QUANTUM COGNITION MODELS OF ETHICAL DECISION-MAKING IN HUMAN BEHAVIOR SIMULATION

Levent Yilmaz

Department of Computer Science and Software Engineering *and*
Department of Industrial and Systems Engineering
Samuel Ginn College of Engineering
Auburn University

yilmaz@auburn.edu

ABSTRACT
Ethical decision-making is a unique aspect of human behavior. When confronted with situations that require careful deliberation over multitude of options that have ethical implications, human behavior tends to resolve dilemmas by resorting to a range of heuristics and principles that view the situation from different perspectives. While constraint and utility-driven decision-making strategies are relevant, the incompatibility among these perspectives can invalidate the underlying premises of models of probabilistic utility-based decisions that rely on classic Kolmogorov axioms. In this paper, it is posited that quantum cognition models can provide an alternative and credible representation of human behavior modeling in simulations that involve ethical decision-making.

Keywords: ethical decision-making, quantum cognition, machine ethics, computational ethics

1. INTRODUCTION

Ethical dilemmas often arise due to conflicts among different facets observed in a situation (Anderson and Anderson, 2011; Wallach and Allen, 2009). The consequences, obligations, duties, and principles present incompatibilities that cannot be simultaneously evaluated to make judgments based on joint assessments. For instance, consider humanitarian response and recovery operations, which require situation awareness and allocation of resources. Is the objective to increase the "number of lives" or "the number of years of life" saved. Also, to what extend the decision-making capacity need to consider the necessity to save the lives of the first responders? Acceptable decisions in such situations rely on the priorities distributed over the principles adhered to (Arkoudas and Bello, 2005) and the consequences expected (Gips, 1995; Gert, 1998). The resolution of conflicts among such competing decisions rely on strategies, which facilitate bringing the uncertainty and ambiguity into a conclusion (Yilmaz et al., 2016; Yilmaz et al., 2017).

The resolution of such conflicts is complicated when the dilemma involves multiple decisions that are framed from the perspective of different stakeholders. In the presence of multiple competing perspectives, the order in which each perspective is evaluated can influence the outcome due to potential interference between judgements. In an ethical dilemma, the decision-maker may start the evaluation from the perspective of an organization that prioritize specific rules and principles, which may be incompatible with the consequentialist view that is prioritized from an individual's perspective. Yet another perspective can be based on the propositions that involve cost-benefit analysis. In the presence of multiple perspectives, judgments can be incompatible in that the set of features necessary to encode and evaluate one perspective may not be shared by the set of attributes used to evaluate another (Busemeyer and Bruza, 2014). Therefore, no common set of features are available to jointly evaluate information in different perspectives.

Quantum causal reasoning and inference models (Bruza et al., 2015) have been put forward to account for a variety of such cognitive phenomena, including interference effects. Quantum models of cognition (Busemeyer and Bruza, 2014) postulate that the state of a cognitive system is undetermined and stays in an indefinite superposition state, reflecting the conflict and uncertainty that is akin to the state of ambiguity observed in ethical dilemmas. Beliefs stay in a superimposed state until a decision must be reached. Making a decision is then the process of transforming a thought wave into a particle in a quantum model. Furthermore, judgements with respect to a perspective create a context that influences the construction of an opinion. As such, quantum models do not require adherence to the principle of unicity. That is, because incompatible perspectives cannot be evaluated on the same basis, they require constructing distinct, but possibly related sample spaces, without complete joint assessments.

In the rest of this paper, the plausibility of using quantum cognition models in modeling ethical dilemmas is examined to bring resolution to cognitive conflicts in accordance with the principles of quantum causal reasoning and inference. We demonstrate the application of quantum causal reasoning using a practical example that involves an ethical dilemma viewed from multiple perspectives. Following the illustration of the strategy, we evaluate the utility of the strategy, discuss its limits,

and conclude by discussing potential areas of further inquiry.

2. BACKGROUND

The field of machine ethics has generated wealth of knowledge, including methods, tools, and strategies for engineering artificial systems that can exhibit the characteristics of moral behavior (Wallach and Allen, 2009; Herman, 2014). As a corollary, the use of computational methods is facilitating both the identification of gaps and the advancement of normative theories of ethical decision-making. Machine ethics has emerged as a field to explore the nature, issues, and computational approaches for ascribing ethical values to agents. Recently, AI-assisted ethics (Etzioni and Etzioni, 2016) is promoted as a perspective to better understand the ethical norms and principles that individuals adhere to so as to provide effective guidance for human-agent interaction.

