Verification of Autonomous Systems: Challenges of the Present and Areas for Exploration

Andrew Bouchard[1] and Richard Tatum [1]

I. ABSTRACT

The words "failure" and "success" are simple binary terms used to describe the outcomes of maritime autonomous behavior, yet how to draw the line separating the two stark realities demonstrates the subtle nature of this problem. If how to judge success or failure is improperly performed, users of systems that rely upon autonomous behavior may quickly find their initial trust misplaced and in some cases, the trust may never be recovered. A loss of trust or an unwillingness to trust newer technologies can greatly hinder the transition and use of such technologies. Acceptance of current and future maritime autonomy so that its use is more ubiquitous and commonplace requires the establishment of rigorous techniques for which clearly defined verification of autonomy can be performed. Therefore, it is critical that verification of maritime autonomous systems be tenably and consistently implemented. For our work, we identify some of the difficulties with verification efforts with respect to general autonomy and not only maritime autonomy. In an attempt to address the difficulties, we present the potential use of techniques found within other fields to assist with some of these difficulties.

II. INTRODUCTION

As humans more and more cede dangerous, dull, and dirty chores to the realm of robotics, it is increasingly important that the community seek assurances that the performance of these systems is effective, reliable, and safe. While the need for performance assurance is important for maritime systems, it is certainly important for general autonomy verification as well. Such assurances are given by verification, which is the process of obtaining proof that a system performs as expected by meeting its performance requirements. Verification provides essential feedback to system developers to direct their improvement efforts, helps consumers of autonomy to compare solutions, and provides information to users to build their trust in the system.

Despite the importance of the verification task, there is currently no widely accepted process for verifying autonomous systems due to the difficulty of the challenge. While there have been a variety of attempts, all so far have fallen short of providing a general approach that can be applied across the field. In this paper, we attempt to chart a path forward to address this need based on work already done not only in the field of autonomous robotics, but also in other relevant fields that have addressed similar challenges.

Before directly addressing the verification task, it is essential first that we define the bounds of what we seek to verify. For the purposes of our work, we define autonomy as a system that makes a decision based on information where its interaction with the world is intractably complex. Further, we define a system as an immobot, robot, group of immobots or group of robots, centralized or decentralized with hardware, software perception, actuation, communications, and decision-making that are abstracted as a unit and act in the world as a whole. These definitions were developed through participation in the Naval Research Labs Verification of Autonomous Systems Working Group.

Our approach to address this need first requires that we outline the challenges particular to the verification of autonomous systems that make current verification methods for other types of technology inadequate, which we do in Section III. In Section IV, we provide the lessons learned from a literature review of attempts already made to verify autonomous systems and in Section V, we give useful analogies from other fields. Finally, in Section VI, we recommend a path forward for developing a general method for verification of autonomous systems.

III. SOFTWARE VERIFICATION AND CHALLENGES TO THE VERIFICATION OF AUTONOMOUS SYSTEMS

As a key component of autonomy is software, it is reasonable to investigate standard verification techniques established in the field of software engineering. These techniques can be classified as either dynamic or static dynamic methods. Dynamic methods are concerned primarily with the development and use of tests to discover issues with respect to verification [1]. The test cases are developed with respect to the software requirements and can be used during the modular, integration, or system test phases of the overall verification phase of development. Typically, the test cases can be traced back to requirements contained in an official requirements document. On the other hand, static methods are those methods that tend to be more analytical in nature. Examples of static methods include anti-pattern detection, software metrics, and formal methods. Over several years, established design patterns with respect to algorithm development have emerged and are implemented to avoid costly programming errors. Anti-pattern methods seek to find and

[1]The authors are autonomy researchers associated with Naval Surface Warfare Center Panama City Division, Panama City, FL andrew.bouchard@navy.mil, richard.d.tatum@navy.mil

correct violations of accepted design patterns [2]. Typical software metrics try to quantify code accuracy and can be measured as number of bugs per line of code or code coverage. While the listed metrics are useful, they are by no means sufficient for measuring software accuracy. Formal methods attempt to logically prove that software satisfies a formal specification [3], typically by using a deterministic model of the system in question to show that the system has some properties for all possible inputs. While outside of the scope of software verification, an analogous application of formal methods has been to produce computer generated proofs of famous mathematics problems such as the four color theorem [4]. Unfortunately, the existing verification techniques cannot be used directly to the verification of autonomy without further augmentation. The dynamic methods rely heavily upon test case generation, which is problematic for autonomy verification. The very nature of autonomous systems requires algorithms and behaviors that respond to the environment. The stochastic properties of the environment introduces temporal and spatial complexities that give rise to the problem of intractable complexity, which makes test case development very difficult. As for the static methods, the anti-pattern methods and various code metrics can be useful in the identification of some initial problems, but of course, they do not always necessarily provide complete coverage with respect to verification. This of course leaves the formal methods, which attempt to address the gap with respect to rigor. As formal methods require determinism and well-characterized systems, the stochastic influences of the environment make formal methods difficult to employ. Many of the most innovative methods applied to autonomous systems are non-deterministic, and because autonomous systems are intended to handle unexpected inputs, by definition all inputs cannot be enumerated. With respect to formal theorem provers, generated proofs also have to be verified as well, which simply transfers the problem to another level.

The problems identified are relevant to the verification of autonomy and any field in which intractable complexity is an issue. Since we are interested in autonomy verification, it is natural to inquire as to the state of the art with respect to autonomy verification, which is discussed in the next section.

IV. Verification Efforts in Autonomous Systems

In order to understand what attempts have already been made in the field of autonomous robots to assess functionality, reliability, and safety, we gathered information from a wide variety of sources. The resulting literature review includes sources from research labs testing individual robot behaviors, formal efforts to define general autonomy frameworks, and reports on methodologies used by specific organizations. Each was analyzed for the specific method that they took to describe or evaluate autonomy, and the results were aggregated to identify trends. The literature review reached saturation; that is, there came a point during the review when additional works added no new information. While significant weaknesses exist that make any of the existing attempts insufficient for the general case of autonomy verification, these attempts contribute valuable insights into how the community of stakeholders conceptualize both autonomy and verification. By collecting the strengths and weaknesses of past efforts, our intent is to craft a way forward that effectively addresses the verification need by taking advantage of those lessons already learned. In our analysis of existing work, we divided contributions into frameworks that describe autonomy, essential attributes that describe autonomy performance, and methods for assessing those attributes.

