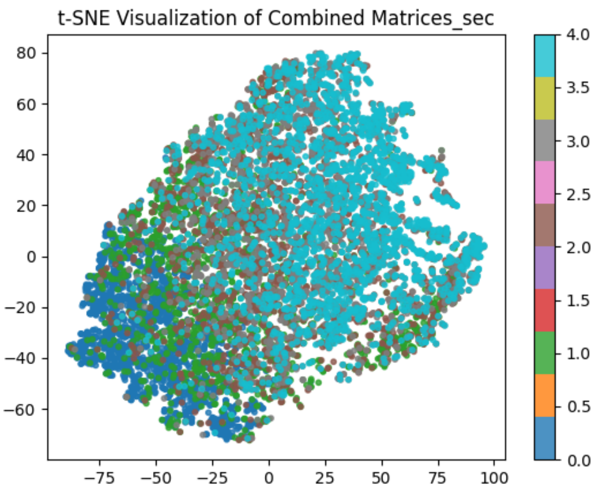


Figure C.2: The t-SNE visualization of combined weights of 1^{st} epoch, 10^{th} epoch, 20^{th} epoch, 30^{th} epoch, 40^{th} epoch from the second train of the ResNet-50 after we subtracted the weights of the 40^{th} epoch of the first run with each of the epochs mentioned above

encompasses 60,000 images of size 32×32, distributed across 10 classes. Throughout the training of the ResNet-50 model, we recorded the weights of the fully connected layer for each epoch, spanning a total of 40 epochs.

Subsequently, we employed the previously trained model, with its stored weights, and retrained it using CIFAR-100, a dataset of similar nature but with increased complexity and a total of 100 classes. This modification in the model reflects potential alterations in the context that can be simulated by changes in the input set of a CNN model. We recorded the weights of the fully connected layer for each epoch during the retraining of the model, extending over a total of 40 epochs. This experiment enables a comparative analysis of the weights across epochs of the second trained model with those of the first trained model, shedding light on any discernible patterns in the changing weights of the fully connected layer.

In this experimental setup, weights serve as a simplified representation of system variety.

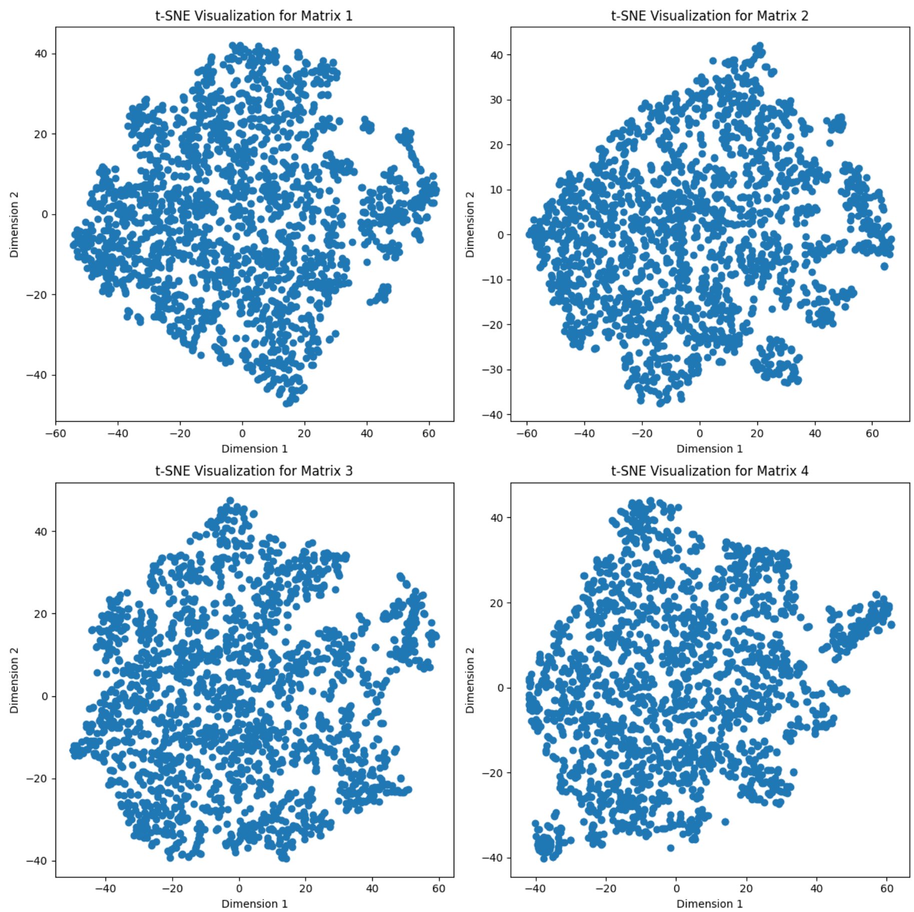
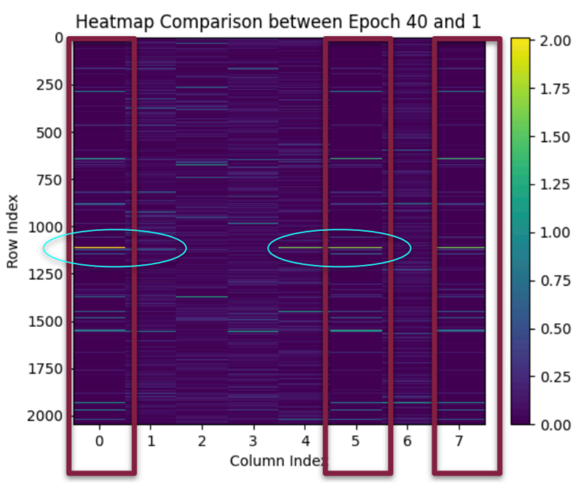
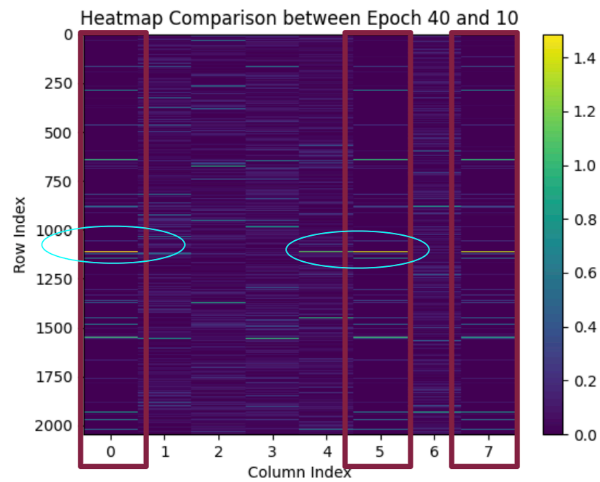


Figure C.3: The Comparison of the weights difference of $1^{st}, 10^{th}, 20^{th}, 30^{th}$ with the weight difference of 40^{th} epoch of the second model. The weights of the epochs in the second model were already subtracted from the 40^{th} epoch of the first model

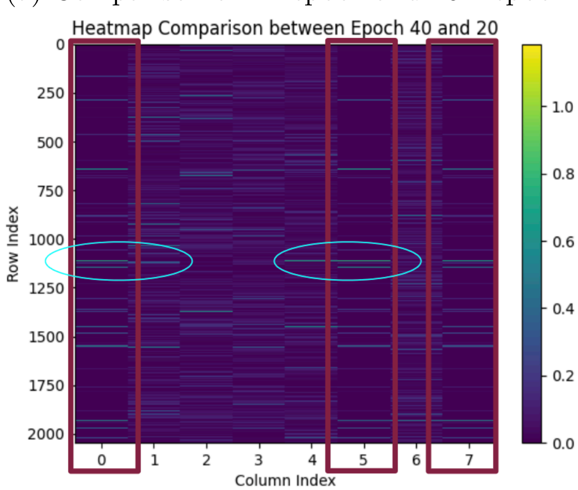
Layers with minimal to no changes in weights can be conceptualized based on the principles of the core precept (open-systems view). Conversely, layers exhibiting significant weight variations in response to changes in the input set (context variety) can be conceptualized through the periphery precept. It is essential to emphasize that this example serves as an illustration of such conceptualization within a CNN model. The primary objective is not to assess the practical utility of modeling the fully-connected layer with core and periphery precepts. As previously mentioned, the core and periphery model is most fitting for types of intelligence that represent a highly coupled relational property between the system and its



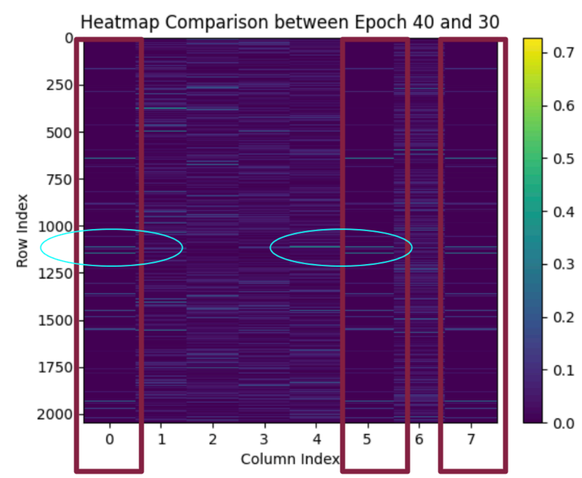
(a) Comparison of 1^{st} epoch and 40^{th} epoch



(b) Comparison of 10^{th} epoch and 40^{th} epoch



(c) Comparison of 20^{th} epoch and 40^{th} epoch



(d) Comparison of 30^{th} epoch and 40^{th} epoch

Figure C.4: The heatmap visualization of comparison of the weights difference of the $1^{st}, 10^{th}, 20^{th}, 30^{th}, 40^{th}$ in second model in relation with the 40^{th} epoch of the first model context.

Figure C.2 presents a visualization of the combined weight distribution of the epochs in the second model. We utilized t-SNE, a dimension reduction technique, to facilitate the visualization of the high-dimensional data from the fully connected layers of ResNet-50. Initially, we selected four epochs from the second training model and computed the weight differences by subtracting the weights of these epochs from the 40^{th} epoch of the first model. Subsequently, we compared these weight differences using t-SNE visualization. As depicted in Figure C.2, the weights shifted from negative to positive dimensions as the epochs progressed. However, a central region with minimal dimension differences between all epochs is observable. For improved visibility, we isolated the weight differences of the $1^{st}, 10^{th}, 20^{th}, 30^{th}$ epochs and compared them with that of the 40^{th} epoch from the second model. Figure C.3 illustrates that the first and second comparisons (top left and top right) exhibit weights shifted to the left, while in the third and fourth comparisons, the weights shifted to the right. Nevertheless, consistent patterns that remained unchanged across all four comparisons are discernible.

To associate the observed patterns of weight change with their respective layers in the fully-connected layer structure of the ResNet-50 model, we represented the weight differences of the mentioned epochs on heatmaps for all eight layers of the fully-connected layer structure. Subsequently, we compared the four heatmaps of weight differences to identify evidence of core and periphery. Specifically, we focused on identifying the layers with the lowest rate of weight change in the heatmaps. To achieve this, we generated four distinct heatmaps based on the t-SNE diagrams presented in Figure C.3.

Figure C.4 displays all four heatmaps, highlighting the three layers with the lowest weight change through red rectangular shapes. These layers predominantly represent the core of the ResNet-50 model in this specific illustration. Nonetheless, even within these layers, there are regions denoted by blue circles that signify areas with significant changes. As the epochs progress, these areas exemplify the periphery that accommodates changes in the context and adapts through higher weight changes.

C.6 Core-Dominant vs Periphery-Dominant Systems

To bound the outcomes of an intelligent system, it is consequential to determine if the system is being modeled as a core-dominant system or periphery-dominant one. Core-dominant model of the system indicates that the characteristics of the system depends more on the behaviors (outputs) of the core whereas periphery-dominant model of the system relies mostly on the behaviors (outputs) of the periphery part. In this section, we will elucidate the conditions for having a core-dominant vs a periphery-dominant model of a system.