2.1. Computational Ethics

The most commonly used AI methods in computational ethics are planning, deductive reasoning with deontic logic, analogical reasoning, constraint satisfaction, decision-theory, and (social) learning. In (Kurland, 1995), the role of theory of planned behavior and reasoned action in predicting ethical intentions towards others is demonstrated in experiments with human subjects. Besides planning and goal-directed behavior, deductive logic is proposed (Saptawijaya and Pereira, 2012) to infer judgments and their ethical implications. In (Arkoudas and Bello, 2005), authors demonstrate the use of deontic logic in formalizing ethical codes and discuss the utility of mechanized formal proofs for instilling trust in autonomous systems. Analogy-based reasoning has also emerged as an effective method to address practical problems. For instance, the Truth-Teller system (McLaren, 2006) implements a computational model of casuistic reasoning, by which a decision is made via comparing the given problem to paradigmatic, real, or hypothetical cases.

Constraint-based methods such as the Ethical Governor model (Arkin, 2009) are proposed promote moral and ethical behavior of autonomous systems by using explicit rules that avoid harmful consequences. Recently, a similar strategy is promoted in (Greene et al., 2016) to manage collective decision-making by filtering behavior per explicitly defined constraints. To address decision-making under uncertainty, Dennis et al. (2013) provide a strategy for generating plausible strategies and preferences in unforeseen situations.

In the absence of pre-defined constraints and preferences, the application of machine learning facilitated discerning rules and principles that are often left implicit in discussions involving moral philosophy and psychology. Among the applications of machine learning to ethics are the MedEthEx (Anderson et al., 2006) and the GenEth (Anderson and Anderson, 2014) systems, both of which use Inductive Logic to discern rules that resolve ethical dilemmas that emerge due to conflicting obligations and duties. Consequentialist theories of ethics are also among the common models for which solutions have been applied in the form of agents that are driven by utility-based cognitive architectures. In consequentialist theories of ethics, actions are evaluated by their consequences; that is, the objective is often to optimize the well-being of the largest group of individuals and achieve the most desirable situation possible among many options (Gips, 1995).

2.2. Machine Ethics

In relation to using ethics for agents, Moor (2006) proposed four types agency: (1) ethical impact agent, (2) implicit ethical agent, (3) explicit ethical agent, and (4) fully ethical agent. Ethical impact agents are indirectly ethical agents, because they are not endowed with models of ethics. Rather, the presence of a computing technology indirectly brings an effect that facilitates the emergence of a situation that can be viewed as morally desirable outcome. On the other hand, if the ethical behavior is intentional, an agent can be either an implicitly or explicitly ethical agent. Implicitly ethical agents are entities, whose actions are constrained to avoid unethical outcomes. Constraints are defined to inhibit undesirable behavior or allow only legal actions permissible by the rules of engagement and general laws of the domain of interest (Arkin, 2009). However, explicitly ethical agents embody knowledge-based models of ethical decision-making that allow representing ethical categories and performing analysis in each situation to select actions recommended by the model. Such models are guided by theories of ethics, including consequentialist, deontological, and virtue-based theories. Therefore, whereas explicit ethical agents are knowledge-based entities, implicit ethical agents can be considered as rule-based entities. Building on knowledge-based ethical agents, fully ethical agents bring a level of expertise, which enables agents to justify their moral reasoning and learn from experience (e.g., failures) to improve their own models of ethics.

3. QUANTUM INFERENCE MODELS

Both classic and quantum probability theories are concerned with the problem of assigning probabilities to events such as the decisions in an ethical dilemma. Quantum approach to reasoning (Trueblood and Busemeyer, 2012) and decision-making is predicated on three major observations: (1) human judgments are constructed from both the current context and the question; hence, the state of a cognitive system is undetermined and the resolution is facilitated by measurements/questions; (2) the ambiguity with respect to multiple beliefs is akin to a thought wave over multiple cognitive states that are superimposed until a final decision is made; (3) the changes in a context driven by a judgement affect later judgements due to incompatibility between events disturbing each other to generate uncertainty.

The quantum inference model has been used to account for order effects in various application domains, including medical diagnostic tasks and jury decision-

making problems. In explaining the order effects, the notion of incompatibility is used to model different aspects of a problem along with their interactions. Specifically, the concept of incompatibility recognizes that the set of features used to evaluate one aspect may not be shared by the set used to model another so that no common set of features can be used to jointly evaluate information from different aspects. This is particularly the case when knowledge about all combinations of information from different sets is not known by the decision-maker. Hence, different perspectives need to be adopted to evaluate different pieces of information.

3.1. Events

Events in quantum theory are modeled as subspaces, which are characterized in terms of basis vectors. Each basis vector corresponds to an elementary outcome. Figure 1 illustrates a hypothetical vector space with three events *X*, *Y*, and *Z*. The basis vectors are orthogonal to each other, indicating that the events that they represent are mutually exclusive.