A. *Frameworks*

It appears to be well understood that, in order to evaluate a system, both the system itself and the method of evaluation must be formally defined and clearly understood. Therefore, it comes as no surprise that the most general work found has been done in the area of defining frameworks to describe autonomous systems and scales for assessment.

1) Levels of Autonomy: Many different groups have proposed conceptual models to describe autonomy in levels, assigning a numerical scale to describe the capabilities of a given robot. These frameworks include the Autonomy Levels for Unmanned Systems (ALFUS) [5]–[7] advanced by National Institute of Standards and Technology (NIST), the Mobility, Acquisition, Protection (MAP) framework from Los Alamos National Lab [8], the 3D Intelligence Space from Draper Labs [9], the US Army Mission Performance Potential (MPP) framework [10], and Air Force Research Labs Autonomous Control Levels (ACL) [11], [12]. While each framework varies in its emphasis depending on the context in which it was developed, each takes essentially the same approach to assign a level of autonomy to a given system: some number of aspects of the system are considered, and some number of levels are given to describe each aspect of the system from least mature to most mature. Furthermore, some subjective guidance is given to aid human evaluators in assigning each aspect of the system to a level. Essentially, levels of autonomy nearly invariably attempt to solve the problem of subjective evaluation of autonomous systems being insufficient by breaking it down into multiple subjective evaluations. Perhaps this is one of the factors that led the Defense Science Board in 2012 to label attempts to define levels of autonomy as not useful [13], favoring instead the definition of system models and frameworks to better describe autonomy without developing a single, concise definition. Although the levels of autonomy appear formally abandoned, they still provide valuable insight into what aspects of systems were deemed critical to the overall notion of its autonomy.

2) System Models: As the field of autonomous systems has expanded, various developers have created useful models for conceptually describing the functionality of their system. Highly general descriptions such as the classic observe, orient, decide, act (OODA) loop [14] or the Cynekin framework [15] can help to organize systems or components under broad headings, but lack the specificity to meaningfully describe the capabilities of most autonomous systems. One attempt to add specificity to the OODA loop is the System Capabilities Technical Reference Model (SC-TRM) [16], which uses five steps to describe the cognitive process and significantly adds components for learning and memory. However, no uses of this framework outside its original application have been found, perhaps due to a lack of guidance in how to apply it to other projects. By contrast, NIST maintains the 4D/RCS reference model architecture [17] based on the general real-time control system (RCS) architecture, which outlines a more thoroughly defined hierarchy of control based on the scope of action at each level. While this solution has more guidance, few implementations of it have been found, likely due to its relative complexity. In a 2012 report on autonomy, the Defense Science Board recommended a Framework for the Design and Evaluation of Autonomous Systems [13] consisting of Cooperative Echelon, Mission Dynamics, and Complex System Trades Space views of systems. In our research, we were unable to find any application of this framework to an actual system, likely because it operates at the level of broad concepts and is therefore difficult to apply to specific technical implementations.

3) Human Survey Scales: Because autonomy has thus far depended on the subjective evaluations of human experts, a number of scales have been developed to help articulate these reviews in a way that is quantitative and somewhat standardized. A number of these tools were presented in [18] along with the metrics to which they have been applied, and are presented here to illustrate what guidance does exist. The most general tool is the Likert scale, a simple linear scale of incremental values or levels of agreement on which the evaluator is asked to indicate their response to a prompting question. This framework is very familiar to most individuals, but requires good design of the prompting questions to be effective. One example in which these questions have been standardized is the Lee & Moray Trust Scale, which presents the evaluator with a defined battery of questions related to their trust of a system answered on a 1-10 Likert scale. More sophisticated frameworks have been developed to evaluate specific system attributes, such as the Cooper-Harper Scale and NASA Task Load Index, which measures burden on a human operator, as well as SART, SAGAT, and SALIANT, all of which measure an operators situational awareness. It is important to note that all of these humans scales are fundamentally subjective, and also that they require careful definition of a taxonomy for evaluation.

B. Attributes

As already discussed, one of the principal challenges of evaluating autonomy is the relative complexity of autonomous systems, making it difficult to determine which attributes of a system are critical to the overall assessment of its independence from a human operator. By considering a large body of work, we have been able to collect many of the attributes that have been deemed important by other efforts, and have divided them into attributes that describe the robot's internal characteristics, the robot's relationship to its human operator, the robot's relationship to its mission, and the robot's relationship to the world.

1) The Robot and Itself: A variety of attributes of the system design independent of any outside interactions have been proposed as influencing the autonomy of the system. Most of the proposed attributes in this category address the complexity of the system, arguing that greater complexity makes the system more difficult to control. Specifically, as a part of its ALFUS framework, NIST includes system dependence as influencing the autonomy of a system [5] and NASA's Jet Propulsion Laboratory includes simplicity among its most important system attributes[19], arguing that simplicity of instantiation results in software that is "maintainable, testable, reusable, and supports interoperability." Other work more broadly defines a criteria of "self-control, characterized by the quantity and complexity of memory addresses and states in the system and also highlights evolution capability, measured by the extent and frequency of variable updates [20]. While these aspects of the system are measurable, their direct link to the autonomy of the system has yet to be demonstrated. However, the larger problem is that to measure any attributes of the system design, the evaluator must have access to the system at a level that is generally not available to anyone but its developers. This limits the usefulness of these attributes to third-party black-box testers.

2) The Robot and Its Operator: Many efforts have attempted to measure autonomy according to the amount of interaction required by the human operator, often measuring the ratio of commands to operational time or events [18], [20], [21]. Other efforts have measured the degree of human intervention specifically with regard to unplanned events, considering the time or level of effort required by a human operator to handle a replan or adapt the system's behavior to complete a task [5], [18]. More generally, other efforts have looked at the operators situational awareness as provided by the autonomous system [18], [22], reasoning that the more the operator understands about the vehicles situation and environment, the more effectively and efficiently he can monitor and control that system. Attributes that relate autonomy of the system back to the operator acknowledge that no system is fully autonomous, a misconception dealt with by several recent reports [13], [23].