C.6.1 Core-Dominant Systems

Core-dominant systems are characterized by having greater variety in their core compared to their periphery, denoted as $V_{\mathcal{C}} \geq V_{\mathcal{P}}$. This classification is attributed to the core's role in impeding more variety from the system's environment [13]. As mentioned earlier, Shannon's entropy can be used to capture the variety as shown in Equation C.3. The entropy of a system will be determined by the joint entropy of the core and periphery parts of the system. From the entropy equation $H(X,Y) = H(X) + H(Y|X)$ [30]; we can derive:

$$\begin{aligned} H(\overline{S}) &= H(\mathcal{C}_{\overline{S}}^{t,t'}) + H(\mathcal{P}_{\overline{S}}^{t,t'}|\mathcal{C}_{\overline{S}}^{t,t'}) \\ &= H(\mathcal{P}_{\overline{S}}^{t,t'}) + H(\mathcal{C}_{\overline{S}}^{t,t'}|\mathcal{P}_{\overline{S}}^{t,t'}) \end{aligned} \tag{C.8}$$

As mentioned earlier, core is not prone to residual changes of inputs and outputs. Without loss of generality, if $\mathcal{C}_{\overline{S}}^{t,t'}$ is the core of our system; $\overline{S}$; we can have a core-dominant model of the system if we have this constraint as follows:

$$H(\mathcal{C}_{\overline{S}}^{t,t'}|\mathcal{P}_{\overline{S}}^{t,t'}) \leq H(\mathcal{C}_{\overline{S}}^{t,t'}) \leq \delta \tag{C.9}$$

δ represents a modeling threshold, specifying that the conditional entropy of the core given the periphery should be less than or equal to a certain value. This condition ensures that there is adequate entropy (i.e., variety) in the core to exhibit core-dominant characteristics in the system's model while maintaining the entropy of $\overline{S}$ at an acceptable level. Equation C.9 is derived from the relation between the system ($\overline{S}$) and its consisting parts ($\mathcal{C}_{\overline{S}}^{t,t'}$, $\mathcal{P}_{\overline{S}}^{t,t'}$). Equation C.9 is a result of simple assumptions and information theoretical equations as follows:

Per Core-dominant Definitions:

$H(\mathcal{C}_{\overline{S}}^{t,t'}) \geq H(\mathcal{P}_{\overline{S}}^{t,t'})$

From Equation C.8 and inequality above:

$H(\mathcal{C}_{\overline{S}}^{t,t'}|\mathcal{P}_{\overline{S}}^{t,t'}) \geq H(\mathcal{P}_{\overline{S}}^{t,t'}|\mathcal{C}_{\overline{S}}^{t,t'})$

From information theory and inequality above:

$H(\mathcal{P}_{\overline{S}}^{t,t'}|\mathcal{C}_{\overline{S}}^{t,t'}) \leq H(\mathcal{P}_{\overline{S}}^{t,t'}) \leq H(\mathcal{C}_{\overline{S}}^{t,t'}|\mathcal{P}_{\overline{S}}^{t,t'}) \leq H(\mathcal{C}_{\overline{S}}^{t,t'})$

According to the definition of core-dominant system provided in Equation C.9, the system's characteristics is more dependant on the the core characteristics. This interpretation infers that not only the core has more entropy (i.e., variety) but also the conditional entropy of the core given periphery; $H(\mathcal{C}_{\overline{S}}^{t,t'}|\mathcal{P}_{\overline{S}}^{t,t'})$; will be greater than the entropy of the periphery part. Therefore, core-dependant systems can be modeled predominantly with the open-view paradigm that depends on the input-output relations between the system and its environment.

C.6.2 Periphery-Dominant Systems

In contrast to core, periphery is prone to residual changes in inputs and outputs sets. Periphery-dependent systems by definition allow for these residual changes and grow and/or modify the possible sets of outcomes based on such changes in periphery. In a periphery-dominant model of systems, changes in the outcomes of systems depend more on the changes in the context than on the bounded structure and behavior of the core. Without loss of generality, consider $\mathcal{C}_{\overline{S}}^{t,t'}$; as the core of the system and the rest of the system will be the periphery part; $\mathcal{P}_{\overline{S}}^{t,t'}$. We can have a periphery-dominant model of the system with a condition as follows.

$$H(\mathcal{P}_{\overline{S}}^{t,t'}|\mathcal{C}_{\overline{S}}^{t,t'}) \leq H(\mathcal{P}_{\overline{S}}^{t,t'}) \leq \gamma \tag{C.10}$$

γ represents a modeling threshold, specifying that the conditional entropy of the periphery given core should be less than or equal to a certain value. This condition ensures that there is adequate entropy (i.e., variety) in the periphery to exhibit periphery-dominant characteristics in the system's model while maintaining the entropy of $\overline{S}$ at an acceptable level.

Similar to Equation C.9, Equation C.10 is a result of simple assumptions and information theoretical equations as follows:

Per Periphery-dominant Definitions:

$H(\mathcal{P}_{\overline{S}}^{t,t'}) \geq H(\mathcal{C}_{\overline{S}}^{t,t'})$

From Equation C.8 and inequality above:

$H(\mathcal{P}_{\overline{S}}^{t,t'} | \mathcal{C}_{\overline{S}}^{t,t'}) \geq H(\mathcal{C}_{\overline{S}}^{t,t'} | \mathcal{P}_{\overline{S}}^{t,t'})$

From information theory and inequality above:

$H(\mathcal{C}_{\overline{S}}^{t,t'} | \mathcal{P}_{\overline{S}}^{t,t'}) \leq H(\mathcal{C}_{\overline{S}}^{t,t'}) \leq H(\mathcal{P}_{\overline{S}}^{t,t'} | \mathcal{C}_{\overline{S}}^{t,t'}) \leq H(\mathcal{P}_{\overline{S}}^{t,t'})$

According to the definition of periphery-dominant systems provided in Equation C.10, the system's characteristics is more dependant on the the periphery characteristics. This interpretation infers that not only the periphery has more entropy but also the conditional entropy of periphery given the core; $H(\mathcal{P}_{\overline{S}}^{t,t'} | \mathcal{C}_{\overline{S}}^{t,t'})$; is greater than the entropy of the core.

In conclusion, as stated in [13], without the need to have assumptions of functional dependence of the components of a system, as commonly done in traditional engineering practices that rely on decomposition and recomposition precepts, the principles of core and periphery can be employed. These principles offer a model that elucidates the aspects of an intelligent system used to constrain environmental variety and regulate outcomes.

C.7 Conclusion

In this paper, we have presented real-world examples demonstrating the practicality of the core and periphery precepts. In our earlier work, we introduced and formalized these precepts, suggesting their potential utility in engineering intelligent systems through the application of open-view and closed-view principles. In this context, we conducted a small experiment aimed at unveiling the structure of a rather simple intelligent system viewed through the core and periphery framework. These examples provided us with evidence indicating that not only can the behaviors and structures of biological systems be elucidated through the core and periphery precepts, but engineered intelligent systems also exhibit analogous structures and behaviors.

Bibliography

[1] W Ross Ashby. *An introduction to cybernetics*. Chapman & Hall Ltd, 1961.

[2] W Ross Ashby. Principles of the self-organizing system. *Modern systems research for the behavioral scientist*, 108, 1968.

[3] George J Klir and W Ross Ashby. Requisite variety and its implications for the control of complex systems. *Facets of systems science*, pages 405–417, 1991.

[4] Markus Schwaninger and Stefan Ott. What is variety engineering and why do we need it? *Systems Research and Behavioral Science*, 2023.

[5] W Ross Ashby and Jeffrey Goldstein. Variety, constraint, and the law of requisite variety. *Emergence: Complexity and Organization*, 13(1/2):190, 2011.

[6] JDR de Raadt. Ashby's law of requisite variety: An empirical study. *Cybernetics and Systems: An International Journal*, 18(6):517–536, 1987.

[7] Elinor Ostrom. Designing complexity to govern complexity. *Property rights and the environment: Social and ecological issues*, pages 33–45, 1995.

[8] Max Boisot and Bill McKelvey. Complexity and organization-environment relations: Revisiting ashby's law of requisite variety. *P. Allen, S. Maguire, & B. McKelvey, Complexity and Management*, pages 279–298, 2011.

[9] Danko Nikolić. Why deep neural nets cannot ever match biological intelligence and what to do about it? *International Journal of Automation and Computing*, 14(5):532–541, 2017.

[10] Danko Nikolic. Testing the theory of practopoiesis using closed loops. In *Closed Loop Neuroscience*, pages 53–65, 2016.

[11] Léon Bottou and Vladimir Vapnik. Local learning algorithms. *Neural computation*, 4(6):888–900, 1992.

[12] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138, 2010.

[13] Tyler Cody, Niloofar Shadab, Alejandro Salado, and Peter Beling. Core and periphery as closed-system precepts for engineering general intelligence. In *International Conference on Artificial General Intelligence*, pages 209–219. Springer, 2022.

[14] W Ross Ashby. Requisite variety and its implications for the control of complex systems. In *Facets of systems science*, pages 405–417. Springer, 1991.

[15] A Wayne Wymore. *Model-based systems engineering*, volume 3. CRC press, 2018.

[16] Niloofar Shadab, Tyler Cody, Alejandro Salado, and Peter Beling. Closed systems paradigm for intelligent systems. In *2022 IEEE International Systems Conference (SysCon)*, pages 1–8. IEEE, 2022.

[17] Andreas Wagner. *Robustness and evolvability in living systems*, volume 24. Princeton university press, 2013.

[18] Jiawei Zhang. Cognitive functions of the brain: perception, attention and memory. *arXiv preprint arXiv:1907.02863*, 2019.

[19] Diane B Paul. The selection of the "survival of the fittest". *Journal of the History of Biology*, 21:411–424, 1988.

[20] F Eugene Yates. Order and complexity in dynamical systems: homeodynamics as a generalized mechanics for biology. *Mathematical and computer modelling*, 19(6-8):49–74, 1994.

[21] Walter Bradford Cannon. Homeostasis. *The wisdom of the body. Norton, Newyork*, 1932.

[22] Stefan Reimann. Homeostasis and stability. *Research Center for Interdisciplinary Studies on Structure Formation*, pages 1–13, 1996.

[23] John S Torday. Homeostasis as the mechanism of evolution. *Biology*, 4(3):573–590, 2015.

[24] David Lloyd, Miguel A Aon, Sonia Cortassa, et al. Why homeodynamics, not homeostasis? *The Scientific World Journal*, 1:133–145, 2001.

[25] Takashi Ikegami and Keisuke Suzuki. From a homeostatic to a homeodynamic self. *BioSystems*, 91(2):388–400, 2008.

[26] Jane Friedman. Suspended judgment. *Philosophical studies*, 162(2):165–181, 2013.

[27] Jane Friedman. Inquiry and belief. *Noûs*, 53(2):296–315, 2019.

[28] Hannes Leitgeb. The stability theory of belief. *Philosophical review*, 123(2):131–171, 2014.

[29] Brett Koonce and Brett Koonce. Resnet 50. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, pages 63–72, 2021.

[30] Thomas M Cover, Joy A Thomas, et al. Entropy, relative entropy and mutual information. *Elements of information theory*, 2(1):12–13, 1991.

Appendix D

From Scenario-Based to Outcome-Based Engineering in SE4AI

Niloofar Shadab, Alejandro Salado, Tyler Cody, Peter Beling

Ready to Submit to IEEE Open Journal of Systems Engineering

Abstract— As intelligent systems become increasingly integrated into our daily lives, it has become evident that there is a mismatch between traditional systems engineering (SE) practices and the nature of intelligent systems. In this paper, we posit that the current SE practices need to change their hyper-focus on scenario-based engineering to a process that is more focused on engineering outcomes in the presence of changing environment. To help enabling such transition, we incorporate closed systems precepts from systems theory into formal SE practices. The concept of closed systems is based on the idea that a system is an entity that is bounded by a physical or conceptual boundary and is isolated from its environment. Closed systems are self-contained and self-regulating, and their behavior can be predicted and controlled based on the system's internal rules and mechanisms. In this paper, we will demonstrate how closed system precepts can be applied to SE practices using a simple example of an intelligent system. We will explore different aspects of applying closed system precepts, including system's boundary definition, system modeling, and system's causal dependencies.

D.1 Introduction

The purpose of this paper is to address the current gaps in SE methods for Artificial Intelligent (AI) systems (SE4AI) [1, 2, 3], focusing on two main aspects: scope and scale. We argue that a key gap in SE4AI lies in these two aspects. While scope and scalability issues are not unique to AI-enabled systems and exist in other complex systems, the nature of these problems and their potential solutions is unique for AI-enabled systems. This is particularly due to the characteristics of intelligence property in AI-enabled systems, which can lead to unforeseen consequences if the scope and scale of the system are not properly defined and managed [1, 4].

The potential impacts of intelligence property on the outcomes of AI-enabled systems and the need to align SE practices with such impacts have been recently investigated [3, 5, 6]. The importance of this line of research amplifies as we transition from engineering "narrow AI" (weak AI) to systems with Artificial General Intelligence (AGI) or "strong AI" [2, 7, 8]. Narrow AI refers to AI-enabled systems that can perform limited tasks, while AGI or strong AI refers to systems that can successfully perform any intellectual tasks that a human being can [9]. Challenges have been identified in SE processes of AI design and deployment in both types of AI-enabled systems, appearing in all the stages of the AI system's life-cycle, from design and engineering to verification, maintenance, monitoring and updates in the later stages of the system's life-cycle [10]. We contend that a multitude of these challenges emanate from the intricate issues related to the scale and scope of AI components within the broader system architecture. These issues are a direct consequence of the inherent fragility of the property of intelligence when considered in the context of the operational environment throughout the entire life-cycle of intelligent systems. Closed systems precepts can potentially offer a significant solution to enable scalability while bounding scope for such systems in their operating environment.