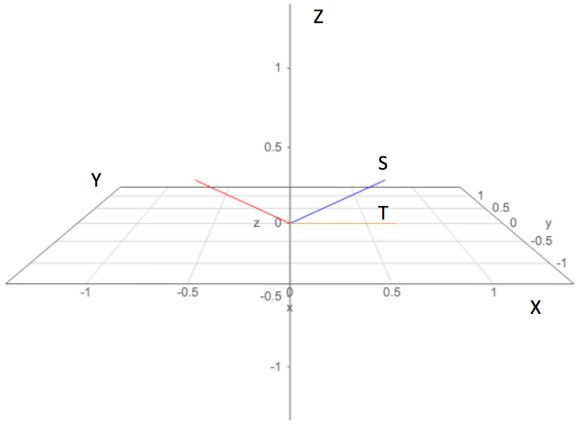

Figure 1: Vector Space

Because events are defined as subspaces, the logic of subspaces are used to combine events so that conjunction of events are defined in terms of the intersection of subspaces. That is, the span of the rays formed by $|X\rangle$ and $|Y\rangle$

3.2. System State

In quantum theory, the system state is defined by a unit-length vector in the *N*-dimensional vector space. The vector, symbolized as $|S\rangle$, is used to map events into probabilities. The state in Figure 1 refers to the line segment labeled as *S*, which is located at 0.696 units along the *x* axis, 0.696 units along the *y* axis, and 0.175 units up the *z* axis. Updating the state based on observations involves projection of the state vector onto the subspace of the observed event.

More specifically, the state is projected onto the subspace of the event, and projected vector is divided by the length of the projection to make sure that the revised state has length one. Also, given the original state vector, *S*, the probability of an event associated with a basis vector is equal to the squared length of the projection. The state $|S\rangle$, shown in Figure 1 is defined with respect to basis vectors $|X\rangle$, $|Y\rangle$, $|Z\rangle$ as follows:

$$|S\rangle = (-06963).|X\rangle + (0.6963).|Y\rangle + (0.1741).|Z\rangle$$

The coordinates of the state vector are defined in terms of the amplitudes of individual basis vectors:

$$|S\rangle \rightarrow \begin{bmatrix} -0.693 \\ 0.693 \\ 0.1741 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \alpha$$

The symbol α denotes the 3×1 (single column) matrix, which is comprised of coordinates of the abstract state vector $|S\rangle$ when defined in terms of the $|X\rangle$, $|Y\rangle$, $|Z\rangle$ basis. These coordinates can also be derived from the canonical coordinate system.

$$|X\rangle \rightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, |Y\rangle \rightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, |Z\rangle \rightarrow \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

The inner product between $|X\rangle$ and $|S\rangle$ computes the transition amplitude from the state $|S\rangle$ to state $|X\rangle$, and it is denoted by the matrix formula.

$$\langle X|S\rangle = [1 \quad 0 \quad 0]. \begin{bmatrix} -0.693 \\ 0.693 \\ 0.1741 \end{bmatrix}$$

$$= 1.(-0693) + 0.(0.693) + 0.(0.1741) = -0.693$$

Next, we use the concepts defined above to model the scenario presented in section 3 to illustrate the process of quantum inference model of causal reasoning to support ethical decision.

4. MULTI-PERSPECTIVE ETHICS CONFLICTS

Ethical dilemmas often involve consideration of decisions from the perspectives of multiple stakeholders. Next, we introduce a motivating scenario, which is adopted from Ethics Case Studies, published by APS Physics (2017). The scenario will be used to demonstrate the basic tenets of the proposed approach.

4.1. Motivating Scenario

Consider a graduate student who is in the final stages of completing her dissertation project and has applied for several tenure-track assistant professor positions. She is invited for interview at her undergraduate alma mater, which is a prestigious research university at which she has connections and would like to work. At the end of the seminar, during the question-and-answer session, the chair of the department asks for specific and detailed information about her research. However, her research group is currently preparing for a patent application that relates to the requested information. Moreover, her research group has agreed to hold information until the paper that is currently being prepared is submitted for publication. Also, her advisor is scheduled to give a presentation at a major international conference to release the details of the technique used in her research.

So, she kindly answers the department chair by indicating that she and her colleagues are in the process of developing a paper and submitting a patent application, and that she would be glad to share the information when an early print of the paper becomes available.

After the seminar, during the private interview session, the search committee presses harder to learn more about the details of her research, highlighting that the department is seeking team players and that a viable candidate for the position should share such information. In this scenario, there are multiple stakeholders: the graduate student applying for the position, her graduate advisor, the colleagues in her research group, the chair, the university where she is interviewing, her graduate university, and the research field. Each one of the stakeholders have different interests. The graduate student is conflicted for the decision made by her research group to delay the release of the information and to keep it confidential reduces her chances of getting hired by the department she is interviewing with. Her options include (1) giving the information and not telling her colleagues, (2) giving the information and telling her colleagues, (3) contacting her supervisor for permission, (4) convincing the chair out of his urgency to acquire the information.

