To appear as a **Short Note** in *New Mathematics and Natural Computation: Special Issue on Engineering of the Mind, Cognitive Science and Robotics*

Behavior Prediction for Decision & Control in Cognitive Autonomous Systems★

Asok Ray axr2@psu.edu
Shashi Phoha sxp26@psu.edu
Soumik Sarkar† soumikpsu@gmail.com

Pennsylvania State University, University Park, PA 16802, USA

***Abstract*—This short note presents an innovative concept of behavior prediction for decision & control in cognitive autonomous systems. The objective is to coordinate human-machine collaboration such that human operators can assess and enable autonomous systems to utilize their experiential & unmodeled domain knowledge and perception for mission execution. The concept of quantum probability is proposed to construct a unified mathematical framework for interfacing between models of human cognition and machine intelligence.**

I. Introduction

Modern human-engineered systems (e.g., power grid, communication, transportation, and smart building) with high degree of autonomy are becoming increasingly complex due to subsystem heterogeneity, uncertain operational & environmental dynamics, and decentralization of decision & control [1]. In spite of embedded intelligence and high degree of autonomy of current and future-generation autonomous systems, it is now recognized that the issues of human factors must be taken into consideration during design, development and operation of such systems [2]. The rationale is that there are several sources of disparities between human cognition and machine intelligence of autonomous systems. These disparities often stem from incompatible internal representations of information, structural characteristics of logic & reasoning, and learning & inference mechanisms. Along this line, Busemeyer et al. [3] have shown that many of the disparities arise due to the use of classical probabilistic axioms in modeling the cognition process. For example, Kolmogorov probabilistic logic is often incompatible with human decision-making if it requires relaxation of the commutative, distributive and closure properties [4] of the underlying cognitive system.

Drawing upon the principles of Quantum Mechanics, multiple perspectives of events can be simultaneously represented as vectors in a Hilbert space over the complex field $\mathbb{C}$. In this setting, if a vector is expressed in different bases, then it may represent different perspectives of an event. It is hypothesized that these events can be synthesized by making use of quantum probability theory to model human perception and decision-making. Recent research publications [3] reveal that quantum-theoretic principles (e.g., superposition of interferences in modeling) and their implementation are compatible with psychological intuitions and concepts of human cognition and decision-making; however, the problem of interfacing between human cognition and machine intelligence is still an open research issue. This short note proposes an innovative concept of behavior prediction for decision & control of cognitive autonomous systems, where a unified mathematical framework is presented from the perspectives of:

- resolving the disparities between human cognition models and machine intelligence models, and
- developing analytical transformations for decision & control at different hierarchical levels of fidelity.

The objective here is to enable human operators to gain insight into behavior prediction and to make machines to unambiguously interpret human instructions, while assessing and enabling autonomous systems to utilize their experiential & unmodeled domain knowledge and perception for mission execution. The major challenge is to achieve and sustain tradeoffs between coherence and performance for operational dependability, which would require addressing unresolved fundamental problems such as aggregation of human and machine decision models.

II. Scientific Approach

Recent literature (e.g., [5]) advocates two main approaches in dealing with interactive and multi-time-scale dynamics of multi-agent autonomous systems that are operated by human agents with diverse training levels and job responsibilities. The first approach, exemplified by Reinforcement Learning and Markov Decision Processes (MDP) [6], attempts to model the computational dynamics in terms of human-understandable heuristics (e.g., reward-based actions) to endow a machine with human-like decision-making capabilities. The second approach [7] is based on mimicking biological cognitive mechanisms through computational intelligence. Both these approaches have their individual limitations.

This short note introduces an interdisciplinary concept that would utilize the complementary processing and synthesizing capabilities of humans and machines in a common mathematical framework. The proposed approach is expected to make autonomous systems trustworthy and dependable members of tightly knit operational units, where humans and machines learn from each other in-situ to improve the mission performance. The underlying framework is depicted in Fig. 1 that entails research issues of model unification, transformation and

★ This work has been supported in part by the Army Research Laboratory (ARL) and the Army Research Office (ARO) under Grant No. W911NF-07-1-0376. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

†Currently with the United Technologies Research Center, East Hartford, CT 06108, USA.