In our previous research, we identified the current gaps in SE4AI using systems-theoretic fundamentals [8]. We posited that due to the high coupling between intelligent systems and their environment, utilizing closed systems precepts in SE is the path to engineering scalable and well-scoped intelligence [8]. As a theoretical foundation to this practical paper, we developed a formalism for closed systems precepts by constructing boundaries of closed systems in multiple perspectives, namely functional and informational perspectives [11]. We derived constraints to achieve two types of closure in systems: functional closure and informational closure [11]. We also demonstrated that functional closure is a special form of informational closure.

In this paper, we aim to develop simple illustrative examples that elaborate practical implications of the closed systems precepts formalism in SE practices. We demonstrate how the functional and informational closure perspectives could be beneficial in bounding the context of intelligent systems, helping systems engineers to better design and engineer the interactions between intelligent systems and their environment, and how informational and functional closure perspectives could provide values in different aspects of SE practices. To transition the closed systems formalism to tangible practical principles, the problem of scope and scale need to be addressed extensively in the context of intelligent systems.

Therefore, the paper is organized as follows: in Section II, we provide a brief theoretical overview of why traditional SE methods have limitations to capture the scope and scale for intelligent systems. In Sections III and IV, we examine the scoping and scaling challenges associated with intelligent systems in practice using simple examples. In Section V, we provide an overview of the implications of how the scope and scale problems are handled in current SE practices for intelligent systems. In Section VI, we dive into the implications of both functionally closed systems and informationally closed systems precepts using simple illustrative examples and how they can provide benefit to overcome the challenges identified in Sections III and IV. Finally, we focus on the need for understanding the causality of information that intelligent systems consume and/or produce in an informationally closed system, as well as the difficulty of designing the boundary of such systems to encompass both the engineered system and its changing environment.

In conclusion, the paper aims to contribute to the development of more robust and reliable intelligent systems that can better meet the needs of their stakeholders by addressing the fragility and sensitivity of such systems in relation to the scope of their context and the scale of their intelligence capability.

D.2 Limitations of Current Paradigms for SE4AI

The application of SE heavily relies on the decomposition and aggregation of components [12]. Each component is modeled based on its inputs and outputs to and from the environment [13]. Based on the current SE paradigms, once the boundaries of the system of interest are determined, the environment is decoupled from the system of interest and modeled through its inputs and outputs to/from the system, and it is assumed that behaviors and structures of the systems outside the boundary of the system of interest cannot be controlled by the system of interest [14]. This approach is built upon the assumption of the behavior-preserving properties of the system and its environment. While decomposition and aggregation work well for traditional engineering systems, intelligent systems are different in that they are highly interconnected with their environment [8, 15]. Due to the high coupling between intelligent systems and their environment, it is difficult to accurately model the system's behavior and its impact on the environment using traditional methods. For example, if a pilot works with an autonomous airplane, the behavior of the airplane is continuously affected by the pilot's behavior, and the pilot's behavior is affected by the autonomous airplane's behavior. Assuming the environment as an external actor is not sufficient for modeling the nature of the tight coupling for intelligent systems. We acknowledge that this high coupling is not unique to intelligent systems (some aspects of traditional systems also face similar levels of high coupling), but suggest that it is a fundamental characteristic of intelligent systems, and hence our interest in this framing.

There have been extensive research to illustrate the fragility of intelligent systems and the ramifications of this fragility, particularly in the application of autonomous vehicles [16, 17, 18]. Let us consider a straightforward example involving autonomous vehicles from [16]. These vehicles are designed to recognize and interpret speed limit signs while on the road, ensuring compliance with the law and adherence to safety protocols. However, even a minor alteration to a speed limit sign can lead to misinterpretation. For instance, a small alteration of the number ”3” in a ”35 miles per hour (mph)” speed limit sign may render it similar to ”85 mph.” Consequently, an autonomous vehicle designed to travel at approximately 35 mph might erroneously operate at 85 mph. This slight alteration in the context of an autonomous vehicle has far-reaching implications, resulting in the violation of safety regulations and protocols. It is essential to grasp that increasing the number of scenarios in use-case models or augmenting training data does not rectify the inherent fragility of these systems, since they would have to account for every minimal variation.

This example underscores a fundamental distinction between traditional SE practices, where the engagement between the system of interest and its environment (including external systems) is reduced to a number of scenarios, and the endeavor to engineer scalable intelligence within the specific context of an intelligent system. Current approaches to address this challenge involve embedding intelligent systems within their operational context and allowing them to operate and learn in that environment [19, 20]. This approach stands in contrast to the conventional practice of designing, engineering, and testing systems in controlled laboratory settings and then expecting them to operate identically once deployed in their operational environment [21]. However, a critical issue persists with this approach: while intelligent systems are deployed in their operational environment well before they reach the market, systems engineers lack a well-defined process for evaluating the confidence they can place in these systems based on observations of their interactions with the environment. Moreover, there is construct on when the scope of the context included in the on-road testing will suffice to scale the capability.

In conclusion, the significant interdependence between the system and its context underscores the challenge of determining the extent of the environment that should be incorporated during the engineering process to capture these interconnections effectively. Current strategies for deploying and gathering training data to address this interdependence face two primary drawbacks. Firstly, not all types of intelligent systems can be readily deployed in their operational environment and tested in the developmental phase. Secondly, even if early deployment and data collection are feasible, there is a lack of clear methods to ensure that the deployed context is sufficient for achieving scaled and robust intelligence capabilities. To elaborate on the second drawback, the problem of **scope** must be addressed first.

For instance, a challenge in the adoption of autonomous vehicles pertains to characterizing their capabilities, thus enabling the enactment of appropriate regulations to govern their operation [22, 23]. A prominent example is the Automatic Cruise Control (ACC) technology, a pivotal feature in autonomous vehicles, which exhibits distinct functionalities across various use cases. These high-level use cases encompass capabilities such as following a leading vehicle, managing cut-in and cut-through, responding to emergency vehicles, and maintaining a consistent cruise speed, among others. Each of these use cases encompasses a multitude of scenarios influenced by environmental factors. Determining which scenarios are pertinent for integrating the ACC capability into the system's intelligence property gives rise to a scoping challenge.

Another main challenge facing intelligent systems is *scalable intelligence*, which refers to the ability to increase intelligence capabilities in a way that is efficient and effective in a variety of contexts. Ongoing research is exploring various approaches to engineer intelligence that can be scaled [24, 25, 26]. One approach involves formalizing trust between AI-enabled systems and humans to scale human interactions with AI-enabled systems [27]. However, trust is just one outcome of an intelligent system and may not be scaled to other potential outcomes such as safety or survivability. The free energy principle is another approach to address scalable (general) intelligence [28]. This principle provides a solution to bound intelligent systems into a set of stable states. However, it does not provide a concrete answer to how engineers can select a set of states and restrict such systems to that particular set. This method also has been subject to criticisms regarding its mathematical consistency [29, 30]. Within the AI research community, the issue of scalability has received considerable attention, particularly regarding the expansion of optimization algorithms within intelligent systems [31, 32, 33]. Nonetheless, it is noteworthy that these methods may or may not guarantee scalability in terms of system outcomes. This is due to the fact that the primary emphasis of such research pertains to the AI component itself, with less immediate consideration for the **scalability** of overall outcomes or behaviors of the system that consists of this AI component. As mentioned earlier, the open-system view (input-output paradigm) in SE practices also fails to characterize the unique features of intelligence property, particularly scalability [8].

So far we posit that the current SE practices need to change their hyper-focus on scenario-based engineering to a process that is more focused on engineering outcomes in the presence of changing environment. To enable this transition of scenario-specific capability to a scaled capability that is outcome-specific, traditional SE, although necessary, may not be sufficient or at best efficient. In the subsequent sections, we will delve deeper into these two issues concerning scale and scope, using the Autonomous Vehicle's ACC capability as an illustrative example.

D.3 Scaling Problem

The challenge of adapting system's capabilities to address a wider array of environmental conditions is fundamentally a matter of **scalability**.

Current SE methods typically rely on scaling systems by increasing the size of their inputs and outputs sets. For example, this might involve raising the number of queries processed per second, expanding the volume of images used for training, or increasing the categories in classification applications, among other such approaches. However, this conventional scaling paradigm encounters limitations in the context of intelligent systems, where constant learning is intrinsic.

In intelligent systems, and by extension, any constituent intelligent component, a perpetual state of learning prevails. This learning aspect is pivotal as every activity undertaken by an intelligent system serves as an opportunity for further learning and adaptation. Consequently, scaling the intelligence property of such systems extends beyond merely expanding inputs and outputs to scaling the outcome of intelligence property in such systems. This phenomenon is especially pronounced in intelligent systems characterized by hierarchical decision-making processes. In such cases, the system is exposed to a multitude of uncontrolled learning events, and the inherent uncertainty regarding the system's true state combines with that of the other integrated systems.

It is common to model the high-level behavior of a system using states. Different sets of behaviors are encapsulated as states and the system may transition between states both endogenously and exogenously [34]. The behavior of behavior-preserving systems can be modeled over a set of finite states that have clear boundaries and can transition from one state to another if they receive proper inputs. Even though the transitions between the states in the set are not deterministic or the environment is only partially observable, the state space of the system remains unchanged.

In dealing with intelligent systems, however, the set of states continuously evolves, some states will be added to the set over the life cycle of the system and some others might be removed from the set as the system cannot get back to those states as a direct result of the learning characteristics of the intelligent systems. In other words, the intelligence property is not only dependant on the system's set of states but also the high coupling between the system's states and the changes in the environment. Yet, SE practices have hyper focus on engineering intelligence through a set of use cases that rely on predefined system's states without considering the paths the system might have taken to achieve these states through its coupling with its environment.

In this context, breaking down intelligence-related capabilities into finer, more detailed components to discern system states and actions may not suffice to achieve the realization of intelligence capabilities at the system level [35]. Intelligent systems are exceptionally sensitive to changes in their environment, and this sensitivity cannot be mitigated by simply increasing the number of use-case models. The addition of more use-cases and/or their refinement fail to alleviate the issues arising from context-induced changes and, in fact, contributes to the unpredictability of intelligent system outcomes.

For instance, in the autonomous vehicle example, the number of routes that the autonomous vehicle can drive or the number of red-light scenarios that the autonomous vehicle handles safely, are considered as scaling intelligence through scaling inputs and outputs of the intelligent system. However, in the same system, scaling a general outcome of "Collision

Avoidance" cannot be necessarily achieved through scaling inputs and outputs; as collision avoidance is an outcome that should be attained in various contexts and in various scenarios at the higher-level in relation with other actors and has direct relation with the fragility of the autonomous vehicles in such scenarios. Such a scale of outcomes cannot be realized solely by adding more scenarios to test or design (scaling inputs-outputs) but by directly scaling outcomes. At some point, scaling the outcomes of an intelligent system relies predominantly on closed relations between the intelligent system and its environment. If we go back to the speed limit example that we provided earlier in this paper, scaling outcome would mean that the autonomous vehicle, although detecting speed limit as 85 mph instead of 35 mph, will be able to drive in a safe manner by evaluating the entire context and other actors and overwritting a probable mistake in its detection. Being able to drive safely would be an outcome that is scalable even if the vehicle misreads the speed limit sign.

It becomes apparent that current SE practices lack a straightforward mechanism for the engineering of scaled intelligence capabilities, particularly in addressing the inherent fragility of these systems. Moreover, a specific and clear principle for ascertaining and confirming scalability across diverse levels of abstraction within intelligent systems remains conspicuously absent.

We suggest that, to attain scalable intelligent capabilities, transitioning from scenario-based engineering to outcome-based engineering is worth exploring. Addressing this transition necessitates a redefinition of the application of current SE practices. This redefinition should encompass the recognition of the high degree of coupling between environmental behaviors and system behaviors, as both are equally indispensable for the emergence of system-level intelligence capabilities.

D.4 Scoping Problem

The scoping process plays a pivotal role in the problem definition phase when designing a system [36]. It often serves as the primary step in comprehending the problem at hand, as well as establishing the essential constraints for the design phase. Typical scoping procedures involve the creation of a context diagram, the development of operational definitions diagrams, and the subsequent specification of use cases.