From the perspective of the graduate applicant, demonstrating that she is a person of her word has the highest importance, but she is also conflicted and uncertain about which judgement is appropriate. The preferences of different stakeholders constitute distinct judgements, which may not be assessed simultaneously or jointly due to the incompatibilities among preferences. That is, for instance, a clear valuation over joint preferences of events from the perspective of her team and the department chair may not be readily available or even be meaningful. Therefore, before reaching a conclusion, she can view the problem from the perspectives of her research group, including her advisor, and the chair of the department to which she is applying. The interference and order effects should be taken into consideration in the reasoning process.

4.2. Quantum State Model

The motivating scenario involves a graduate student applying for a tenure-track position. She is expected to decide whether to give the requested information. There are three perspectives. The personal perspective of the graduate student, and the viewpoints of her research team and the department chair that will make the hiring decision. There are two judgments, giving the information or not, represented by events G versus $\overline{G}$ The evidence or observation that influences the judgement is whether it is permissible to release the information. The permissibility of releasing the information is represented by the event P, and its complement $\overline{P}$. In the context of classical probability or classical Bayesian model, the decision process can be modeled with at least 8 events, which consider combinations of giving the information, its complement, as well as the permissibility (and its complement) from the perspectives of the team and the department chair. Yet, these perspectives are incompatible, and the order in which each perspective is considered affects the final moral judgment.

To define a quantum inference model for this problem, we need to define a vector space, which is comprised of orthogonal and mutually exclusive events. One approach is to define eight basis vectors, one for each one of the combinations of three features: Giving the information (G), permissible from the perspective of the research team (P_t), and permissible from the perspective of the chair (P_c). Including the complements of these features, there are eight distinct joint events. Instead, the joint event space can be reduced to a four-dimensional space based on four combinations of two hypotheses (i.e., judgments of giving information or not) and two types of effects (permissible or not).

These four dimensions can be framed from three perspectives: applicant's perspective, team's perspective, and chair's perspective. The applicant's perspective represents the student's state of belief before any facts are considered from other perspectives. Considering her own perspective, the applicant generates the first judgment. If the research team's view is considered next, then the team's perspective denotes the state of belief when the applicant views the dilemma from the lens of the team, resulting in the second judgment. If the chair's arguments are considered next, then the department chair's perspective defines the applicant's state of belief when viewed from the chair's point of view, generating the third and final judgment.

To determine the probabilities associated with the judgments at each step of the process as different perspectives are considered, we need to clarify the relationships between the applicant basis, team basis, and chair basis. The same belief state $|S\rangle$ can be defined from the perspective of these three basis configurations as follows:

$$\begin{aligned} |S\rangle \quad &= a_{GP}.|A_{GP}\rangle + a_{G\overline{P}}.|A_{G\overline{P}}\rangle + a_{\overline{G}P}.|A_{\overline{G}P}\rangle \\ &\quad + a_{\overline{GP}}.|A_{\overline{GP}}\rangle \\ &= t_{GP}.|T_{GP}\rangle + t_{G\overline{P}}.|T_{G\overline{P}}\rangle + t_{\overline{G}P}.|T_{\overline{G}P}\rangle \\ &\quad + t_{\overline{GP}}.|T_{\overline{GP}}\rangle \\ &= c_{GP}.|T_{GP}\rangle + c_{G\overline{P}}.|T_{G\overline{P}}\rangle + c_{\overline{G}P}.|T_{\overline{G}P}\rangle \\ &\quad + c_{\overline{GP}}.|T_{\overline{GP}}\rangle \end{aligned}$$

For each basis, the index GP stands for the pattern "Giving the information and Permissible", $G\overline{P}$ for "Giving and not Permissible", $\overline{G}P$ for not "Giving and Permissible", and $\overline{GP}$ for not "Giving and not Permissible." Each basis can be represented by 3×1 coordinate matrix defining the amplitudes for each basis vector.

$$a = \begin{bmatrix} a_{GP} \\ a_{G\overline{P}} \\ a_{\overline{G}P} \\ a_{\overline{GP}} \end{bmatrix}, t = \begin{bmatrix} t_{GP} \\ t_{G\overline{P}} \\ t_{\overline{G}P} \\ t_{\overline{GP}} \end{bmatrix}, c = \begin{bmatrix} c_{GP} \\ c_{G\overline{P}} \\ c_{\overline{G}P} \\ c_{\overline{GP}} \end{bmatrix}$$

Transformation from one basis to another requires mapping the amplitude distribution of one perspective to the other so that events can be evaluated from a different viewpoint. Figure 2 illustrates a mapping from the perspective of the applicant to the perspective of the research team.