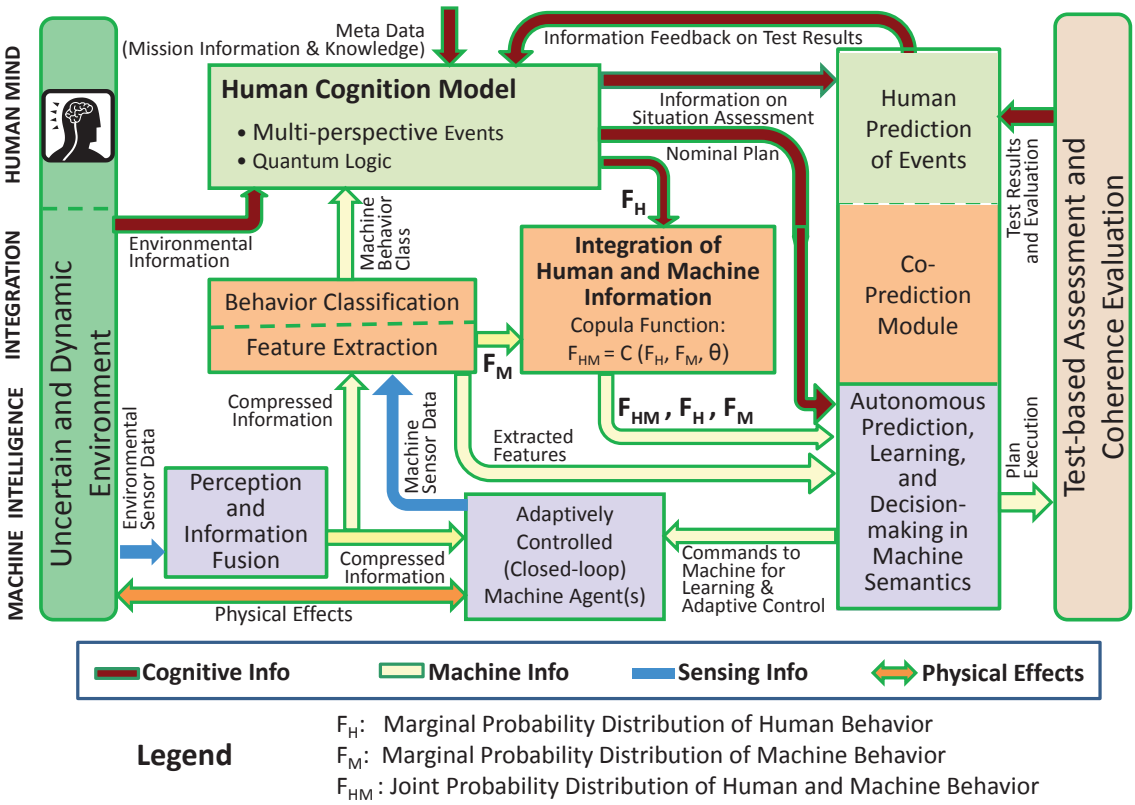

Fig. 1. Schematic for Decision & Control in Cognitive Autonomous Systems

construction of co-dependence and coherence measures as well as identification of precursors for emergent discords.

A. Human Cognition and Machine Intelligence Modeling

The concept of human cognition modeling, proposed in this note, draws upon recent developments in Cognitive Science, which rely on quantum dynamic principles for constructing human cognition models of autonomous systems [3]. For example, in quantum probability theory, the state of a system is determined by a complex-valued function, called the amplitude function $\mathcal{A}$, on the outcome (or sample) space. If the outcomes x_i of an event $E = \{x_1, x_2, \cdots\}$ do not interfere with each other, then the probability of the event E is computed as:

$$P(E) = \sum_k \left|\mathcal{A}(x_k)\right|^2 \tag{1}$$

otherwise (i.e., in the presence of interferences),

$$P(E) = \left|\sum_k \mathcal{A}(x_k)\right|^2 \tag{2}$$

In classical probability theory, the probability of an event is obtained by Eq. (1) as the sum of the probabilities of the sample outcomes composing the event. In quantum probability theory, the probability of an event is obtained by Eq. (2), where the event probability $P(E)$ can be decomposed in two parts. The first part is classical probability in Eq. (1) and the second part consists of the cross terms that represent constructive and destructive interferences similar to the characteristics of wave phenomena in Quantum Mechanics [4]. It is hypothesized that the human cognition model in Fig. 1 can be developed in the framework of Eq. (2) [3].