In SE practices, the scope must strike a balance between being comprehensive enough to encompass all relevant external stimuli and being small enough to enable the system's realization in a cost-effective and timely manner. This process involves capturing all external systems that may affect and/or being affected by the system's input-output relations. Consequently, the scope is defined based on the desired set of system's inputs-outputs within their respective environment.

Scoping the environment in intelligent systems presents unique challenges due to the emergence of the system's behaviors, particularly in terms of functional input/output relations, in relation to its environment [37]. In the context of SE4AI, there currently exists no established framework or guiding principle for determining the extent to which the environment should be considered to realize intelligent capabilities. To model scope using input-output relations, one needs to capture an instance of a pre-condition in the environment to establish a clear system boundary. For example, ACC capability may depend on weather conditions (rainy, snowy, sunny, etc.) in the environment. Each scenario must be modeled under specific pre-defined environmental conditions. This process increases the number of scenarios required to engineer the scope of the ACC capability and yet it does not directly focus on the scope of the relational property.

In intelligent systems, the dissolution of boundaries between the system and its environment in the systems modeling processes results in capabilities derived from the system's intelligence being dependent on the relational property between the system and its environment. This relational property serves as an indicator of the scope of such capabilities. However, current SE methods tend to enforce a clear boundary around the system, leading to scoping the system solely through input-output interactions with the environment [38].

There are situations that need to be modeled with more than one use case in the autonomous vehicle's environment. There are situations where these defined use case models dissolve boundaries. For instance, a "cut-in" scenario can be considered as "follow vehicle" scenario and vice versa. The dissolution of the boundaries of the use-case in the systems modeling processes is another reason of why these SE techniques might not be sufficient for modeling intelligence capability.

As a direct consequence of the learning aspects, intelligent systems have fragile behaviors when facing different sets of inputs; and this is the cause for the need of way many more predefined sets than a behavior-preserving system may need. When combined with the problem of scale, it becomes evident that current SE methods are not practical to address the fragility of intelligent systems for most forms of intelligence capabilities.

Therefore, engineering intelligent systems through a set of inputs and outputs relations faces a significant challenge when it comes to scoping problem. There are two aspects to this challenge. First, in engineering intelligent systems, systems engineers need to capture many boundary conditions that is a direct result of deriving an outcome-based problem through an inputs-outputs-based problem. Second, decomposing intelligence capabilities into smaller, more detailed capabilities would result in derivation of system's outcome through aggregation of smaller outcomes which brings similar problems as the first aspect, and it is a direct result of decomposition and aggregation through input-output relations. All in all, scoping problems when it comes to intelligent systems require paradigms that are complimentary to the input-output relations.

Effectively addressing the scoping challenges in intelligent systems requires capturing the closed coupling between the boundary of environmental conditions and future system states.

D.5 Scale and Scope with Traditional SE Practices

As previously discussed in this paper, we asserted that deriving outcome-based capabilities directly from input-output relationships is not a viable approach for intelligent systems. This raises the question of how we can, in fact, derive outcomes in such systems without relying on this input-output paradigm.

To gain a more comprehensive understanding of how current SE practices address the challenges of scale and scope, we will delve into the examination of existing SE methodologies, specifically through the lens of input-output relations concerning ACC capabilities.

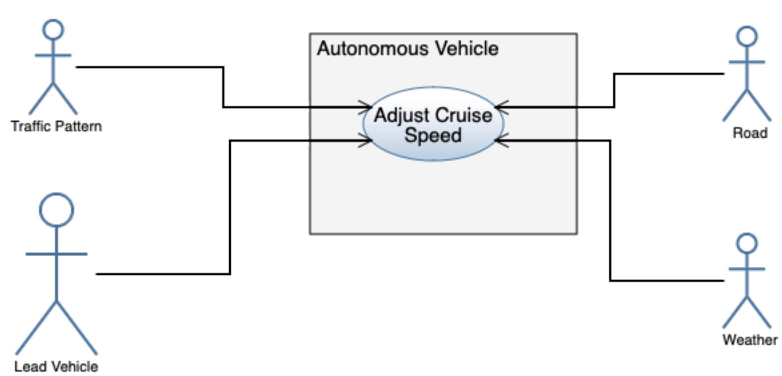


Figure D.1: Block diagram of the autonomous vehicle (system of interest) utilized in realization of ACC capability in the current SE practices. In this diagram, all the external systems that have association with this capability are captured.

The conventional system boundary within the framework of SE is typically represented by the input-output relations with its external systems (ref. Figure D.1 for an example of an autonomous vehicle limited to those relationships of interest to the ACC capability). To elucidate this challenge, consider Figures D.2 and D.3. In these illustrations, the dynamic interactions of the traffic pattern, which encompasses other roadway users, become abstracted out in the form of input-output to/from the system of interest. One might argue that it is possible to break down traffic pattern into more detailed actors and external systems and capture their behaviors through separate input sets to the autonomous vehicle. The problem with this approach is that it leads to scalability and scoping issues that were discussed previously while making the system more fragile by adding more interfaces. For instance, traffic pattern can be broken down into cars, trailers, emergency vehicles, police, bicycles, motorcycles and pedestrians. And each one of those have further variations: color, shape, plate number, cleanliness... The autonomous vehicle needs to exhibit different behaviors based on the combination of such systems in its environment, while ignoring certain variations between such systems that should be irrelevant for some driving capabilities. Pull over scenarios by police are significantly different from scenarios to give the right of way to an ambulance. These two scenarios are handled through two separate use cases. However, in scenarios that include both ambulance and police cars, separate use cases interface with each other, creating a power set of the number of use cases that the autonomous vehicle will face. As depicted in Figure D.3, the traffic patterns provide essential information related to speed and distance of other roadway users. However, it is not possible to come up with the requisite abstraction levels for systematically modeling the traffic pattern as an entirety without breaking it down to an input-output problem. Capturing outcomes through input vectors from the system's environment is the main relation to the fragile nature of the intelligent systems. The governance of traffic pattern as an entity would have significant effect on modeling the nature of ACC capability in the autonomous vehicle.

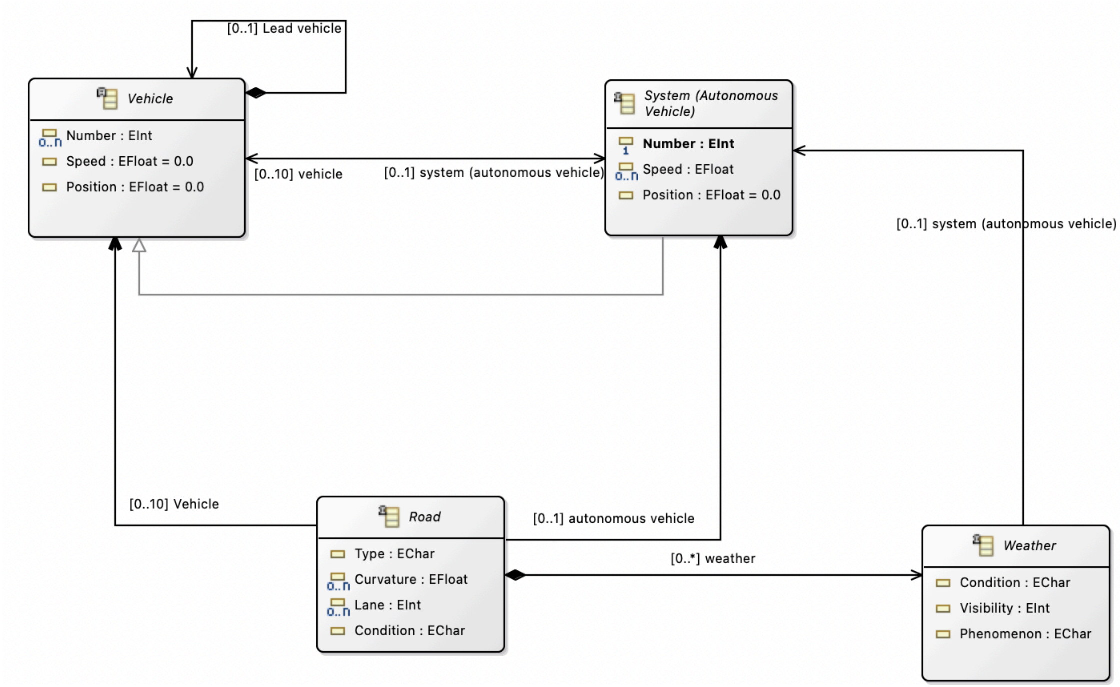


Figure D.2: Class diagram in open SE that depicts the relations between the autonomous vehicle (system of interest) and the external systems.

Having highlighted the limitations of modeling an outcome-based problem through an input-output framework, we propose that engineers often need to utilize multiple models to cover various scenarios that lead to the desired outputs for intelligent capabilities. The specific desired outputs depend on the system's behavior, which, in turn, is influenced by its context. For instance, the response of an autonomous vehicle in a scenario with only the lead vehicle on the road would differ from its behavior in a scenario with high traffic density. We posit that closed systems precepts enable systems engineers to find new approaches in SE practices that allow direct engineering of the outcome-based problems which can circumvent the need to manage numerous separate use-case instances. In the next section, we delve into the potential advantages of employing closed systems precepts in providing this missing layer of abstraction.

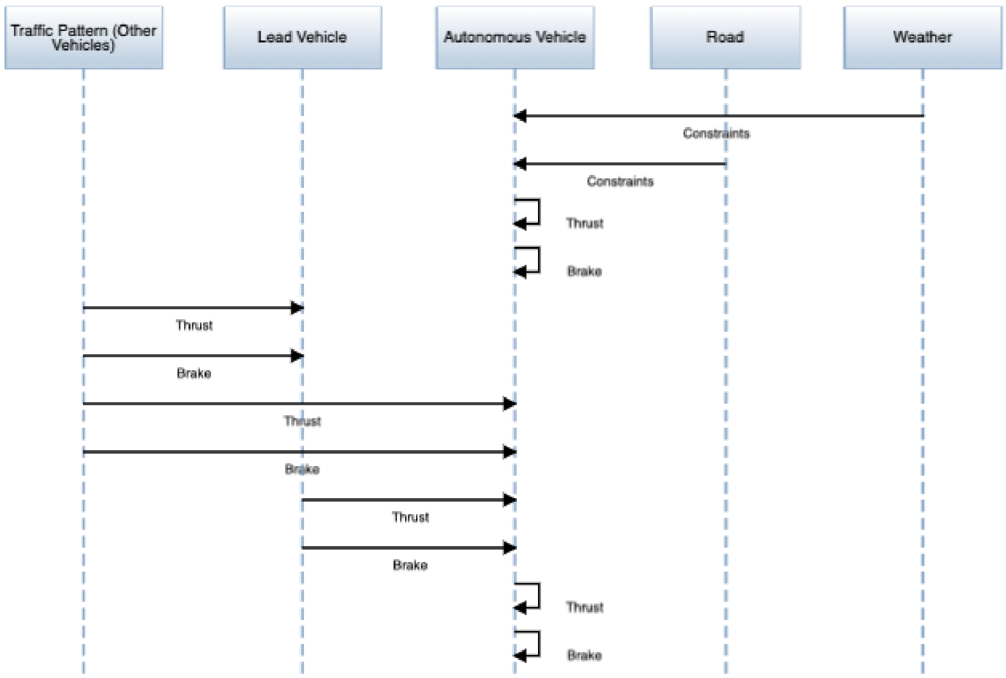


Figure D.3: Sequence diagram in open SE that models the signals that are being transferred between the autonomous vehicle (system of interest) and the external systems.

D.6 Application of Closed SE

In our previous discussion, we highlighted the limitations of open systems precepts in SE of intelligent systems using the autonomous vehicle example. We then emphasized the importance of capturing the closed relations between the system and its environment. In this section, we will elaborate how the closed systems precepts that we formalized in our previous work can be utilized to enable direct engineering of outcome-based problems in intelligent systems.

D.6.1 Functional Closure Precept

In our previous work, to formalize closure, we considered part of the environment and the system of interest within the boundary of a context system to accommodate the closed relations between the system of interest and its environment. Therefore, different types of closure can be achieved for the context system that consists of the intelligent system and part of its environment. Based on this convention, we defined functionally closed system as follows [11]:

Definition D.1 (Functionally Closed System). A context system, S^C is functionally closed from its outer environment, E^O, if and only if,

1) There exists a minimal set of inputs and outputs, , M, such that S^C is functionally dependent on M. This condition can be shown as: $S^C \subseteq \times\{\mathcal{X}_M, \mathcal{Y}_M\}$, and

2) There are no additional inputs from E^O beyond M that can influence the behavior of S^C. and

3) There are no additional outputs from S^C beyond M that can affect the behavior of S^C.