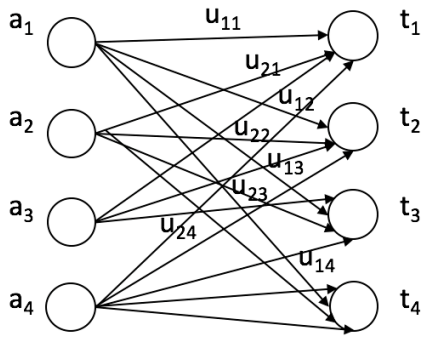

Figure 2: Transformation of the Amplitudes

The connection weights u_{ij} are used to map input amplitude distribution $\boldsymbol{a}$ across $a_1$to a_4 into the output distribution $\boldsymbol{t}$ across $t_1$to t_4. The mapping uses a linear combination strategy $t_j = \sum_i u_{ij}\, a_j$. This mapping can be characterized in terms of the following matrix operation:

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} = \begin{bmatrix} u_{11} & u_{21} & u_{31} & u_{41} \\ u_{12} & u_{22} & u_{32} & u_{42} \\ u_{13} & u_{23} & u_{33} & u_{43} \\ u_{14} & u_{24} & u_{34} & u_{44} \end{bmatrix} . \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

This can be abbreviated as

$$t = U.a$$

where U is a unitary matrix that preserves probability amplitudes in such a way that $t^T.t = 1$ just as $a^T.a = 1$. That is, sum of probabilities across events in each basis equal to 1. Therefore, a unitary matrix U maps one set of coordinates defined by the first basis to another set of coordinates defined by the second basis. Because we need the basis vectors to be of unit length, pairwise orthogonal, a unitary matrix need to preserve the lengths and inner products. Therefore, U needs to satisfy the property $U^T.U = I = U.U^T$. Any unitary matrix can be expressed in terms of matrix exponential of a Hamiltonian matrix, $U = e^{-iH}$, where $H^T = H$ is an$N \times N$ Hermitian matrix. The elements of an Hermitian matrix has elements $h_{jk} = h_{jk}^*$.

Next, we assume the presence of unitary transformation matrices among the perspectives. One unitary matrix U_{at} is needed to map the initial personal perspective of the applicant to the viewpoint of her research team. By using the unitary matrix, the coordinates of the belief state vector of the applicant is mapped on to the coordinates $t = U_{ta}.a$ of what the applicant thinks the team's perspective is. Similarly, the mapping from the applicant's perspective to the department chair's viewpoint requires the unitary matrix U_{ca}.

4.3. Applicant's Perspective

The initial belief state is defined in terms of the applicant's perspective and its orthogonal basis vectors. The initial state is akin to prior probability in a Bayesian network.

$$|S\rangle = a_{GP}.|A_{GP}\rangle + a_{G\overline{P}}.|A_{G\overline{P}}\rangle + a_{\overline{G}P}.|A_{\overline{G}P}\rangle + a_{\overline{GP}}.|A_{\overline{GP}}\rangle$$

The coordinates of the initial state are based on the probability amplitudes, which can be used to determine the probabilities of individual events. Probabilities of events are equal to square of the amplitudes defined in the coordinate matrix.

$$a = \begin{bmatrix} a_{GP} \\ a_{G\overline{P}} \\ a_{\overline{G}P} \\ a_{\overline{GP}} \end{bmatrix} = \begin{bmatrix} \sqrt{0.643} \\ \sqrt{0.643} \\ \sqrt{0.643} \\ \sqrt{0.643} \end{bmatrix}$$

Given this coordinate matrix of probability amplitudes for the applicant perspective, before the evaluation of other perspectives, the probability of giving the information can be derived by projecting the state vector onto the subspace spanned by the give-information basis vectors: $a_{GP}\,.\,|A_{GP}\rangle + a_{G\overline{P}}\,.|A_{G\overline{P}}\rangle$. Using the respective amplitudes, the probability of the decision (D) of giving the information equals $D(G) = |a_{GP}|^2 + |a_{G\overline{P}}|^2$.

4.4. Research Team's Perspective

To view the ethical dilemma from the perspective of the research team, the applicant needs to rotate her perspective and align it with the team's viewpoint. This can be done by either rotating the state vector or the basis. In the previous section, we defined the team's viewpoint from a distinct set of orthogonal basis vectors to interpret the applicant's initial state $|S\rangle$ from a different viewpoint.

$$|S\rangle = t_{GP}.|T_{GP}\rangle + t_{G\overline{P}}.|T_{G\overline{P}}\rangle + t_{\overline{G}P}.|T_{\overline{G}P}\rangle + t_{\overline{GP}}.|T_{\overline{GP}}\rangle$$

The amplitudes for the team's basis can be derived from the applicant's basis via an appropriate unitary matrix that meaningfully relates probability amplitudes across the source and target coordinate matrices.