The tools of quantum probability for human cognition modeling are to be developed on a Hilbert space, defined over the complex field $\mathbb{C}$ and spanned by (finitely many) orthonormal basis vectors [3]. As seen in Eqs. (1) and (2), the probability mass functions are obtained as the squared moduli of the coefficients of the quantum state vectors. The resulting family of (multivariate) distributions, denoted as $F_H(\bullet)$, represents marginal probabilities of human behavior as depicted in Fig. 1. The objective here is to construct a mathematical framework for interfacing behaviorial models of human agents with those of machine agents.

Referring to Fig. 1, two of the inputs to the human cognition model, namely, environmental information and machine behavior class, are combined to form a tensor product space [3], [4], and the human belief state is represented as a vector on this space. It is hypothesized that the remaining inputs to the human cognition model, namely, meta data (e.g., mission information and knowledge base), and the feedback information from test results, can be combined to assign amplitudes to each basis vector [3]. A unitary operator could be used to transform one set of basis vectors to another for evaluating the state vectors from different perspectives to represent multiple mission objectives. The key idea here is to map a state vector, defined on the space of a human behavior model, into the space of machine systems by projecting the state vector onto

the subspaces representing the pertinent action categories.

Machine intelligence models have been developed as probabilistic finite state automata (PFSA) that belong to a class of Markov decision processes (MDPs) [8]. Wen et al. [9] and Adenis et al. [10] have shown that a class of PFSA forms a Hilbert space over the real field $\mathbb{R}$. In this setting, the norm of a PFSA (induced by the inner product) is interpreted as a measure of the information content of the (time series) data represented by the PFSA. This formalism has established mathematical properties that make it inherently suitable for in-situ behavior prediction of autonomous systems. In particular, such models are robust to noise and exogenous disturbances, provide order reduction in the sense of maximum entropy, and capture structural nonlinearities with no significant loss of information (e.g., the loss at an infinite horizon tending to zero). Therefore, it is logical to build machine intelligence models in the PFSA setting instead of solely relying on quantum probability theory (QPT). The family of multivariate distributions of machine PFSA models, denoted by $F_M(\bullet)$ as depicted in Fig. 1, is suitable for representing interactions with human models.

It is necessary to establish a common mathematical framework for interactions of the human and machine models, where $F_H(\bullet)$ and $F_M(\bullet)$ represent respective families of marginal distributions of human and machine behaviors. It would require integration of the above two classes of Hilbert spaces, which entail different algebraic structures as they are defined over two different fields $\mathbb{C}$ and $\mathbb{R}$, respectively. The integration could be accomplished by constructing a topological imbedding [11] from the low-dimensional space of the human cognition model to the high-dimensional space of machine intelligence model, where the image of the imbedding function is a (closed) subspace of PFSA models that are (unique) fixed points in the machine dynamics under a given mission plan. The imbedding could be used to construct a homeomorphism between these two subspaces. High co-predictability measures identify corresponding subspaces of the two Hilbert spaces as regions of coherence, where humans can dependably predict behaviors of the autonomous system that, in turn, can reliably follow human instructions. In this setting, identification of thresholds for varying degrees of trust would require domain-specific testing and training as explained later in Section III.

B. Coupling of Human and Machine Decision Spaces

Analytical relationships need to be formulated between event/action sequences (equivalently, the state transition vectors in the respective Hilbert spaces) in the human cognition and machine intelligence models. It would also require in-situ discovery of regions of model coherence for achieving acceptable performance under noisy and uncertain conditions with incomplete information. Specifically, regions of structural coherence need to be defined in the human and machine event/action spaces and the critical parameters must be identified. It is recognized that subsequent behaviors of autonomous systems may significantly deviate from human predictions.