3) The Robot and Its Mission: Although there have been efforts to evaluate autonomy independently of any specific mission context [10], the majority of work in the field to date acknowledges either explicitly or implicitly that the ability of a robot to successfully operate independently of a human operator will depend on its capabilities specific to the task at hand. However, a number of efforts have identified common attributes that cross most if not all mission domains, and hold some promise for a generalized evaluation. The most abstract of these attempts is a general definition of mission success and system errors [18], [19]. While this method requires criteria such as the percentage of correct decisions or a false alarm rate to be defined for each mission, having consistent criteria notionally offers some level of standardization between robots and mission sets. Other work suggests using concepts of precision, recall, and accuracy [24], measures of the robots ability to detect and process events [25], or the efficiency with which it uses resources like time or energy [25] as factors in evaluating mission autonomy. All of these are important factors in the successful completion of a mission, and useful points of comparison, but so far no widely accepted framework has emerged to unify these individual characteristics into a cohesive system evaluation. Perhaps the most ambitious and comprehensive effort to date in this area has been the effort to define Mission Complexity, one of the three axes used in the ALFUS framework [5]. While the resulting guidance is useful in its consideration of task structure, coordination among components, world modeling, and general performance, it requires specification of both the robot and the mission with excruciating detail such that the benefit of a generalized framework for evaluation is essentially lost.

4) The Robot and the World: The definition of autonomy selected by the Verification of Autonomous Systems Working Group makes the intractable complexity of the robots operating environment one of the key defining characteristics of an autonomous system. Therefore, that autonomy will be heavily enabled or constrained by the systems ability to operate in, understand, and react to its surroundings. One approach characterizes the environments that the robot can operate in, developing a variety of ways to assess its size, complexity, and dynamics [5], [25]. A few efforts have factored in the robustness of the platform itself [25], considering the ability of the system to survive in a given environment to be dependent to a large extent on its physical design and construction. This is an important consideration, and should not be neglected, but the majority of efforts have focused less on survivability and more on characterizing the vehicles understanding of its environment through sensing, perception, and reasoning about geospatial knowledge. This is most often approached through a notion of situational awareness [14], [22] analogous to human awareness of their operational environment and taking advantage of work already done in its assessment. Because the notion of autonomy is essentially to mimic human performance with software, this human analogy appears to work well, and suggests its utility in other areas as well, an opportunity explored in Section V.

C. Methods

Once a framework is chosen to define the structure of the evaluation and the critical attributes are selected, what remains is how to measure those attributes in the context of the framework. Many different methods of measurement have been proposed depending on what is being measured, and we have categorized our findings into four major categories: ratings based on human evaluation, application of existing formal methods, development of a specific heuristic, and testing structures and procedures.

1) Human-Based Ratings: Due to the lack of an accepted evaluation framework and quantitative metrics for describing autonomous systems, assessment efforts for autonomy to date have largely relied on subjective human evaluation, as demonstrated in a variety of examples [7], [10], [12], [18], [19], [22]. Depending on the purpose and priorities of the effort, the human evaluators may be experts in the development of autonomous system, expert operators, or other stakeholders in the creation, acquisition, and use of autonomy. However, even with guidance based on the frameworks already described, expert opinion varies and the performance of a given system changes with context, leading to inconsistencies in ratings even between evaluators of the same profession [7], [10], [13], [19], [21]. What is needed is a means of evaluating autonomous systems whose results are not dependent on the subjective opinions and preferences of a few individuals, but on the objective performance and characteristics of the system.

2) Formal Methods: Existing tools for proving software typically involve the creation of a system model for which all possible states can be enumerated and the system verified as stable. However, due to the problem of intractable complexity already outlined, this approach is difficult to apply to autonomous systems. Variations applied to autonomy generally involve testing a subset of possible situations in one way or another.

Modelling the system and simulating a subset of scenarios is a popular solution for planetary rovers, where the intended mission is well-defined and evaluators have access to the system design and implementation [19], [26], [27]. The well-defined nature of the systems and missions have also allowed for automation of testing and assessment based on human-provided scenarios [28], [29]. While this process has proven effective at debugging and hardening robust vehicle controllers for safely navigating alien terrain, its application to other systems is limited when considering more complex mission areas with greater chances for unknown inputs.

Another technique suggests the use of formal methods applied to a subset of the system in time. Just-in-time assurance considers possible states for the robot reachable within some reasonable time horizon, and verifies only that subset of the entire state space [30], [31]. This seems to be generally proposed for safety contexts, in which failure criteria can be easily defined (e.g. robot collides with human), due to the need for automated on-line verification.

Some researchers focus on applying formal methods to a subset of the robot capabilities, for instance, navigation using cost-to-go maps [32], [33]. Because the resulting metrics can be well-defined according to the specific function, they can be rigorously quantitative and objective. However, this approach is limited in its broader applicability, since not all autonomy capabilities are so well-defined as navigation, and this fails to account for interactions between components of a system.

While formal methods are useful in specific, well-defined subsets of the autonomy evaluation problem, they fall short of offering a complete solution for overall evaluation because they rely on limiting system inputs and outputs and having access to the complete definition of the system. By contrast, objective evaluation of autonomous systems typically is performed through black-box testing with no or limited knowledge of system implementation, and based on the definition of autonomy selected for this effort, the interactions of the system with its environment are required to be intractably complex.

3) Heuristic Models: As the name implies, heuristics tend to forgo the optimal solutions in place of suboptimal approaches to solutions due to the complexity of what is being modeled. With respect to autonomy verification, we have identified four distinctive approaches and they are dynamical systems based approaches, error growth factors, multi-criteria evaluation methods, and on board autonomy verification.

One method to obtain metrics for verification involves measuring some quality of the underlying dynamical system. In [34], the author uses a dynamical system to model the agent-environment interaction. The author shows how to estimate the attractor dimensionality by using correlation dimensionality. With estimates of the dimension of the attractor, different agent-environment systems can be compared more quantitatively. Of course, if a dynamical system does not effectively model the agent-environment interaction, which may be possible in some cases, then this and any of the dynamical systems heurisitics is of limited use.

Another dynamical systems related approach involves the use of error growth factors (EGF). The EGFs measure how sensitive a robots trajectory is with respect to initial conditions [35]. The EGF is approximated as the first Lyapunov exponent of a 1-d dynamical system. In particular, the author compared a simulation model to an actual robot through the use of EGF. The positive aspect of the dynamical systems heuristics is that these approaches try to measure the properties that are inherently associated with the complexity of the underlying system, which may provide a crucial first step to dealing with intractable complexity.