Where S^C is a context system that is aimed to be conceived as functionally closed, E^O is the environment outside the context system, M is a minimal set of inputs; X_M, and outputs Y_M. This definition of a functionally closed system is a relaxation of the systems-theoretic definition of a closed system [39]. It allows for the system to have interactions with the environment, but these interactions must not affect the behavior of the system or its outputs, and the outputs of the system do not affect the behavior of the environment [11].

Functional closure assumptions are informally utilized in various engineering applications. For instance, in the implementation of the ACC capability, road vibrations are often disregarded when designing functions for ACC capability. This is because the vibration effects can be considered negligible for this particular functionality, and similarly, the impact of the

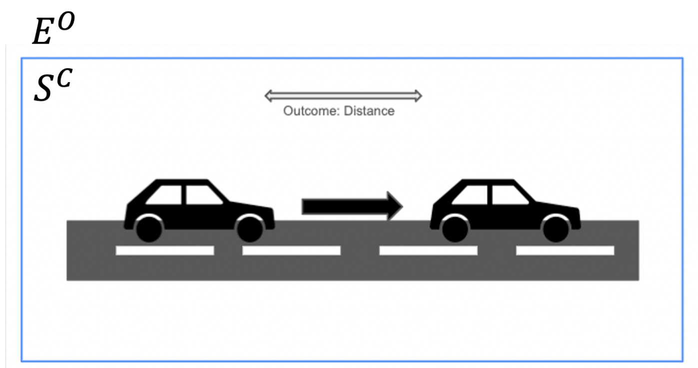


Figure D.4: Application of functional closure to engineer the ”cruise speed adjustment” functionality in the presence of a lead vehicle

function's output (movement on the road) on road vibrations can be overlooked. To illustrate this concept, we present a straightforward scenario for the ACC capability, as depicted in Figure D.4. In this representation, we focus on the follow lead vehicle sub-capability of ACC that can be developed using the principles of functional closure.

To engineer such a capability, we consider the system of interest (the autonomous vehicle that is following the lead vehicle), the lead vehicle, and the road surface within the functionally closed system boundary. As seen in Figure D.4, there are no inputs or outputs to and from the closed context system, delineated by the solid blue boundary. The underlying assumption is that any external inputs, such as those related to weather, visibility conditions, road vibrations, or interactions with other road users, do not significantly affect the vehicle's output. Similarly, the vehicle's capability does not affect the outer environment.

To derive functional requirements and predict the next state of the system illustrated in Figure D.4, no information from the environment is necessary, as long as we regard the distance between the two vehicles of interest within the boundary of the closed system as the desired outcome. The modeling assumption of abstracting out all other information transmitting between the context system and its environment, denoted as E^O, empowers us to engineer the outcome of the closed system through the execution of the set of functions by the two vehicles over the system's life cycle. This modeling assumption enables us to engineer distance as a scalable outcome in any scenario, as long as we consider the lead vehicle within the boundary of the functionally closed system. Formally, we can represent the execution of the functions of the system within the blue boundary in Figure D.4 as follows[1]:

$$
\begin{aligned}
&O_{FCS} = \text{Outcome of The Functionally Closed System}\\
&F_{FCS} \subseteq \times\{F_{Following}, F_{Lead}\}\\
&S^C \subseteq F_{FCS}\\
&F_{FCS} : \mathcal{X}_M \rightarrow \mathcal{Y}_M\\
&\text{Where:}\quad \mathcal{X}_M \subseteq M \quad \mathcal{Y}_M \subseteq M\\
&X_M \subseteq \{V_L, a_L, J_L, F_{fric}\}\\
&Y_M \subseteq \{V_F, a_F, J_F\}\\
&X = X_E \setminus M \text{ and } Y = Y_E \setminus M\\
&\text{ We have: } S^C \not\subseteq \times\{\mathcal{X}, \mathcal{Y}\}\\
&\Rightarrow O_{FCS} : M \times F_{FCS} \rightarrow \mathcal{D}
\end{aligned}
\tag{D.1}
$$

Where F_{FCS} is the set of functions to be executed within the functionally closed system, $F_{Following}, F_{Lead}$ are the actuation functions of the following vehicle and lead vehicle, respectively. M is the minimal set of inputs and outputs, X_E, Y_E are the inputs and outputs sets of the closed system's environment, F_{fric} represents the friction force of the road surface that enables the vehicles to move. $\mathcal{D}$ is the outcome of the functionally closed system.

This modeling approach offers several benefits to systems engineers. Firstly, it allows them to initially focus on the static properties of an outcome without being overwhelmed by

[1]The representation of the functional closure in Equation D.1 is for demonstration only

the dynamic effects of other inputs from the environment. For example, wind speed may have a considerable impact on the speed, acceleration, and jerk of both vehicles at each time stamp, but from a functional perspective, it does not play a role in the execution of actuation functions. Secondly, this abstraction level provides a benchmark for understanding what type of information is needed to evaluate and calculate the parameters of the functions executed within the functionally closed system. Distance, for instance, is an outcome achieved through the minimal set of inputs and outputs defined in Equation D.1. The next level of abstraction in the modeling process would provide the information that affects the values of the elements in the minimal set of inputs and outputs. Such information is required to be gathered to calculate such elements in the minimal set. Thirdly, it identifies those external systems that have functional relations with the system of interest. In Figure D.4, the road surface and lead vehicle exhibit this characteristic. Lastly, this modeling process can be conducted without the need to break down the system to any scenario-based problems at this level of abstraction. For example, there is no need to break down the outcome into logical scenarios with different distance values (such as comfort distance, emergency distance, safe distance, etc.).

In conclusion, functional closure criteria can be utilized to derive the follow lead vehicle capability for ACC. These criteria offer a foundation for systems engineers to establish the boundaries of the environment around the autonomous vehicle in order to derive distance as the outcome of the follow lead vehicle capability. In general, functional closure precept can help systems engineers to scale the functional outcomes of a system-level capability by identifying and working with the minimal set of inputs and outputs while ignoring those environmental factors that do not have any effects on the capability from the functional perspective.

D.6.2 Informational Closure Precept

In addition to functional closure, SE can employ informational closure for its practices at the next level of abstraction. An informationally closed system can be achieved when there is no *new* information exchanged between the context system S^C and its environment E^O [40]. In informationally closed systems, information can be transmitted through the closed system's boundary in the form of "mutual information" [11]. Therefore, functionally closed systems are a special case of informational closure where this "mutual information" between the closed system and its environment will be zero. As a result, a systems-theoretic definition of informationally closed systems can be described as follows [11]:

Definition D.2 (Interpretation of An Informationally Closed Systems Using Information). A Context System that transitions through states $1, 2, ..., n, n+1$; is informationally closed if there is no joint information between S^C_{n+1} and $E^O_n|S^C_n$.

$$I(S^C_{n+1}; E^O_n|S^C_n) = 0$$

We also derived a theorem to calculate the minimum level of mutual information required between the system and its external environment to achieve closure [11]:

Theorem D.3 (Inequality for mutual information in closure).

$$I(S^C_n; E^O_n) \geq H(S^C_{n+1}, S^C_n) - H(S^C_{n+1}, S^C_n|E^O_n)$$

Theorem D.3 provides the relation for the level of mutual information being presented in the boundary of an informationally closed system. To maintain closure at state n, the output of the closed system to the environment at state n (i.e., the amount of information transmitted from the system to its environment) should follow Theorem D.3.

Informational closure, adds a dynamic nature to functional closure where the system can be considered closed relative to its set of states and the outer environment set of states. In simpler terms, informational closure pertains to the relationship between the current state (n) of the system and its next state ($n+1$), considering the state of its environment at the current time (n). Sustaining informational closure involves maintaining a specific level of mutual information between the system and its environment, as outlined in Theorem D.3. Informational closure implies that an informationally closed system is not derived from the constraints on inputs set. It is rather derived from the ability to predict or expect such an inputs set from its environment, E_n^O.

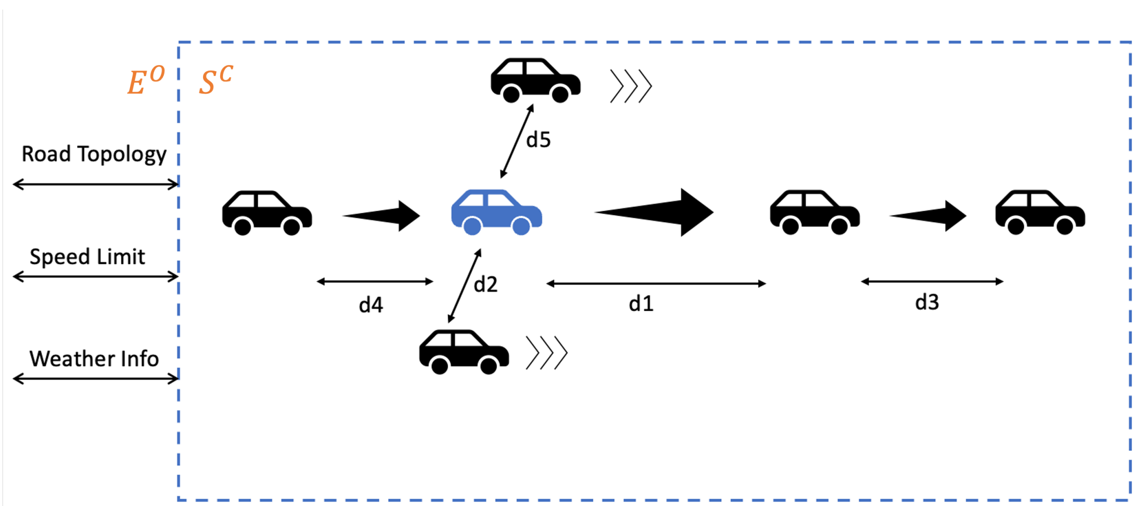


Figure D.5: Application of informational closure to scope the ACC capability in its operational environment

Similar to the concept of functional closure, we will now delve into informational closure, using a simple scenario example to illustrate its application in the context of the ACC capability. In Figure D.5, the system of interest, depicted as the blue vehicle, is tasked with demonstrating ACC capability in its driving context. However, what sets the scenario in Figure D.5 apart from the one in Figure D.4 is the inclusion of additional systems within the closed system boundary, indicated by the blue dashed line. In this context, the system of interest requires supplementary information to evaluate and quantify the parameters in the minimal set defined in Equation D.1.

In certain situations, contingent on the traffic dynamics and driving context, the autonomous vehicle must be capable of adjusting its cruise speed to accommodate various ACC use cases, such as handling cut-in and cut-through use cases. The advantage of employing this principle is that systems engineers need not concern themselves with the dissolution of use case models, such as transitioning from a cut-in use case to a follow lead vehicle use case, as long as they incorporate the principles of informational closure. In informational closure, the ACC capability is designed in its entirety, rather than through the aggregation of individual use case models.

According to the definition of informational closure, it is imperative to encompass enough of the environment within the closed system's boundary to establish the required level of mutual information between the closed system and its outer environment. As illustrated in Figure D.5, the exchange of information through the closed system's boundary, which includes elements like Road Topology, Speed Limit, Weather Information, and so forth, is shared between the closed system and its environment. This type of information is characterized as mutual information between the closed system and its environment in its current state. The process of selecting mutual information is usually based on the modeling assumptions of what is inside the closed system and what could be shared with the outer environment. Speed limit sign, for example, can fall either inside the boundary in Figure D.5 or outside of it. The important assumption is, either way, the system has enough knowledge of the speed limit no matter where the speed limit sign is located at each state n.

Informational closure plays a pivotal role in addressing the scoping challenges of intelligent systems by imposing constraints on the amount of information required from the environment to construct an informationally closed system capable of functioning across various states of its external environment, denoted as E^O, as long as it maintains informational closure at each system state. To formally outline the process of determining how many vehicles in the environment should be encompassed within the closed context system, denoted as S^C (as depicted in Figure D.5), we propose that we follow Theorem D.3 and determine joint and conditional information and entropy of the context system with respect to its outer environment.

To have an informationally closed context system, S^C, depicted in Figure D.5, the inequality $I(S_n^C; E_n^O) \geq H(S_n^C, S_{n+1}^C) - H(S_{n+1}^C, S_n^C | E_n^O)$ needs to be satisfied. The entropy of the context system should be reduced by reducing uncertainties. The reduction of uncertainties only occurs with increasing the level of information. The level of information depends on where we put the boundary of the context system. In Equation D.2, we show how the terms in Theorem D.3 can be calculated. This process helps systems engineer to lock down a desired context with all the necessary information they need to predict the next state of the system with confidence. The process of quantifying Theorem D.3 is based on adding more information inside the context to both satisfies the inequality and be cost-effective in terms of how much uncertainty we accept to tolerate for the system.