The framework depicted in Fig. 1 adopts a constructive method of measuring co-dependence of two emerging sequences. This approach has been developed to formulate co-prediction algorithms that measure causal co-dependency between symbol sequences via finite state probabilistic transducers to capture statistically significant inter-dependencies [12][13]. It is noted that human and machine system dynamics take place at different time scales and may use different alphabets for symbolic representations that are not synchronized. Yet they both model the same physical phenomena, albeit as models of different order. Co-dependence coefficients of event/action sequences, represented by vectors in each of the two above-mentioned Hilbert spaces, are defined as reduction in entropy of the next symbol distribution of one trajectory by observing the other.

Statistical characterization of human-machine interactions requires derivation of the joint distribution between such heterogeneous systems. Conventional models, such as multivariate Gaussian distributions, are inadequate for two main reasons: (i) F_H and F_M are likely to be disparate distribution families, and (ii) the dependence structure between F_H and F_M may be nonlinear. Recently, Iyengar et al. [14] have shown how the copula theory can be used to construct models of heterogeneous random variables that have disparate distributions. Besides the copula theory, there are other viable tools (e.g., cross machines [13]) that should be investigated for this purpose.

C. Prediction, Learning, and Adaptation

Autonomous systems considered in this note are expected to be endowed with capabilities of distributed learning and adaptive control. This would require development of algorithms for consistency of machine intelligence models with human cognition and belief models. Furthermore, the algorithms should be able to predict and mitigate emerging discord if the human operator alters the reference trajectory under unmodeled disturbances. Referring to Fig. 1, there are feedback loops at three levels of hierarchy. The outermost loop involves four inputs, namely, environmental information, meta data & knowledge, machine performance, and test results. Based on these inputs, the human may alter his/her actions that, in turn, may cause changes in the outputs of the human cognition model as seen in Fig. 1. The middle loop provides adaptive control of machines that operate with the built-in inner most closed-loop control of the autonomous agents. Cross-disciplinary concepts borrowed from diverse disciplines (e.g., Statistical Mechanics and Multi-fractals), can be used to address adaptation issues (e.g., performance and stability robustness) in the multi-time-scale operations of interacting autonomous/manned teams. The built-in control system in the innermost closed-loop is not explicitly addressed in this note, because this issue has been extensively dealt with in the scientific literature on decision & control theory.

III. Test Methods & Assessment

Test-based verification & validation is a crucial part of the integrated human-machine framework. Tests need to be performed in distributed settings under realistic assumptions of cognitive limitations in human decision-making. For example,

humans may maintain quantized priors regarding objects or events, which have implications on making inferences in a Bayesian framework. Some of the key objectives of the testing and assessment process are identified below.

A. Emergent Behavior Characterization

A crucial step towards verification & validation is characterization of emergent behaviors of distributed human-machine systems. Observable changes may take place in the interactive dynamics of an autonomous system prior to the appearance of a critical emergent behavior (e.g., a phase transition). Therefore, the operational parameters need to be extended during testing to capture various phase transition phenomena and to determine the critical thresholds at which desirable behaviors of the autonomous system may begin to break down unpredictably. It is hypothesized that these thresholds represent the boundaries of coherence regions beyond which the system behavior becomes more unpredictable. Co-prediction analysis with corresponding trajectories in the quantum space of human cognition is an open area of research to resolve these hypotheses and to determine event/action trajectories that provide precursors for the emergence of such phase transition phenomena. For example, probabilistic finite-state automata (PFSA) have been constructed from quasi-stationary time series data for early detection of incipient faults in diverse dynamical systems [12][13].

B. Assessment of Transfer Learning

Transfer learning, studied extensively in Cognitive Psychology [2][7], is another relevant issue in human-machine systems that is more complex than conventional learning. In transfer learning, the cognitive outputs of previously learnt elements are evoked and subsequently applied to somewhat different situations with sufficiently similar stimulus characteristics. For example, system behaviors are less predictable when algorithms learnt in one context do not transfer well to other contexts. The main objectives of testing from these perspectives are:

- Determination of the contexts in which the expected transfer does not happen.
- Analysis of action inconsistencies that are likely to occur.
- Determination of environmental factors for which the autonomous system has transfer variations.