In [36], a set-theoretic framework is proposed and used for the task of plan generation, validation, execution, and monitoring. The authors call the implementation of their framework the Autonomous Reasoning Engine (ARE) and it can be used to provide on-board autonomy using model-based reasoning. This type of approach tries to use inferences during runtime about the model and newly available data to create a more accurate representation of the interactions of the autonomy and its environment.

The final heuristic we considered is known as multicriteria evaluation [26]. Typically, objective functions are 1-d and are used to provide solutions in a wide range of areas such as scheduling or sensor placement. Rather than optimize over scalar values, multicriteria evaluation seeks to optimize over a set of vectors, which allows for the relationships between parameters to be represented during the optimization process.

4) Testing Structures: As already discussed, the most common method of verification is scenario testing [25]–[27], [33], though its application is rarely formally discussed [15], [37]. The key challenge of scenario testing is that it cannot be exhaustive as the intractable complexity prevents this. Thus far, no efforts have defined a clear way to select a meaningful and generalizable set of scenarios a priori that will sufficiently represent the overall operation of the system. As a result, proposed solutions focus either on maximizing the number of scenarios performed and assessed or on producing some technique to intelligently generate relevant test cases on-line. NASA JPL has found success in the automated generation and evaluation of test cases for well-defined missions [29] based on manually defined criteria. While this automation allows many more tests to be run than would otherwise be possible, it is clearly limited by the ability of the criteria to represent relevant scenarios a priori and the quality of the design of those criteria. Alternatively, some groups have found success in identifying challenging areas for scenario testing by using extensive logging and randomization of scenarios. In [21], logs were reviewed a posteriori based on qualitative identification of challenges to the autonomous system in order to find and further test quantitative explanations for those challenges. The most robust proposal along those lines proposes the creation of an intelligent testing agent that can autonomously identify and generate system stressors to find the operational boundaries of the system within an intractably complex state space [16]. While an autonomous verification agent would itself require verification, the concept is that the one verification effort would then ease or solve every future verification. Currently, this idea is completely theoretical, but offers a promising answer to the complexity problem.

In summary, current efforts to evaluate autonomous systems were found to be reliant primarily on ad-hoc scenario tests and subjective human evaluation. While some approaches varied, the overwhelming majority depended at some level on one or both of these elements. The first indicates that the system being tested was put in some limited scenario that was relevant to the aspect of the system being tested. Generally, this scenario was intuitively difficult with regard to what was being tested, but simplified with regard to other aspects of the system or environment. However, in no case was any attempt made to generalize the results of the specific scenario or family of scenarios to the broader performance of the robot in other scenarios. With few exceptions, the final evaluation of the performance of an autonomous system was described by a subjective evaluation given by an autonomy expert or system user. While in many cases a framework was given to guide that evaluation, multiple papers still demonstrate variance among multiple experts, suggesting a lack of reliability in such methods.

V. Analogous Topics

An idea that garnered a significant amount of attention at the first meeting of the Verification of Autonomous Systems Working Group was that researchers look for analogies in other fields to aid in evaluation. The argument is that many other fields already address the problem of evaluating and assessing complex systems; therefore, there should be some lessons learned that can be applied from these past experiences. With that objective, we expanded our literature survey to include a variety of other disciplines to find useful analogies to robotic autonomous systems, which include the fields of philosophy, psychology, and mathematics.

A. Approach

An initial survey was done to explore possible fields for further research based on recommendations from the working group and use of general search terms, which helped to identify the most promising fields. Evaluation of service animals and animal behavior was expected to be a particularly fruitful area for exploration, but our initial analysis indicated similar problems to those in autonomous system evaluation analysis by highly subjective criteria and simple, non-standard questionnaires that assume expertise on the part of the evaluators [38]–[40].

Based on this broad search for opportunities, potential analogies for evaluation of autonomy were identified in the fields of philosophy, psychology, music, business, statistics, mathematics, and actuarial science. We decided to seek direction for further research from professionals in the field to either confirm perceived commonalities and guide further work or to cut short any investigations likely not to be fruitful.

B. Initial Exploration

These professional opinions were found by contacting professors at Florida State University in Tallahassee, Florida. Over the course of two days, researchers interviewed ten professors in these fields to determine what commonalities existed to autonomous systems that might be useful to the development of assessment and evaluation. While the area of music offered some interesting methods to handle the problems of subjective evaluation [41]–[43], it was determined that the techniques require an ontology, or a common understanding of meaning and taxonomy that does not yet exist for autonomy. Therefore, these techniques have been tabled for later reference while earlier work is done. Based on these discussions, it was decided to further pursue the areas of philosophy, psychology, and the several mathematical disciplines.

C. Philosophy

According to philosopher Isaiah Berlin, "The task of philosophy…is to extricate and bring to light the hidden categories and models in terms of which human beings think;" [44] or more colloquially, to define the fundamental nature of things. To the engineer or technologist, focused on the practical and implementable solution, to address such fundamental principles can seem an esoteric waste of time better spent developing and testing new systems. Perhaps this is why, consistently throughout the literature review of work provided above, the challenge arises of a lack of consistent definitions and frameworks to understand autonomous systems. Berlin charts through history a trend in which formal and empirical sciences developed by first being philosophical, without formal axioms or rules of observation by which to answer questions, and then divorcing themselves from philosophy once those frameworks were developed. Perhaps this is what is needed in this arena to first address autonomous systems as new entities needing clear definitions and frameworks of understanding before techniques can be developed for their assessment.

The challenge of defining the conditions for symbolic interaction within a culture is traditionally the realm of hermeneutics [45], a branch of philosophy focused on the understanding and interpretation of language. Even a brief foray into the field quickly shows how much work has been put in over centuries into the definition of language and communication through hermeneutics; philosophy can clearly contribute to the challenge of defining autonomy across the discipline of robotics. In fact, philosophy already offers its own notion of autonomy, and though it traditionally considers it in terms of human autonomy, its use of shared terms like "autonomous agents" and "determinism" suggest some shared space of understanding. Alfred Mele, in a general book on the topic, defines autonomy roughly as self-rule or self-government [46]. It seems that greater specificity is controversial within philosophy as well, but Mele argues in his book that self-control alone is insufficient for autonomy, and argues for an entire chain of conditions that must be met before an agent can be said to act autonomously. Others, perhaps more usefully to this purpose, define autonomy as a set of competences - capacities to choose rationally and objectively [47]. The analogy to robotics breaks down at the points that philosophy assumes the presence of a will or emotional desire, for a robot has neither of these; therefore, the work to be done in applying the centuries of philosophical exploration to robotics is in translating the ideas already developed into a new context with a different type of agent.