Initial Modelling Assumptions:

$$I(S_n^C; E_n^O) = \{V_{limit}, R^0, W^0\}$$

$$I(E_n^O) = \{V_{limit}, R^0, W^0, RU\}$$

Calculation of Entropy in The Proposed Model:

$$H(S_n^C) = -\sum_{x \in S_n^C} P(x) \log P(x) = -\sum_{d_n^1}^{d_n^5} p(d_n^i) \log(d_n^i)$$

where: $p(d_n^i) = p(V_n^i, a_n^i, V_n^0, a_n^0)$

$$H(S_{n+1}^C) = -\sum_{y \in S_{n+1}^C} P(y) \log P(y)$$

$$= -\sum_{d_{n+1}^1}^{d_{n+1}^5} p(d_{n+1}^i) \log(d_{n+1}^i)$$

where:

$$p(d_{n+1}^i) = p((V_{n+1}^i, a_{n+1}^i | V_n^i, a_n^i), (V_{n+1}^0, a_{n+1}^0 | a_n^0, V_n^0))$$

Final Calculation of Inequality:

$$\begin{aligned}
&\text{First: } I(S_n^C, E_n^0) = H(S_n^C) - H(S_n^C | E_n^0) \\
&\text{Second: } H(S_n^C, S_{n+1}^C) = H(S_n^C) + H(S_{n+1}^C | S_n^C) \\
&\text{Third: } H(S_{n+1}^C, S_n^C | E_n^O) = \\
&\qquad H(S_{n+1}^C) + H(S_n^C | E_n^O) - I(S_{n+1}^C; S_n^C | E_n^O)
\end{aligned} \tag{D.2}$$

Where $I(S_n^C; E_n^O)$ is mutual information between the outer environment and the closed context system at state n, $H(S_n^C)$ is entropy of the context system at state n, $H(S_{n+1}^C)$ is entropy of the context system at state $n+1$, V_{limit}, R^0, W^0, RU are speed limit, road topology, and weather info, and other road users[2] respectively. d^i is the distance between the ego vehicle (blue car) and i^{th} vehicle, V^0, a^0 are ego vehicle's velocity and acceleration, and V^i, a^i are i^{th} vehicle's velocity and acceleration.

The number of vehicles depicted in Figure D.5 is purely for illustrative purposes. In practical engineering applications, this quantity may fluctuate in accordance with the probability distribution (entropy) characterizing the current and subsequent states of the context system and the level of acceptable mutual information that is both feasible and cost-effective. The objective behind ascertaining the extent to which the environment should be encompassed within the closed context system is to define the scope of the environment, thereby ensuring the scalability of the intelligence property within the given context. In this evaluation process, systems engineers must assess the degree of mutual information between the context system and its external environment, aiming for a level that sufficiently satisfies the inequality $I(S_n^C; E_n^O) \geq H(S_n^C, S_{n+1}^C) - H(S_{n+1}^C, S_n^C | E_n^O)$. Attaining a desired level of mutual information hinges directly on the decision regarding the extent to which the inner environment is encompassed within the context system. One example of such process was demonstrated by Equation D.2. At the end, all the elements inside the boundary are counted to quantify the First, Second, and Third terms in Equation D.2.

In this section, we have explored how informational and functional closure can serve distinct purposes within SE practices, effectively addressing scoping and scaling issues in the context of intelligent systems. Our examples have demonstrated that the principles of closed systems

[2]Other road users refers to all road users outside of what is already within the boundary of the context system. For example, in Figure D.5, all other vehicles on the road except the six vehicles depicted in the picture fall into the category of "other road users".

precepts offer systems engineers a valuable framework for devising methodologies to define the boundaries around the system of interest and its operational context. We have delved into the concept of the desired level of mutual information between the closed system and its environment. However, we have yet to discuss how this mutual information can be interpreted in the context of an engineered system. This matter will be addressed in Subsection D.6.3, where we will delve into the causality of mutual information, shedding light on its interpretability within the framework of informational closure.

D.6.3 Causality of Mutual Information

The subsequent step in integrating informationally closed systems precepts into SE practices for intelligent systems involves the interpretation of mutual information. As previously discussed in this section, mutual information embodies the kind of information shared between the informationally closed system and its environment. However, to leverage this information for engineering purposes, it is crucial to thoroughly examine and understand its origin, causality, and its engineering implications. The causality provides engineers with insights for making alternative design decisions, and ultimately, it becomes a decisive factor in determining how mutual information is acquired by the closed system. Table D.1 provides a concise overview of various interpretations of mutual information.

Within the environment of the informationally closed system, two categories of information exist. The first category is relevant information, essential for the closed system to foresee its future states. The second category is irrelevant information, which does not contribute to predicting the current and future states of the closed system and can be disregarded.

Mutual information is in the category of relevant information to the closed system that is shared with the environment; we can mathematically show this information as $I(S_n; E_n)$.

To engineer or design the relevant information mentioned above, it is crucial to identify the causality and source of the information between the closed system and its environment.

As delineated in Table D.1, relevant information can be sub-categorized as confirmation or redundant information. These classifications establish a causal connection between the environment and the closed system, facilitating the engineering of mutual information across the system's boundary. Confirmation information serves to verify existing data within the closed system and can originate from either the closed system itself or external sources in its environment.

For instance, an autonomous vehicle may validate the weather temperature by measuring it in the environment and cross-referencing it with information from an external source, such as the National Oceanic and Atmospheric Administration (NOAA). In contrast, redundant information pertains to data that is superfluous for the closed system. An example of redundant information would be road geometry that the autonomous vehicle is already aware of and is also present in the environment.

To illustrate these categories further, let's consider the statement: "it is sunny outside." For an observer inside a room, this information can be interpreted as redundant if they can directly see the sun through a window. In this case, the observer already has the information within their closed system. Redundant information is a form of mutual information.

Now, let's imagine a scenario where a weather channel in radio predicts that it will rain in two hours, and the observer can already see the formation of rain clouds. In this case, the information from the radio acts as confirmation, confirming what the observer already anticipates. Confirmation information is another form of mutual information between the closed system and its environment.

Information Type	Interpretation	Causal Dependency
Relevant Information	Type of information that is required to predict the future state of the closed system and is shared between the current state of the closed system and its environment.	**Confirmation**: Type of relevant information that confirms the information existing inside of the closed system. **Redundancy**: Type of relevant information that adds redundancy to the information that already exists inside of the closed system.
Irrelevant Information	Type of information that exists in the environment of the closed system and is irrelevant to the prediction of the next state of the closed system and needs to be **Ignored** by the current state of the closed system.	**Unnecessary**: Type of irrelevant information in the environment that is not necessary in computing entropy or information of the next state of the closed system. **Unavailable**: Type of irrelevant information in the environment that is not available to the current state of the closed system and cannot be available through the boundaries of the closed system.

Table D.1: Types of Information Relative to The Closed System's Boundary

Additionally, there is another type of information that we call "Irrelevant Information". This information is irrelevant to the context of the closed system and adds no value for predicting the closed system's next state. Therefore, it would be *ignored* by the closed system. All information outside the boundary of the informationally closed system that is not part of the mutual information falls into the category of irrelevant information. In other words, it is the difference between the total environmental information and the mutual information: $I(E_n) - I(E_n; S_n)$.

By appropriately defining the boundary of the closed system, all the information in the outer environment becomes irrelevant, with $I(S_{n+1}; E_n|S_n) = 0$, indicating no impact on the closed system's predictions of its next state; S_{n+1}. Irrelevant information can also be categorized into different types. In this paper we identified two types; unnecessary and unavailable. If we go back to the "it is sunny outside" example, we can conclude that information of day or night time is unnecessary as "it is sunny" in and of itself indicates daytime. Similarly, information about the number of sunny days in the next month is not available and can be considered as irrelevant information. Identification of irrelevant information reduces noise in stakeholder needs, and systems requirement derivation. Such information can also be useful in system design choices.

To evaluate the mutual information of an informationally closed system, systems engineers need to examine both relevant and irrelevant information. Relevant information can be identified by analyzing confirmation and redundant information exchanged between the closed system and its environment. In the context of the autonomous vehicle scenario depicted in Figure D.5, the classification of speed limit information can vary based on the design choices and stakeholder requirements. If the autonomous vehicle's navigation system already contains pre-stored speed limit information based on location data, then the incoming speed limit information from the environment can be considered redundant information. This is because the information is already available within the closed system.

On the other hand, if the speed limit information does not pre-exist in the autonomous vehicle's navigation system, the sensor system of the vehicle detects the speed limit in real-time. In this case, when the dispatcher transmits new speed limit information to the autonomous vehicle, it serves as confirmation information to validate and confirm the speed limit detected by the vehicle's sensors. This confirmation information completes the existing knowledge within the closed system.

As illustrated in this simple example, the categorization and causality of mutual information can influence and be influenced by design preferences as well as the specific needs of the system's stakeholders. Systems engineers need to consider these factors when determining and/or engineering level of mutual information within an informationally closed system.

D.7 Conclusion

In this paper, we have elucidated the scoping and scaling challenges that arise when addressing outcome-based issues within the framework of current SE practices for intelligent systems. Subsequently, we have presented a comprehensive delineation of how different forms of closure, including informational and functional closure, can be effectively applied. These approaches empower systems engineers to devise SE methodologies that directly tackle outcome-based problems, obviating the need for decomposition and aggregation through input-output relationships. Our proposition is that this approach will assist systems engineers in surmounting the hurdles associated with engineering scalable intelligence while ensuring its precise scoping within the operational context.

Bibliography

[1] Thomas A McDermott, Mark R Blackburn, and Peter A Beling. Artificial intelligence and future of systems engineering. In *Systems Engineering and Artificial Intelligence*, pages 47–59. Springer, 2021.

[2] James Llinas, Hesham Fouad, and Ranjeev Mittu. Systems engineering for artificial intelligence-based systems: A review in time. *Systems Engineering and Artificial Intelligence*, pages 93–113, 2021.

[3] Niloofar Shadab, Aditya U Kulkarni, and Alejandro Salado. Shifting paradigms in verification and validation of ai-enabled systems: A systems-theoretic perspective. In *Systems Engineering and Artificial Intelligence*, pages 363–378. Springer, 2021.

[4] William D MILLER. The future of systems engineering: Realizing the systems engineering vision 2035. 2022.

[5] Tom McDermott, Dan DeLaurentis, Peter Beling, Mark Blackburn, and Mary Bone. Ai4se and se4ai: A research roadmap. *Insight*, 23(1):8–14, 2020.

[6] Kara Pepe and Nicole Hutchison. Ai4se and se4ai: Setting the roadmap toward human-machine co-learning. *INSIGHT*, 25(4):80–84, 2022.

[7] Tyler Cody, Niloofar Shadab, Alejandro Salado, and Peter Beling. Core and periphery as closed-system precepts for engineering general intelligence. In *International Conference on Artificial General Intelligence*, pages 209–219. Springer, 2022.

[8] Niloofar Shadab, Tyler Cody, Alejandro Salado, and Peter Beling. Closed systems paradigm for intelligent systems. In *2022 IEEE International Systems Conference (SysCon)*, pages 1–8. IEEE, 2022.

[9] Ben Goertzel. Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 5(1):1, 2014.

[10] James Llinas. Motivations for and initiatives on ai engineering. *Engineering Artificially Intelligent Systems: A Systems Engineering Approach to Realizing Synergistic Capabilities*, pages 1–18, 2021.

[11] Niloofar Shadab, Tyler Cody, Alejandro Salado, and Peter A. Beling. A systems-theoretical formalization of closed systems. 2023.

[12] David Long and Zane Scott. *A primer for model-based systems engineering*. Lulu. com, 2012.

[13] Dominique Luzeaux. A formal foundation of systems engineering. In *Complex Systems Design & Management*, pages 133–148. Springer, 2015.

[14] Alexander Kossiakoff, William N Sweet, Samuel J Seymour, and Steven M Biemer. *Systems engineering principles and practice*, volume 83. John Wiley & Sons, 2011.

[15] J Michael Hanratty, Robert H Lightsey, and Arvid G Larson. Open systems and the systems engineering process. In *INCOSE International Symposium*, volume 9, pages 1343–1348. Wiley Online Library, 1999.

[16] Steve Povolny and Shivangee Trivedi. Model hacking adas to pave safer roads for autonomous vehicles. *McAfee Blogs*, 2020.

[17] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. Robust physical-world attacks on deep learning visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1625–1634, 2018.

[18] Adith Boloor, Xin He, Christopher Gill, Yevgeniy Vorobeychik, and Xuan Zhang. Simple physical adversarial examples against end-to-end autonomous driving models. In *2019 IEEE International Conference on Embedded Software and Systems (ICESS)*, pages 1–7. IEEE, 2019.