C. Analysis of Dependability

Tests are needed for dependability analysis perceived as the dual to reinforcement learning [6] that searches over actions in a given state for an optimal reward policy. Among the available test methods, morphological analysis is apparently a viable method that alternates between analysis of related patterns in two data streams and synthesis of the solution region obtained by pruning the large-dimensional input space [15]. It focuses on totality of the relationships contained in multi-dimensional, non-quantifiable problem complexes, where traditional quantitative methods (e.g., causal modeling and simulation) may not be sufficient, because inherent uncertainties may not be readily reducible and are often ill-defined.

IV. Summary, Conclusions and Future Work

Despite significant advances in both fields of Artificial Intelligence and Machine Learning, autonomous systems remain heavily dependent on human operators for reasoning and decision-making, which add nontrivial cognitive loads to (possibly overburdened) human operators. On the other hand, humans may not trust machine's decisions in critical situations, more so when they deviate from their own perception. This issue is addressed in the short note, where the concept of a mathematically structured interface is aimed at integrating autonomous systems as dependable members of tightly knit operational units in diverse applications. Apart from an introduction to basic human-machine integration, this note also presents pertinent functional requirements of testing and validation, which are indispensable for robustness and resilience of human-machine collaboration.

Future research should continue on the confluence of computational, physical and neuro-sciences, for meaningful man/machine collaboration in the direction that bears the promise for providing solutions to unresolved problems such as persistent surveillance in critical missions.

References

[1] E. Lee, "Computing foundations and practice for cyber-physical systems: A preliminary report," University of California, Berkeley, CA, USA, UCB/EECS-2007-72,, Tech. Rep., May 2007.

[2] G. Salomon and D. Perkins, "Rocky roads to transfer: rethinking mechanism of a neglected phenomenon," *Educational Physiologist*, vol. 24, no. 2, pp. 113–142, 1989.

[3] J. Busemeyer and P. Bruza, *Quantum Models of Cognition and Decision*. Cambridge University Press, Cambridge, UK, 2012.

[4] S. Gudder, *Quantum Probability*. Academic Press, Boston, MA, USA, 1988.

[5] W. Knox and P. Stone, "Interactively shaping agents via human reinforcement," in *Proceedings of The Fifth International Conference on Knowledge Capture (K-CAP 09), Redondo Beach, California, USA*, September 2009.

[6] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA, 1998.

[7] K. Holyoak and R. Morrison, *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press, Cambridge, U.K., 1988.

[8] E. Vidal, F. Thollard, C. de la Higuera, F. Casacuberta, and R. Carrasco, "Probabilistic finite-state machines– Part I," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, pp. 1013–1025, 2005.

[9] Y. Wen, A. Ray, I. Chattopadhyay, and S. Phoha, "Modeling of Symbolic Systems: Part II - Hilbert space construction for model identification and order reduction," in *Proceedings of American Control Conference, San Francisco, CA, USA*, June-July 2011, pp. 5139–5144.

[10] P. Adenis, Y. Wen, and A. Ray, "An inner product space on irreducible and synchronizable probabilistic finite state automata," *Math. Control Signals Syst.*, vol. 23, no. 4, pp. 281–310, 2012.

[11] J. Munkres, *Topology, 2nd ed.* Prentice Hall, upper Saddle River, NJ, USA, 2000.

[12] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Processing*, vol. 84, no. 7, pp. 1115–1130, July 2004.

[13] S. Sarkar, S. Sarkar, K. Mukherjee, A. Ray, and S. Srivastav, "Multi-sensor information fusion for fault detection in gas turbine engines," *I. Mech E Part G; Journal of Aerospace Engineering*, p. doi: 10.1177/0954410012468391, 2013.

[14] S. Iyenger, P. Varshney, and T. Damarla, "A parametric copula-based framework for hypothesis testing using heterogeneous data," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2308–2319, May 2011.

[15] J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Trans. Graph.*, vol. 5, no. 2, pp. 110–141, 1986.