Whereas autonomy as self-control considers autonomy as a property of the agent itself, there is also a notion in philosophy of autonomy with respect to the surrounding conditions. The general idea is that humans are only truly autonomous if they both have the internal capacity to act and are free from outside strictures that prevent or constrain their action. These issues are often discussed in the context of moral and political autonomy, which require freedom from manipulation that are counter to "one's values, character, and commitments," whether by other individuals, by society, or by some authority [48]. However, an important concept that philosophy also brings to the table is that autonomy and free will are generally considered by philosophers as the requirements for moral responsibility [49], [50]. If this is true not merely as an academic theory but as a human common knowledge expectation, it imposes important expectations on autonomous systems for not only reliable or predictable action, but for moral action as well.

D. Psychology

The American Psychological Association defines cognitive psychology as the study of "how people acquire, perceive, process and store information" [51]. Because autonomous systems at some level attempt to replicate human performance, the utility of understanding human cognitive systems to the design of an artificial cognitive system is clear. While there are many ways in which psychology can improve our description and verification of autonomous systems, three major areas with particular promise are in the area of goal-setting, expert evaluation, and trust.

How humans set, manage, and achieve goals has been a topic of study for many decades, with the intention of improving human performance. The fundamental work in the mechanics of goal setting is useful because it provides a model with defined inputs and outputs based on observations of a complex system (humans) as a black box [52]–[54]. Therefore, it provides both a notional goal-setting framework and a process of evaluation from which robotics may borrow. The challenge in applying the more practical work in the field is that much of it implicitly assumes exclusively human psychology [55], or is focused on how to set goals in a way that improves specifically human performance [56]. One specific contribution that this body of work offers is a number of definitions for goals themselves and the factors that moderate them. Notions of the distinction between reactive and proactive goals [57], task difficulty and goal difficulty, and goal complexity and goal levels [56] are useful in ultimately defining a framework for evaluation of higher-level tasking. One consistent underlying need among all this work, however, is a common understanding of goals, requiring the definition of a taxonomy of specific goals and sub-goals with relationships between them [57], [58].

Another useful analogy for the evaluation of autonomous systems is the evaluation of expertise in humans; autonomous systems are designed to mimic expert performance by humans, so why should they not be assessed in the same way? Human models of general intelligence that were originally anticipated to hold great promise for analogous assessment, turned out to be based on a human-specific taxonomy of fundamental capabilities [59]–[61], and therefore of limited applicability. It also appears that no consensus has been reached on how to measure or represent human intelligence[62], [63], and researchers were seeking more solid models on which to develop tools. More important, however, is the question of whether a general model of evaluation is possible or if each measure of performance must be developed specifically for the system and its intended application. Some work suggests separate evaluation efforts for domain-dependent and domain-independent characteristics [64], while others argue that expert evaluation is only meaningful when applied specifically to the capability under consideration [65], [66]. While both views provide valuable contributions to the notion of assessing autonomous systems, both agree that the majority of the evaluation must be specific to the capability being defined, and that this requires the establishment of a common taxonomy for common evaluation.

The problem of trust is not an analogy for evaluation, but rather one of the few generally agreed-upon characteristics required of an effective autonomous system. The field of psychology has considered both the establishment and maintenance of trust in general and trust between users and robots specifically, and has identified a number of important factors. Several frameworks describing the key attributes of trust both between humans and between human and robot have been proposed [67], [68], and these factors should be a key focus of verification efforts. Other existing work has attempted to measure operator trust in an autonomous system in a collaborative environment, either using a metric based on the fraction of mission time the operator spends directly controlling the system [69] or directly reported based on a questionnaire [70]. The latter case used self-reported trust to track the impact of different types of error on user trust, and showed that the initial impression of performance is persistent, that the degree to which trust is lost correlates with the severity of the error introduced, and that there is an inertial component to trust - a transient error in a proven system will only briefly affect trust, while a persistent error causes trust to decrease to some minimum before gradually recovering. These sorts of dynamics are important to take into account when designing verification methods, since the process is intended to instill trust in a systems operation.

E. Mathematics

Mathematics provides tools to describe and analyze systems under consideration, and as such, this allows us to formally make and justify claims about these particular systems. Once a claim has been proven, the resulting theorem can be used to prove other theorems, or it may be used in an application to assist with the reduction of computational complexity. With respect to intractable complexity, various techniques within mathematics may be used to help reduce the number of test cases in very rigorous and justifiable ways, which are now discussed.

As intractable complexity implies, the number of test cases that one would have to consider in a brute-force manner is uncountably infinite. In mathematics, the notion of equivalence relations can be useful for drastically reducing the number of cases to consider. Essentially, the set that contains all of the possible inputs required by the autonomy is partitioned into subsets that represent inputs which are similar with respect to the equivalence relation. Based upon the equivalence relation, the inputs which are similar can be grouped together into subsets known as equivalence classes, from which only a few elements need only be selected to represent the entire equivalence class. For verification, this means a significant reduction in test cases as only a few elements from each equivalence class need to be selected for testing. The key to this method is defining the equivalence relation, which will determine how the set of inputs will be grouped. This technique has in fact been applied as software verification technique [71]. This technique was applied and found to be better than branch testing and code reading techniques. However, the technique was not found to be as effective as Boundary Value Analaysis [72]. In the boundary value analysis, the boundary is formed using the equivalence relation, suggesting that another important problem becomes edge case identification,which we now discuss.

Another way in which the lessons of mathematics can assist with the intractable complexity problem is to try to identify edge cases. As discussed in [72], the edge cases with respect to the inputs need to be considered. Certain inputs can produce catastrophic effects, and some of these are mere perturbations of perfectly valid inputs that produce stable behavior with respect to autonomy. Catastrophe theory, a subfield of bifurcation theory, is concerned with studying those events in which discrete jumps in behavior occur from small continuous changes in the inputs [73]. This characteristic, known as sensitivity to initial conditions, is one of defining components of a dynamical systems [74]. In particular perturbations of equilibria can produce completely different solutions when compared to the equilibrium solutions, especially if those solutions are unstable. Application of this analogy to autonomy verification then suggests that if some inputs are close to those inputs that produce equilibria solutions for some underlying dynamical system describing the agent-environment interaction, and if those inputs produce drastically different results in autonomy, then these so-called edge cases should be considered when testing. However, more work needs to be performed to show that the dynamical systems assumption is correct. Furthermore, the precise nature of the dynamical system describing the agent-environment interaction needs to be determined as well.