$$t = \begin{bmatrix} t_{GP} \\ t_{G\overline{P}} \\ t_{\overline{G}P} \\ t_{\overline{GP}} \end{bmatrix} = U_{ta}.a$$

From the perspective of the team, the probabilities associated with each event are the squared magnitudes of the coordinate matrix. These probabilities are

$$\begin{bmatrix} |t_{GP}|^2 \\ |t_{G\overline{P}}|^2 \\ |t_{\overline{G}P}|^2 \\ |t_{\overline{GP}}|^2 \end{bmatrix}$$

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

Suppose that the team responds with "No Permission". The state vector $|S\rangle$ is then projected onto the subspace, which is spanned by the set of vectors $\{|T_{G\overline{P}}\rangle, |T_{\overline{G}\overline{P}}\rangle\}$. This is followed by normalization to produce the following unit-length state vector.

$$|S_{\overline{P}}\rangle = \frac{0.|T_{GP}\rangle + t_{G\overline{P}}.|T_{G\overline{P}}\rangle + 0.|T_{\overline{G}P}\rangle + t_{\overline{G}\overline{P}}.|T_{\overline{G}\overline{P}}\rangle}{\sqrt{|t_{G\overline{P}}|^2 + |t_{\overline{G}\overline{P}}|^2}}$$

The probability amplitudes of the projected and normalized vector, which is predicated on the research team's permission of the release of the information, is as follows:

$$t_{\overline{P}} = \frac{1}{\sqrt{|t_{G\overline{P}}|^2 + |t_{\overline{G}\overline{P}}|^2}} \begin{bmatrix} 0 \\ t_{G\overline{P}} \\ 0 \\ t_{\overline{G}\overline{P}} \end{bmatrix}$$

Following the selection of the "No permission" event from the perspective of the research team, the probability of releasing information, *D(G)*, is derived by projecting the state $|S_{\overline{P}}\rangle$ onto the subspace spanned by $\{|T_{GP}\rangle, |T_{G\overline{P}}\rangle\}$ and then squaring its length.

$$\begin{aligned} D(G|S_{\overline{P}}) &= || \; |T_{GP}\langle T_{GP}|S_{\overline{P}}\rangle\rangle + |T_{G\overline{P}}\langle T_{G\overline{P}}|S_{\overline{P}}\rangle\rangle \; || \\ &= || \frac{t_{G\overline{P}}}{\sqrt{|t_{GP}|^2 + |t_{G\overline{P}}|^2}} \; |T_{G\overline{P}}\rangle || \\ &= \frac{|t_{G\overline{P}}|^2}{|t_{GP}|^2 + |t_{G\overline{P}}|^2} \end{aligned}$$

4.5. Department Chair's Perspective

After evaluation of the research team's perspective, suppose the applicant considers the events of giving the information and permissibility of releasing the information, along with their complements, from the perspective of the department chair. The revised belief state, which was derived based on the evaluation of the team's perspective, is now viewed from the perspective of the department chair.

$$\begin{aligned} |S_{\overline{P}}\rangle = c_{GP}.|C_{GP}\rangle + c_{G\overline{P}}.|C_{G\overline{P}}\rangle + c_{\overline{G}P}.|C_{\overline{G}P}\rangle \\ + c_{\overline{G}\overline{P}}.|C_{\overline{G}\overline{P}}\rangle \end{aligned}$$

The belief state, as viewed from the chair's perspective, is characterized by the probability amplitudes, which can be defined by the following 4×1 matrix.

$$c_p = \begin{bmatrix} c_{GP} \\ c_{G\overline{P}} \\ c_{\overline{G}P} \\ c_{\overline{G}\overline{P}} \end{bmatrix} = U_{tc}.t_{\overline{P}}$$

The probability amplitudes are derived by using U_{ct}, which maps the calculated probability amplitudes of the team's perspective onto the department chair's perspective. The probabilities for each event are then defined as the squares magnitudes of the coordinates of c_p.

$$\begin{bmatrix} |c_{GP}|^2 \\ |c_{G\overline{P}}|^2 \\ |c_{\overline{G}P}|^2 \\ |c_{\overline{G}\overline{P}}|^2 \end{bmatrix}$$

The unitary matrix U_{tc} is expected to be designed to reflect the fact that the department chair is more likely to choose the "Permission" event than not. That is $c_{GP} + c_{\overline{G}P}$ should be significantly larger than $c_{G\overline{P}} + c_{\overline{G}\overline{P}}$. Following the research team's decision, if the applicant considers the department chair to choose to request the information, the belief state $|S_{\overline{P}}\rangle$ is projected onto the subspace spanned by $\{|C_{GP}\rangle, |C_{\overline{G}\overline{P}}\rangle\}$. The projection produces the following belief state

$$|S_{P\overline{P}}\rangle = \frac{c_{GP}.|C_{GP}\rangle + 0.|C_{G\overline{P}}\rangle + c_{\overline{G}P}.|C_{\overline{G}P}\rangle + 0.|C_{\overline{G}\overline{P}}\rangle}{\sqrt{|c_{GP}|^2 + |c_{\overline{G}P}|^2}}$$