[19] Dasom Lee and David J Hess. Regulations for on-road testing of connected and automated vehicles: Assessing the potential for global safety harmonization. *Transportation Research Part A: Policy and Practice*, 136:85–98, 2020.

[20] Xingyu Zhao, Valentin Robu, David Flynn, Kizito Salako, and Lorenzo Strigini. Assessing the safety and reliability of autonomous vehicles from road testing. In *2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE)*, pages 13–23. IEEE, 2019.

[21] Avner Engel. *Verification, validation, and testing of engineered systems*, volume 73. John Wiley & Sons, 2010.

[22] Akhil Shetty, Mengqiao Yu, Alex Kurzhanskiy, Offer Grembek, Hamidreza Tavafoghi, and Pravin Varaiya. Safety challenges for autonomous vehicles in the absence of connectivity. *Transportation research part C: emerging technologies*, 128:103133, 2021.

[23] Sebastian Krügel and Matthias Uhl. The risk ethics of autonomous vehicles. 2022.

[24] Douwe Kiela, Luana Bulat, Anita L Vero, and Stephen Clark. Virtual embodiment: A scalable long-term strategy for artificial intelligence research. *arXiv preprint arXiv:1610.07432*, 2016.

[25] Ikas Kassabalidis, MA El-Sharkawi, RJ Marks, P Arabshahi, and AA Gray. Swarm intelligence for routing in communication networks. In *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270)*, volume 6, pages 3613–3617. IEEE, 2001.

[26] Zhe Zou, Rong Zhao, Yujie Wu, Zheyu Yang, Lei Tian, Shuang Wu, Guanrui Wang, Yongchao Yu, Qi Zhao, Mingwang Chen, et al. A hybrid and scalable brain-inspired robotic platform. *Scientific reports*, 10(1):1–13, 2020.

[27] Shili Sheng, Erfan Pakdamanian, Kyungtae Han, Ziran Wang, John Lenneman, and Lu Feng. Trust-based route planning for automated vehicles. In *Proceedings of the*

ACM/IEEE 12th International Conference on Cyber-Physical Systems, pages 1–10, 2021.

[28] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138, 2010.

[29] Miguel Aguilera, Beren Millidge, Alexander Tschantz, and Christopher L Buckley. How particular is the physics of the free energy principle? *Physics of Life Reviews*, 2021.

[30] Martin Biehl, Felix A Pollock, and Ryota Kanai. A technical critique of some parts of the free energy principle. *Entropy*, 23(3):293, 2021.

[31] Henggang Cui, Gregory R Ganger, and Phillip B Gibbons. Scalable deep learning on distributed gpus with a gpu-specialized parameter server. *Cmu. edu.[Online]. Available: https://www. pdl. cmu. edu/PDLFTP/BigLearning/CMU-PDL-15-107. pdf.[Accessed: 01-Sep-2020]*, 2015.

[32] Mohammad Khan, Didrik Nielsen, Voot Tangkaratt, Wu Lin, Yarin Gal, and Akash Srivastava. Fast and scalable bayesian deep learning by weight-perturbation in adam. In *International Conference on Machine Learning*, pages 2611–2620. PMLR, 2018.

[33] Matthias Feurer, Benjamin Letham, and Eytan Bakshy. Scalable meta-learning for bayesian optimization. *stat*, 1050(6), 2018.

[34] Jon Holt and Simon Perry. *SysML for systems engineering*, volume 7. IET, 2008.

[35] Ralph W Gerard. Entitation, animorgs and other systems. In *Views on General System Theory: Proceedings of the 2nd Systems Symposium at Case Institute, John Wiley and Sons*, 1964.

[36] Megan Blakemore. Problem scoping design thinking and close reading: Makerspaces in the school library. *Knowledge Quest*, 46(4):66–69, 2018.

[37] Saurabh Mittal and Andreas Tolk. The complexity in application of modeling and simulation for cyber physical systems engineering. *Complexity Challenges in Cyber Physical Systems: Using Modeling and Simulation (M&S) to Support Intelligence, Adaptation and Autonomy*, page 103, 2019.

[38] Dennis M Buede and William D Miller. The engineering design of systems: models and methods. 2016.

[39] Anatol Rapoport Bertalanffy, Kenneth E Boulding, William Ross Ashby, Margaret Mead, and Gregory Bateson. *L. von Bertalanffy, General System Theory.* New York: George Braziller, 1968.

[40] Nils Bertschinger, Eckehard Olbrich, Nihat Ay, and Jürgen Jost. Information and closure in systems theory. In *Explorations in the Complexity of Possible Life. Proceedings of the 7th German Workshop of Artificial Life*, pages 9–21. IOS Press Amsterdam, The Netherlands, 2006.

Appendix E

Global Bibliography

Bibliography

[1] Tom McDermott, Dan DeLaurentis, Peter Beling, Mark Blackburn, and Mary Bone. Ai4se and se4ai: A research roadmap. *Insight*, 23(1):8–14, 2020.

[2] Niloofar Shadab, Aditya U Kulkarni, and Alejandro Salado. Shifting paradigms in verification and validation of ai-enabled systems: A systems-theoretic perspective. In *Systems Engineering and Artificial Intelligence*, pages 363–378. Springer, 2021.

[3] Alejandro Salado and Hanumanthrao Kannan. Elemental patterns of verification strategies. *Systems Engineering*, 22(5):370–388, 2019.

[4] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 427–436, 2015.

[5] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. arxiv e-prints, page. *arXiv preprint arXiv:1312.6199*, 2013.

[6] F Chollet. On the measure of intelligence. cs, 2019.

[7] Pei Wang. On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37, 2019.

[8] Sofia Samoili, Montserrat Lopez Cobo, Emilia Gomez, Giuditta De Prato, Fernando Martinez-Plumed, and Blagoj Delipetrev. Ai watch. defining artificial intelligence. towards an operational definition and taxonomy of artificial intelligence. 2020.

[9] Shili Sheng, Erfan Pakdamanian, Kyungtae Han, Ziran Wang, John Lenneman, and Lu Feng. Trust-based route planning for automated vehicles. In *Proceedings of the ACM/IEEE 12th International Conference on Cyber-Physical Systems*, pages 1–10, 2021.

[10] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138, 2010.

[11] Miguel Aguilera, Beren Millidge, Alexander Tschantz, and Christopher L Buckley. How particular is the physics of the free energy principle? *Physics of Life Reviews*, 2021.

[12] Martin Biehl, Felix A Pollock, and Ryota Kanai. A technical critique of some parts of the free energy principle. *Entropy*, 23(3):293, 2021.

[13] Ludwig Von Bertalanffy. General systems theory. *The Science of Synthesis: Exploring the Social Implications of General Systems Theory*, 103, 2010.

[14] Fremont E Kast and James E Rosenzweig. General systems theory: Applications for organization and management. *Academy of management journal*, 15(4):447–465, 1972.

[15] Norman S Nise. *Control systems engineering*. John Wiley & Sons, 2020.

[16] A Wayne Wymore. *Model-based systems engineering*, volume 3. CRC press, 2018.

[17] Anatol Rapoport Bertalanffy, Kenneth E Boulding, William Ross Ashby, Margaret Mead, and Gregory Bateson. *L. von Bertalanffy, General System Theory.* New York: George Braziller, 1968.

[18] Ludwig Von Bertalanffy. General system theory: foundations, development, applications. george braziller. *Inc., New York*, 1968.

[19] Yingxu Wang. Towards the abstract system theory of system science for cognitive and intelligent systems. *Complex & Intelligent Systems*, 1(1):1–22, 2015.

[20] Milan Zeleny. What is autopoiesis. *Autopoiesis: A theory of living organization*, pages 4–17, 1981.

[21] Francisco G Varela, Humberto R Maturana, and Ricardo Uribe. Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4):187–196, 1974.

[22] John Bednarz Jr. Autopoiesis: The organizational closure of social systems. *Systems Research*, 5(1):57–64, 1988.

[23] Nils Bertschinger, Eckehard Olbrich, Nihat Ay, and Jürgen Jost. Information and closure in systems theory. In *Explorations in the Complexity of Possible Life. Proceedings of the 7th German Workshop of Artificial Life*, pages 9–21. IOS Press Amsterdam, The Netherlands, 2006.

[24] Nils Bertschinger, Eckehard Olbrich, Nihat Ay, and Jürgen Jost. Autonomy: An information theoretic perspective. *Biosystems*, 91(2):331–345, 2008.

[25] Angel Mora, Pablo Cordero, Manuel Enciso, Inmaculada Fortes, and Gabriel Aguilera. Closure via functional dependence simplification. *International Journal of Computer Mathematics*, 89(4):510–526, 2012.

[26] Mihajlo D Mesarovic and Yasuhiko Takahara. *General systems theory: mathematical foundations*. Academic press, 1975.

[27] Stephanie Hoffmann. An evaluation of the autopoietic account of interests. *Synthese*, 200(3):1–18, 2022.

[28] W Ross Ashby and Jeffrey Goldstein. Variety, constraint, and the law of requisite variety. *Emergence: Complexity and Organization*, 13(1/2):190, 2011.

[29] Carlos Gershenson. Requisite variety, autopoiesis, and self-organization. *Kybernetes*, 2015.

[30] W Ross Ashby. *An introduction to cybernetics*. Chapman & Hall Ltd, 1961.

[31] W Ross Ashby. Requisite variety and its implications for the control of complex systems. In *Facets of systems science*, pages 405–417. Springer, 1991.

[32] Niloofar Shadab, Tyler Cody, Alejandro Salado, and Peter Beling. Closed systems paradigm for intelligent systems. In *2022 IEEE International Systems Conference (SysCon)*, pages 1–8. IEEE, 2022.

[33] Tyler Cody, Niloofar Shadab, Alejandro Salado, and Peter Beling. Core and periphery as closed-system precepts for engineering general intelligence. In *International Conference on Artificial General Intelligence*, pages 209–219. Springer, 2022.

[34] Niloofar Shadab, Tyler Cody, Alejandro Salado, and Peter A. Beling. A systems-theoretical formalization of closed systems. 2023.

[35] Jie M Zhang, Mark Harman, Lei Ma, and Yang Liu. Machine learning testing: Survey, landscapes and horizons. *IEEE Transactions on Software Engineering*, 2020.

[36] Niloofar Shadab and Alejandro Salado. Towards an interface description template for ai-enabled systems. *arXiv preprint arXiv:2007.07250*, 2020.

[37] Niloofar Shadaba, Tyler Codya, Peter Belinga, and Alejandro Saladob. Product herding for intelligent systems. 2023.

[38] Niamat Ullah Ibne Hossain, Raed M Jaradat, Michael A Hamilton, Charles B Keating, and Simon R Goerger. A historical perspective on development of systems engineering discipline: A review and analysis. *Journal of Systems Science and Systems Engineering*, 29(1):1–35, 2020.

[39] Andrew P Sage. *Systems engineering*, volume 6. John Wiley & Sons, 1992.

[40] Gordon Baxter and Ian Sommerville. Socio-technical systems: From design methods to systems engineering. *Interacting with computers*, 23(1):4–17, 2011.

[41] Koen H Van Dam, Igor Nikolic, and Zofia Lukszo. *Agent-based modelling of socio-technical systems*, volume 9. Springer Science & Business Media, 2012.

[42] M Kudret Yurtseven and Walter W Buchanan. Socio-technical system design: a general systems theory perspective. In *Proc. The International Conference on Engineering and Computer Education-ICECE'2013*, 2013.

[43] Louis S Metzger, George Rebovich Jr, Robin A Cormier, Spurgeon Jr T Norman, Deborah L Schuh, Peter A Smyton, Robert S Swarz, and Frederick C Wendt. Systems engineering guide: Collected wisdom from mitres systems engineering experts. Technical report, MITRE CORP BEDFORD MA BEDFORD United States, 2014.

[44] Charles Keating, Ralph Rogers, Resit Unal, David Dryer, Andres Sousa-Poza, Robert Safford, William Peterson, and Ghaith Rabadi. System of systems engineering. *Engineering Management Journal*, 15(3):36–45, 2003.

[45] John Boardman and Brian Sauser. System of systems-the meaning of of. In *2006 IEEE/SMC International Conference on System of Systems Engineering*, pages 6–pp. IEEE, 2006.

[46] Ryszard Tadeusiewicz. Introduction to intelligent systems. In *Intelligent systems*, pages 1–1. CRC Press, 2018.

[47] Alexander Kossiakoff, William N Sweet, Samuel J Seymour, and Steven M Biemer. *Systems engineering principles and practice*, volume 83. John Wiley & Sons, 2011.

[48] Antonio R Damasio. Descartes' error and the future of human life. *Scientific American*, 271(4):144–144, 1994.

[49] Jie Lu, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang. Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12):2346–2363, 2018.