F. Statistics

The field of statistics is concerned with data and how it can be analyzed with respect to scientific processes. In particular, statistical analysis can try to detect which data components are dependent and independent. Once the dependent variables of an experiment or process have been identified, we can restrict our samples to consider only those independent variables, and thus reduce the complexity of our samples. One such technique well known in the field of statistics is known as principal component analysis (PCA). Essentially, PCA identifies which of the possible correlated variables into a set of linearly uncorrelated variables, which are known as principle components. In [75], the authors discuss PCA with an application to chemistry. While their application is specific, the authors give details describing the method, as well as a historical description of the method and which fields use the technique.

With respect to autonomy verification, we may use PCA to help identify which variables of the test cases should be considered. Of course this is only useful when data is available, and when the interactions between the state variables are so complex so as to merit the use of this technique to identify those independent variables that cannot be easily identified otherwise. While this does not necessarily attack the computational complexity directly, the reduction of the redundant or statistically unnecessary variables can help winnow down the unnecessary variables to produce smaller test cases for consideration.

VI. Future Work

By means of an extensive literature review, we have explored the field of autonomy verification, identifying the key needs several opportunities both within the field of robotics and through analogies to other areas of study. Taking all of this into account, we close by charting a path forward, laying out the work that needs to be done to build a meaningful and broadly applicable verification framework for autonomous systems.

A. Understand Stakeholders

Although the technical challenges of autonomy verification are interesting, it is essential to never forget that the end of verification is not the creation of some metric or number, but to communicate information to stakeholders in such a way that it instills trust in the correct operation of the system. As demonstrated in the psychology research, some of the broad ideas of trust are already understood, but specifics exist according to application. We have identified three groups of stakeholders who use the results of verification in different ways: developers, who use it to guide development; customers, who use it to compare options and inform purchasing decisions; and users, who rely on it to determine how much trust to invest in the system during operation. In order to meet the needs of these various stakeholders with any proposed solutions, we intend first to interview representatives from each group and identify their particular needs in the verification process.

B. Define Taxonomy

Much of the literature reviewed in this effort point to the necessity of a common language and organization of ideas to this end, we suggest that the first step of an effort to create a method for verification of autonomous systems be to define a taxonomy of goals, tasks, and capabilities for a current set of common missions, as well as define a framework for managing and adding to that taxonomy. Such an effort must be accomplished in collaboration with other stakeholders in order to avoid the frequent problem of stovepipe solutions that was so frequent in the definition of levels of autonomy. As members of the Verification of Autonomous System Working Group hosted by Naval Research Lab (NRL) and NATO SCI-288 Research Task Group (RTG), we have ongoing and collaborative relationships with other stakeholders in autonomous systems around the world that will provide both feedback as this taxonomy is developed and aid in its adoption once complete.

C. Build Framework

As a taxonomy provides a common language and organization of concepts, evaluation of autonomous systems will require some logical organization of capabilities according to their different functions. We propose that, by developing well-defined categories for specific capabilities defined in the taxonomy, components which are similar can use the same tools for evaluation, decreasing the effort needed for verification. Because there have been numerous frameworks proposed for the technical design of autonomous systems, we are careful to specify that this conceptual framework will be designed to represent the capabilities of the system and guide evaluation rather than represent the technical implementation and guide system design.

D. Apply Metrics

Using the taxonomy and framework, we plan to identify meaningful cases and apply any derived or developed metrics to those cases. In most cases, we will consider metrics from our analogies work, and apply them to a few cases under consideration. The use of the framework may prove to be very critical for this stage. The framework will allow us to identify those well-defined categories based upon categorization defined by the taxonomy and potentially consider individual metrics for each category, which can then be aggregated into some other metric representing total system performance. This step should allow us to exercise and verify the metrics, as well as our approach to autonomy verification that relies heavily upon the taxonomy and framework.

VII. Conclusion

As the problem of autonomy verification is a difficult problem, we have presented findings from literature reviews of the field of verification, as well as other analogous fields with the goal of developing a way forward with respect to autonomy verification. As intractable complexity makes traditional methods for autonomy verification difficult, we have sought examples and analogies from the fields of philosophy, psychology, and mathematics. These fields have provided useful guidance that we intend to employ by developing a taxonomy of concepts related to verification, as well as a conceptual framework. Further investigation of existing metrics and new metrics will allow us to test and verify our approach to autonomy verification and is the subject of future work.

References

[1] A. Dasson and A. Funes, *Verification, Validation and Testing in Software Engineering*. Idea Publishing Group, 2007.

[2] W. J. Brown, *AntiPatterns: Refactoring Software, Architectures and Projects in Crisis*. Wiley, 1998.

[3] M. Newborn, *Automated Theorem Proving Theory and Practice*. Springer-Verlag, 2001.

[4] K. Appel and W. Haken, “Every planar map is four colorable part 1: Discharging,” *Illinois J. Math*, vol. 21, pp. 429–490, 1977.

[5] H.-M. Huang, J. S. Albus, E. R. Messina, R. L. Wade, and R. English, “Specifying autonomy levels for unmanned systems: Interim report,” in *Defense and Security*. International Society for Optics and Photonics, 2004, pp. 386–397.

[6] H.-M. Huang, “Autonomy levels for unmanned systems (ALFUS) framework: safety and application issues,” in *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems*. ACM, 2007, pp. 48–53.

[7] G. T. McWilliams, M. A. Brown, R. D. Lamm, C. J. Guerra, P. A. Avery, K. C. Kozak, and B. Surampudi, “Evaluation of autonomy in recent ground vehicles using the autonomy levels for unmanned systems (alfus) framework,” in *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems*. ACM, 2007, pp. 54–61.

[8] B. Hasslacher and M. W. Tilden, “Living machines,” *Robotics and autonomous systems*, vol. 15, no. 1, pp. 143–169, 1995.

[9] B. T. Clough, “Metrics, schmetrics! how the heck do you determine a UAV’s autonomy anyway,” DTIC Document, Tech. Rep., 2002.