The probability amplitudes with respect to the department chair's perspective as viewed by the applicant is predicated on department chair's preference (i.e., permission), which follows the research team's negative response. The amplitudes is represented by the following coordinate matrix.

$$c_{P\overline{P}} = \frac{1}{\sqrt{|c_{GP}|^2 + |c_{\overline{G}P}|^2}} \begin{bmatrix} c_{GP} \\ 0 \\ c_{\overline{G}P} \\ 0 \end{bmatrix}$$

Following the provision of the choices from the perspective of both the research team and the department chair, the judged probability of releasing the information is derived from the squared magnitude of the projection of the belief state onto the subspace spanned by $\{|C_{GP}\rangle, |C_{G\overline{P}}\rangle\}$. This probability equals

$$\begin{aligned} (G|T_{\overline{P}}, C) &= || \; |C_{GP}\langle C_{GP}|S_{P\overline{P}}\rangle\rangle + |C_{G\overline{P}}\langle C_{G\overline{P}}|S_{P\overline{P}}\rangle\rangle \; || \\ &= || \frac{c_{GP}}{\sqrt{|c_{GP}|^2 + |c_{G\overline{P}}|^2}} \; |C_{GP}\rangle || \\ &= \frac{|c_{GP}|^2}{|c_{GP}|^2 + |t_{G\overline{P}}|^2} \end{aligned}$$

Given the probabilities of distinct events derived from the resultant state vector, a final judgment is made in accordance with the probability amplitudes of the computed coordinate matrix.

5. EVALUATION AND DISCUSSION

The quantum probability model of the inference process shown in the previous section illustrates how an ethical dilemma can be viewed from the perspectives of different stakeholders. By evaluating different perspectives in the preference order imposed by the decision-maker alternative conclusions can be drawn. The value of the quantum approach and the use of the vector-based projections of the belief states are consistent with the

notion of ambiguity and uncertainty involved in ethical conflicts.

At any point in the process, the decision-maker does not have a definite answer; rather, by asking questions and making evaluations, the uncertainty is reduced or projected onto relevant dimensions. Also, the order in which the perspective considerations are made influence the outcome. This is akin to the notion of giving priorities to principles, consequences, or obligations, as well as the perspectives of the stakeholders in a context-sensitive manner. For instance, in one context, the applicant may choose to evaluate the situation from the perspective of the department chair, followed by an assessment from the point of view of the research team. Depending on the transition amplitudes or the rotation structure between the basis vectors of the subspaces, which represent distinct perspectives, a choice made in one perspective may either positively or negatively interfere with one or more events in the other perspective.

5.1. Learning Unitary Transformation Matrices

One of the critical challenges in devising a quantum approach to with such an interference mechanism is the need for defining unitary transition matrices between pairs of coordinate systems that represent different perspectives. Learning the connection weights between basis vectors using a gradient descent optimization approach is one plausible approach for deriving the relations between perspectives.

5.2. Hybrid Modeling of Quantum Cognition Models

The quantum inference model presented in this paper utilizes the Hilbert space to model belief state of an individual in terms of vector projections. The revision of belief states is driven by manipulation via unitary matrices of the amplitudes of the basis vectors in a coordinate system. The explicit representation of symbolic propositions and the revision of their degree of expectancy facilitate clear communication of the cognitive state of an agent. These symbolic propositions are not only succinct, but are also accessible to the decision-makers. However, the symbolic representation lacks a holistic view of the emergence of the probability amplitudes associated with the basis vectors. A sub-symbolic or implicit cognition layer can provide a unified mechanism to support the implementation of quantum cognition model.

The symbolic versus sub-symbolic representation dichotomy in cognitive systems is well-recognized. At the symbolic level, situations and objects are represented by atomic symbols that define labels and categories, which can be modified by syntactic manipulation. On the other hand, such categories and types emerge from a complex pattern of distributed and interactive sub-symbols that characterize the formation of the symbolic types and concepts. At the symbolic level, entities are atomic, discrete, and content-defined, whereas in the sub-symbolic level, there is a distributed and complex adaptive mechanism that characterize the emergence of categories and types at the symbolic level. The coupling of the explicit symbolic level and the implicit sub-symbolic level aims to mitigate the symbol grounding problem and variable-binding problem of the sub-symbolic phenomena.