[50] Pauching Yap, Hippolyt Ritter, and David Barber. Addressing catastrophic forgetting in few-shot problems. In *International Conference on Machine Learning*, pages 11909–11919. PMLR, 2021.

[51] Hong Wang, Yanjun Huang, Amir Khajepour, Dongpu Cao, and Chen Lv. Ethical decision-making platform in autonomous vehicles with lexicographic optimization based model predictive controller. *IEEE transactions on vehicular technology*, 69(8):8164–8175, 2020.

[52] Cristina Mele, Jacqueline Pels, and Francesco Polese. A brief review of systems theories and their managerial applications. *Service science*, 2(1-2):126–135, 2010.

[53] A Wayne Wymore. *A mathematical theory of systems engineering: the elements*. Wiley, 1967.

[54] Tyler Cody, Stephen Adams, and Peter A Beling. A systems theoretic perspective on transfer learning. In *2019 IEEE International Systems Conference (SysCon)*, pages 1–7. IEEE, 2019.

[55] Alejandro Salado. A systems-theoretic articulation of stakeholder needs and system requirements. *Systems Engineering*, 24(2):83–99, 2021.

[56] W Ross Ashby. Principles of the self-organizing system. In *Facets of systems science*, pages 521–536. Springer, 1991.

[57] George A Hazelrigg and Donald G Saari. Toward a theory of systems engineering. *Journal of Mechanical Design*, 144(1), 2022.

[58] JE Kasser and DK Hitchins. Unifying systems engineering: Seven principles for systems engineered solution systems, proceedings of the 20th international symposium of the incose. 2011.

[59] David D Walden and Robert E McKillip. Intelligent systems: a systems engineering approach. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3):871–882, 1991.

[60] Russell C Eberhart and James S Albus. Systems engineering for artificial intelligence. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 32(4):435–440, 2002.

[61] Thomas A McDermott, Mark R Blackburn, and Peter A Beling. Artificial intelligence and future of systems engineering. In *Systems Engineering and Artificial Intelligence*, pages 47–59. Springer, 2021.

[62] James Llinas, Hesham Fouad, and Ranjeev Mittu. Systems engineering for artificial intelligence-based systems: A review in time. *Systems Engineering and Artificial Intelligence*, pages 93–113, 2021.

[63] Ernesto Exposito, Juan R Rabu nal, and Rafael S Barbado. Systems engineering for artificial intelligence and cognitive systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(6):827–834, 2016.

[64] Defense Acquisition Guidebook DAG and Workplace Ethic. *Introduction to systems engineering*. Citeseer, 2000.

[65] David D Walden and John B Goodenough. Mission engineering: A framework for the systems engineering of complex systems. *Systems Engineering*, 7(1):1–18, 2004.

[66] John B Goodenough and Richard E Schatzki. The concept of closure in systems engineering. *Systems Engineering*, 5(2):89–106, 2002.

[67] John B Goodenough and David D Walden. A closure-based approach to systems engineering. *Systems Engineering*, 8(1):1–17, 2005.

[68] Robert K Logan. What is information?: Why is it relativistic and what is its relationship to materiality, meaning and organization. *Information*, 3(1):68–91, 2012.

[69] Christoph Adami. What is information? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2063):20150230, 2016.

[70] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[71] Hans Christian Von Baeyer. *Information: The new language of science.* Harvard University Press, 2004.

[72] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55, 2001.

[73] Thomas M Cover, Joy A Thomas, et al. Entropy, relative entropy and mutual information. *Elements of information theory*, 2(1):12–13, 1991.

[74] Erik G Learned-Miller. Entropy and mutual information. *Department of Computer Science, University of Massachusetts, Amherst*, page 4, 2013.

[75] Mai Vu. Entropy and mutual information. *Retrieved March*, 2, 2014.

[76] Jeff A Estefan et al. Survey of model-based systems engineering (mbse) methodologies. *Incose MBSE Focus Group*, 25(8):1–12, 2007.

[77] Alejandro Salado and Aditya U Kulkarni. An assessment of the adequacy of common definitions of the concept of system. In *INCOSE International Symposium*, volume 31, pages 510–521. Wiley Online Library, 2021.

[78] Mihajlo D Mesarovic and Yasuhiko Takahara. Complex systems. *Abstract Systems Theory*, pages 330–414, 2006.

[79] Robert Laurini and Derek Thompson. *Fundamentals of spatial information systems*, volume 37. Academic press, 1992.

[80] MIHAJLO D MESAROVIC. Conceptual basis for a mathematical theory of general systems. *Kybernetes*, 1(1):35–40, 1972.

[81] David Krakauer, Nils Bertschinger, Eckehard Olbrich, Jessica C Flack, and Nihat Ay. The information theory of individuality. *Theory in Biosciences*, 139(2):209–223, 2020.

[82] W Ross Ashby. Principles of the self-organizing system. *Modern systems research for the behavioral scientist*, 108, 1968.

[83] George J Klir and W Ross Ashby. Requisite variety and its implications for the control of complex systems. *Facets of systems science*, pages 405–417, 1991.

[84] Markus Schwaninger and Stefan Ott. What is variety engineering and why do we need it? *Systems Research and Behavioral Science*, 2023.

[85] JDR de Raadt. Ashby's law of requisite variety: An empirical study. *Cybernetics and Systems: An International Journal*, 18(6):517–536, 1987.

[86] Elinor Ostrom. Designing complexity to govern complexity. *Property rights and the environment: Social and ecological issues*, pages 33–45, 1995.

[87] Max Boisot and Bill McKelvey. Complexity and organization-environment relations: Revisiting ashby's law of requisite variety. *P. Allen, S. Maguire, & B. McKelvey, Complexity and Management*, pages 279–298, 2011.

[88] Danko Nikolić. Why deep neural nets cannot ever match biological intelligence and what to do about it? *International Journal of Automation and Computing*, 14(5):532–541, 2017.

[89] Danko Nikolic. Testing the theory of practopoiesis using closed loops. In *Closed Loop Neuroscience*, pages 53–65, 2016.

[90] Léon Bottou and Vladimir Vapnik. Local learning algorithms. *Neural computation*, 4(6):888–900, 1992.

[91] Andreas Wagner. *Robustness and evolvability in living systems*, volume 24. Princeton university press, 2013.

[92] Jiawei Zhang. Cognitive functions of the brain: perception, attention and memory. *arXiv preprint arXiv:1907.02863*, 2019.

[93] Diane B Paul. The selection of the "survival of the fittest". *Journal of the History of Biology*, 21:411–424, 1988.

[94] F Eugene Yates. Order and complexity in dynamical systems: homeodynamics as a generalized mechanics for biology. *Mathematical and computer modelling*, 19(6-8):49–74, 1994.

[95] Walter Bradford Cannon. Homeostasis. *The wisdom of the body. Norton, Newyork*, 1932.

[96] Stefan Reimann. Homeostasis and stability. *Research Center for Interdisciplinary Studies on Structure Formation*, pages 1–13, 1996.

[97] John S Torday. Homeostasis as the mechanism of evolution. *Biology*, 4(3):573–590, 2015.

[98] David Lloyd, Miguel A Aon, Sonia Cortassa, et al. Why homeodynamics, not homeostasis? *The Scientific World Journal*, 1:133–145, 2001.

[99] Takashi Ikegami and Keisuke Suzuki. From a homeostatic to a homeodynamic self. *BioSystems*, 91(2):388–400, 2008.

[100] Jane Friedman. Suspended judgment. *Philosophical studies*, 162(2):165–181, 2013.

[101] Jane Friedman. Inquiry and belief. *Noûs*, 53(2):296–315, 2019.

[102] Hannes Leitgeb. The stability theory of belief. *Philosophical review*, 123(2):131–171, 2014.

[103] Brett Koonce and Brett Koonce. Resnet 50. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, pages 63–72, 2021.

[104] William D MILLER. The future of systems engineering: Realizing the systems engineering vision 2035. 2022.

[105] Kara Pepe and Nicole Hutchison. Ai4se and se4ai: Setting the roadmap toward human-machine co-learning. *INSIGHT*, 25(4):80–84, 2022.

[106] Ben Goertzel. Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 5(1):1, 2014.

[107] James Llinas. Motivations for and initiatives on ai engineering. *Engineering Artificially Intelligent Systems: A Systems Engineering Approach to Realizing Synergistic Capabilities*, pages 1–18, 2021.

[108] David Long and Zane Scott. *A primer for model-based systems engineering.* Lulu. com, 2012.

[109] Dominique Luzeaux. A formal foundation of systems engineering. In *Complex Systems Design & Management*, pages 133–148. Springer, 2015.

[110] J Michael Hanratty, Robert H Lightsey, and Arvid G Larson. Open systems and the systems engineering process. In *INCOSE International Symposium*, volume 9, pages 1343–1348. Wiley Online Library, 1999.

[111] Steve Povolny and Shivangee Trivedi. Model hacking adas to pave safer roads for autonomous vehicles. *McAfee Blogs*, 2020.

[112] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. Robust physical-world attacks on deep learning visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1625–1634, 2018.

[113] Adith Boloor, Xin He, Christopher Gill, Yevgeniy Vorobeychik, and Xuan Zhang. Simple physical adversarial examples against end-to-end autonomous driving models.

In *2019 IEEE International Conference on Embedded Software and Systems (ICESS)*, pages 1–7. IEEE, 2019.

[114] Dasom Lee and David J Hess. Regulations for on-road testing of connected and automated vehicles: Assessing the potential for global safety harmonization. *Transportation Research Part A: Policy and Practice*, 136:85–98, 2020.

[115] Xingyu Zhao, Valentin Robu, David Flynn, Kizito Salako, and Lorenzo Strigini. Assessing the safety and reliability of autonomous vehicles from road testing. In *2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE)*, pages 13–23. IEEE, 2019.

[116] Avner Engel. *Verification, validation, and testing of engineered systems*, volume 73. John Wiley & Sons, 2010.

[117] Akhil Shetty, Mengqiao Yu, Alex Kurzhanskiy, Offer Grembek, Hamidreza Tavafoghi, and Pravin Varaiya. Safety challenges for autonomous vehicles in the absence of connectivity. *Transportation research part C: emerging technologies*, 128:103133, 2021.

[118] Sebastian Krügel and Matthias Uhl. The risk ethics of autonomous vehicles. 2022.

[119] Douwe Kiela, Luana Bulat, Anita L Vero, and Stephen Clark. Virtual embodiment: A scalable long-term strategy for artificial intelligence research. *arXiv preprint arXiv:1610.07432*, 2016.

[120] Ikas Kassabalidis, MA El-Sharkawi, RJ Marks, P Arabshahi, and AA Gray. Swarm intelligence for routing in communication networks. In *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270)*, volume 6, pages 3613–3617. IEEE, 2001.

[121] Zhe Zou, Rong Zhao, Yujie Wu, Zheyu Yang, Lei Tian, Shuang Wu, Guanrui Wang, Yongchao Yu, Qi Zhao, Mingwang Chen, et al. A hybrid and scalable brain-inspired robotic platform. *Scientific reports*, 10(1):1–13, 2020.

[122] Henggang Cui, Gregory R Ganger, and Phillip B Gibbons. Scalable deep learning on distributed gpus with a gpu-specialized parameter server. *Cmu. edu.[Online]. Available: https://www. pdl. cmu. edu/PDLFTP/BigLearning/CMU-PDL-15-107. pdf.[Accessed: 01-Sep-2020]*, 2015.

[123] Mohammad Khan, Didrik Nielsen, Voot Tangkaratt, Wu Lin, Yarin Gal, and Akash Srivastava. Fast and scalable bayesian deep learning by weight-perturbation in adam. In *International Conference on Machine Learning*, pages 2611–2620. PMLR, 2018.

[124] Matthias Feurer, Benjamin Letham, and Eytan Bakshy. Scalable meta-learning for bayesian optimization. *stat*, 1050(6), 2018.

[125] Jon Holt and Simon Perry. *SysML for systems engineering*, volume 7. IET, 2008.

[126] Ralph W Gerard. Entitation, animorgs and other systems. In *Views on General System Theory: Proceedings of the 2nd Systems Symposium at Case Institute, John Wiley and Sons*, 1964.

[127] Megan Blakemore. Problem scoping design thinking and close reading: Makerspaces in the school library. *Knowledge Quest*, 46(4):66–69, 2018.

[128] Saurabh Mittal and Andreas Tolk. The complexity in application of modeling and simulation for cyber physical systems engineering. *Complexity Challenges in Cyber Physical Systems: Using Modeling and Simulation (M&S) to Support Intelligence, Adaptation and Autonomy*, page 103, 2019.

[129] Dennis M Buede and William D Miller. The engineering design of systems: models and methods. 2016.