[10] P. J. Durst, W. Gray, A. Nikitenko, J. Caetano, M. Trentini, and R. King, “A framework for predicting the mission-specific performance of autonomous unmanned systems,” in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 1962–1969.

[11] H. Chen, X.-m. Wang, and Y. Li, “A survey of autonomous control for uav,” in *Artificial Intelligence and Computational Intelligence, 2009. AICI’09. International Conference on*, vol. 2. IEEE, 2009, pp. 267–271.

[12] E. Sholes, “Evolution of a uav autonomy classification taxonomy,” in *Aerospace Conference, 2007 IEEE*. IEEE, 2007, pp. 1–16.

[13] R. Murphy and J. Shields, “The role of autonomy in DoD systems,” DoD Defense Science Board, Tech. Rep., 2012.

[14] J. D. Gehrke, “Evaluating situation awareness of autonomous systems,” in *Performance Evaluation and Benchmarking of Intelligent Systems*. Springer, 2009, pp. 93–111.

[15] F. Macias, “The test and evaluation of unmanned and autonomous systems,” DTIC Document, Tech. Rep., 2008.

[16] M. Castillo-Effen and N. A. Visnevski, “Analysis of autonomous deconfliction in unmanned aircraft systems for testing and evaluation,” in *Aerospace conference, 2009 IEEE*. IEEE, 2009, pp. 1–12.

[17] J. S. Albus, H. Hui-Min, M. E. R., K. Murphy, M. Juberts, A. Lacaze, S. B. Balakirsky, M. O. Shneier, T. H. Hong, H. A. Scott, F. M. Proctor, W. P. Shackleford, J. L. Michaloski, A. J. Wavering, T. R. Kramer, N. G. Dagalakis, W. G. Rippey, k. A. Stouffer, and S. legowik, “4d/rcs version 2.0: A reference model architecture for unmanned vehicle systems,” NIST, Tech. Rep., 2002.

[18] L. Billman and M. Steinberg, “Human system performance metrics for evaluation of mixed-initiative heterogeneous autonomous systems,” in *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems*. ACM, 2007, pp. 120–126.

[19] D. F. Woerner, “The selection and infusion of autonomy for mars rovers,” in *Intelligent Systems, 2002. Proceedings. 2002 First International IEEE Symposium*, vol. 1. IEEE, 2002, pp. 86–91.

[20] F. Alonso, J. L. Fuertes, L. Martínez, and H. Soza, “Towards a set of measures for evaluating software agent autonomy,” in *Eighth Mexican International Conference on Artificial Intelligence, 2009. MICAI 2009.* IEEE, 2009, pp. 73–78.

[21] D. Gertman, C. McFarland, T. Klein, A. Gertman, and D. Bruemmer, “A methodology for testing unmanned vehicle behavior and autonomy,” in *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems*. ACM, 2007, pp. 62–69.

[22] M. Steinberg, “Intelligent autonomy for unmanned naval systems,” in *Defense and Security Symposium*. International Society for Optics and Photonics, 2006, pp. 623 013–623 013.

[23] J. M. Bradshaw, R. R. Hoffman, M. Johnson, and D. D. Woods, Florida Institute for Human and Machine Cognition, Tech. Rep., 2013.

[24] J. F. Porter, K. Cheung, J. A. Giampapa, and J. M. Dolan, “A reliability analysis technique for estimating sequentially coordinated multirobot mission performance,” in *PRIMA 2013: Principles and Practice of Multi-Agent Systems*. Springer, 2013, pp. 276–291.

[25] A. Lampe and R. Chatila, “Performance measure for the evaluation of mobile robot autonomy,” in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 4057–4062.

[26] G. S. Sukhatme and G. A. Bekey, “An evaluation methodology for autonomous mobile robots for planetary exploration,” in *Proc. of the First ECPD International Conference on Advanced Robotics and Intelligent Automation*. Citeseer, 1995, pp. 558–563.

[27] F. Ingrand, S. Lacroix, S. Lemai-Chenevier, and F. Py, “Decisional autonomy of planetary rovers,” *Journal of Field Robotics*, vol. 24, no. 7, pp. 559–580, 2007.

[28] C. Pecheur, “Verification and validation of autonomy software at NASA,” 2000.

[29] K. J. Barltrop, K. H. Friberg, and G. A. Horvath, “Automated generation and assessment of autonomous systems test cases,” in *Aerospace Conference, 2008 IEEE*. IEEE, 2008, pp. 1–10.

[30] J. Rushby, “Just-in-time certification,” in *Engineering Complex Computer Systems, 2007. 12th IEEE International Conference on*. IEEE, 2007, pp. 15–24.

[31] K. Eder, C. Harper, and U. Leonards, “Towards the safety of human-in-the-loop robotics: Challenges and opportunities for safety assurance of robotic co-workers’,” in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*. IEEE, 2014, pp. 660–665.

[32] Z. Kong, V. Korukanti, and B. Mettler, “Mapping 3d guidance performance using approximate optimal cost-to-go function,” in *AIAA Navigation, Guidance and Control Conference, Chicago, IL*, 2009.

[33] Z. Kong and B. Mettler, “Evaluation of guidance performance in urban terrains for different uav types and performance criteria using spatial ctg maps,” *Journal of Intelligent & Robotic Systems*, vol. 61, no. 1-4, pp. 135–156, 2011.

[34] T. Smithers, "On quantitative performance measures of robot behaviour," in *The Biology and Technology of Intelligent Autonomous Agents*. Springer, 1995, pp. 21–52.
[35] U. Nehmzow, "Quantitative analysis of robot–environment interaction-towards scientific mobile robotics," *Robotics and Autonomous Systems*, vol. 44, no. 1, pp. 55–68, 2003.
[36] M. Bozzano, A. Cimatti, M. Roveri, and A. Tchaltsev, "A comprehensive approach to on-board autonomy verification and validation," in *IJCAI*, vol. 11, 2011, pp. 2398–2403.
[37] B. Smith, W. Millar, J. Dunphy, Y.-W. Tung, P. Nayak, E. Gamble, and M. Clark, "Validation and verification of the remote agent for spacecraft autonomy," in *Aerospace Conference, 1999. Proceedings. 1999 IEEE*, vol. 1. IEEE, 1999, pp. 449–468.
[38] R. E. Brown, L. Stanford, and H. M. Schellinck, "Developing standardized behavioral tests for knockout and mutant mice," *ILAR Journal*, vol. 41, no. 3, pp. 163–174, 2000.
[39] L. Asher, S. Blythe, R. Roberts, L. Toothill, P. J. Craigon, K. M. Evans, M. J. Green, and G. C. England, "A standardized behavior test for potential guide dog puppies: Methods and association with subsequent success in guide dog training," *Journal of Veterinary Behavior: Clinical Applications and Research*, vol. 8, no. 6, pp. 431–438, 2013.
[40] J. M. Koolhaas, C. M. Coppens, S. F. de Boer, B. Buwalda, P. Meerlo, and P. J. Timmermans, "The resident-intruder paradigm: a standardized test for aggression, violence and social stress," *Journal of visualized experiments: JoVE*, no. 77, 2013.
[41] J. M. Geringer and M. D. Worthy, "Effects of tone-quality changes on intonation and tone-quality ratings of high school and college instrumentalists," *Journal of Research in Music Education*, vol. 47, no. 2, pp. 135–149, 1999.
[42] E. Ekholm, G. C. Papagiannis, and F. P. Chagnon, "Relating objective measurements to expert evaluation ofvoice quality in western classical singing: Critical perceptual parameters," *Journal of Voice*, vol. 12, no. 2, pp. 182–196, 1998.
[43] J. Wapnick and E. Ekholm, "Expert consensus in solo voice performance evaluation," *Journal of Voice*, vol. 11, pp. 429–436, 1997.
[44] I. Berlin, "Concepts and categories," 1978.
[45] B. Ramberg and G. K., "Hermeneutics," 2005.
[46] A. R. Mele, *Autonomous Agents: From Self-Control to Autonomy*. Wiley, 1995.
[47] B. Bernofsky, "Liberation from self: A theory of personal autonomy," *The Philosphical Review*, vol. 106, pp. 599–601, 1997.
[48] J. Christman, "Autonomy in moral and political philosophy," *Stanford encyclopedia of philosophy*, 2008.
[49] G. Dworkin, *The Theory and Practice of Autonomy*. Cambridge University Press, 1988.
[50] R. Clarke, "Free will and the conditions of moral responsibility," *Philophical Studies*, vol. 66, pp. 53–72, 1992.
[51] "Brain science & cognitive psychology explores our mental processes," http://www.apa.org/action/science/brain-science/index.aspx, 2015, accessed: 2015-07-17.
[52] A. Juarrero, *Dynamics in action: Intentional behavior as a complex system*. Citeseer, 1999, vol. 31.
[53] E. A. Locke and G. P. Latham, "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey." *American psychologist*, vol. 57, no. 9, p. 705, 2002.
[54] L. Horváth and I. J. Rudas, "Human intent representation in knowledge intensive product model," *Journal of Computers*, vol. 4, no. 10, pp. 954–961, 2009.
[55] A. Sloman, "Motives, mechanisms, and emotions," *Cognition and Emotion*, vol. 1, no. 3, pp. 217–233, 1987.
[56] E. A. Locke, K. N. Shaw, L. M. Saari, and G. P. Latham, "Goal setting and task performance: 1969–1980." *Psychological bulletin*, vol. 90, no. 1, p. 125, 1981.
[57] T. J. Norman and D. Long, "Goal creation in motivated agents," in *Intelligent Agents*. Springer, 1995, pp. 277–290.
[58] K. Ishii and M. Sugeno, "A model of human evaluation process using fuzzy measure," *International Journal of Man-Machine Studies*, vol. 22, no. 1, pp. 19–38, 1985.
[59] R. W. Woodcock, "Theoretical foundations of the wj-r measures of cognitive ability," *Journal of Psychoeducational Assessment*, vol. 8, no. 3, pp. 231–258, 1990.
[60] E. Benson, "Intelligent intelligence testing," *Monitor on Psychology*, vol. 34, no. 2, pp. 48–58, 2003.
[61] W. J. Schneider and K. S. McGrew, *The Cattell-Horn-Carroll model of intelligence*. Guilford Press, 2012.
[62] T. Z. Keith, J. H. Kranzler, and D. P. Flanagan, "What does the cognitive assessment system (cas) measure? joint confirmatory factor analysis of the cas and the woodcock-johnson tests of cognitive ability," *School Psychology Review*, vol. 30, no. 1, pp. 89–119, 2001.
[63] M. C. Frey and D. K. Detterman, "Scholastic assessment or g? the relationship between the scholastic assessment test and general cognitive ability," *Psychological science*, vol. 15, no. 6, pp. 373–378, 2004.
[64] R. O'Keefe and D. E. O'Leary, "Expert system verification and validation: a survey and tutorial," *Artificial Intelligence Review*, vol. 7, pp. 3–42, 1993.
[65] K. A. Ericsson and N. Charness, "Expert performance: Its structure and acquisition." *American psychologist*, vol. 49, no. 8, p. 725, 1994.
[66] K. A. Ericsson *et al.*, "The influence of experience and deliberate practice on the development of superior expert performance," *The Cambridge handbook of expertise and expert performance*, pp. 683–703, 2006.
[67] H.-W. Bierhoff and B. Vornefeld, "The social psychology of trust with applications in the internet," *Analyse & Kritik*, vol. 26, no. 1, pp. 48–62, 2004.
[68] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 53, no. 5, pp. 517–527, 2011.
[69] E. Freedy, E. DeVisser, G. Weltman, and N. Coeyman, "Measurement of trust in human-robot collaboration," in *Collaborative Technologies and Systems, 2007. CTS 2007. International Symposium on*. IEEE, 2007, pp. 106–114.
[70] J. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, no. 10, pp. 1243–1270, 1992.
[71] N. Juristo, S. Vegas, M. Solari, S. Abrahao, and I. Ramos, "Comparing the effectiveness of equivalence partitioning, branch testing and code reading by stepwise abstraction applied by subjects," *2012 IEEE Fifth International Conference on Software Testing, Verification and Validation*, pp. 330–339, 2012.
[72] S. C. Reid, "An empirical analysis of equivalence partitioning, boundary value analysis and random testing," *Fourth International Software Metrics Symposium*, pp. 64–73, 1997.
[73] A. G. Wilson, *Catastrophe Theory and Bifurcation*. Croom Helm, 1981.
[74] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley, 1989.
[75] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, pp. 37–52, 1987.