Metaphorically, the implicit cognition layer is consistent with the view of making decision as the process of transforming a thought wave into a particle in a quantum model (Busemeyer and Bruza, 2014). Because even classical dynamical systems such as connectionist models could exhibit quantum-like properties (Blutner and Graben, 2014), the basic idea is to find a general and abstract formulation of quantum models that align with and validate against the structures of quantum probabilities and vectors in a Hilbert space. The wave propagation can be mimicked via coherence maximization in the constraint network in a way similar to simultaneous firing of dendrites in the connectionist model of the Pattern Recognition Theory of Mind (Kurzweil, 2012). The activations of nodes in the network is analogous to acceptability of the respective propositions.

This strategy is akin to viewing state as a unit-length vector in an N-dimensional vector space. Each node receives input from every other node that it relates to. The inputs can then be moderated by the weights of the link from which the input arrives. These weights can be construed as the elements of a unitary matrix that connect the probability amplitudes of propositions in one perspective to the probability amplitudes of the propositions in another perspective. The activation value of a unit is updated as a function of the weighted sum of the inputs it receives. The process continues until the activation values of all units settle. The activation values at the equilibrium state need to be projected and mapped onto probability amplitudes so that the probability of events sum up to 1.

In quantum cognition models, each coordinate basis is considered in sequence, resulting in order and interference effects that are consistent with observed human behavior. The provision of an implicit connectionist and sub-symbolic processing layer can help demonstrate a process by which beliefs as viewed from different perspectives can be adjusted in both directions simultaneously without following a strict sequence in evaluation. That is, evaluation from the view of multiple perspectives can interfere with each other. In the most general sense, the equilibrium can be construed as an attractor state in a complex adaptive system. The attractor state emerges at the end of a deliberation process by which we reflect on and revise our beliefs.

6. CONCLUSIONS

Ethical decision-making is emerging as a critical challenge as autonomous systems and software agents continue to immerse in our daily lives. Such systems face ethical dilemmas and conflicting situations when their course of actions may have both desirable and undesirable consequences when viewed from different perspectives. Moreover, the conflicts between obligations, duties, and consequences require viewing

situations from different perspectives to reach a decision that considers multiple points of view. The use of quantum probabilistic and causal inference models is promoted as a plausible strategy to frame the perspectives in terms of distinct coordinate systems, which are related to each other in terms of vector orientations. Furthermore, the capacity of quantum cognition to specify judgements in terms of indefinite states facilitates capturing psychological experience of conflict, ambiguity, confusion, and uncertainty. By using a simple and hypothetical example, we illustrated how quantum probabilistic reasoning can resolve conflict by reducing an indefinite state to a certainty and decision via projection over subspaces defined by multitude of basis vectors.

REFERENCES

Anderson, M. and Anderson, S.L., 2014. GenEth: A General Ethical Dilemma Analyzer. In *AAAI* (pp. 253-261).

Anderson M. and Anderson, S.L. eds., 2011. Machine ethics. Cambridge University Press.

Anderson M., Anderson S.L. and Armen C., 2006. An approach to computing ethics. *IEEE Intelligent Systems*, *21*(4), pp.56-63.

APS Physics, 2017. Ethics Case Studies. https://www.aps.org/programs/education/ethics/

Arkin R., 2009. Governing lethal behavior in autonomous robots. CRC Press.

Arkoudas S. B. and P. Bello, 2005. Toward ethical robots via mechanized deontic logic. In AAAI Fall Symposium on Machine Ethics.

Blutner R. and beim Graben, P., 2016. Quantum cognition and bounded rationality. Synthese, 193(10), pp.3239-3291.

Bruza P.D., Wang, Z. and Busemeyer, J.R., 2015. Quantum cognition: a new theoretical approach to psychology. Trends in cognitive sciences, 19(7), pp.383-393.

Busemeyer J.R. and Bruza P.D., 2012. Quantum models of cognition and decision. Cambridge University Press.

Gert B., 1998. Morality: Its nature and justification. Oxford University Press on Demand.

Greene, J., Rossi, F., Tasioulas, J., Venable, K.B. and Williams, B., 2016, February. Embedding ethical principles in collective decision support systems. In AAAI (pp. 4147-4151).

Dennis L., Fisher, M., Slavkovik, M., and Webster, M., 2013. Ethical choice in unforeseen circumstances. In Conference Towards Autonomous Robotic Systems (pp. 433-445). Springer Berlin Heidelberg.

Etzioni A. and Etzioni, O., 2016. AI assisted ethics. Ethics and Information Technology, 18(2), pp.149-156.

Gips J., 1995. Towards the ethical robot. Android epistemology, pages 243–252, 1995.

Herman M., 2014. Moral heuristics and biases. Journal of Cognition and Neuro-ethics, 73(1), pp.127-142.

Kurland N.B., 1995. Ethical intentions and the theories of reasoned action and planned Behavior1. Journal of applied social psychology, 25(4), pp. 297-313.

Kurzweil R., 2012. How to create a mind: The secret of human thought revealed. Penguin.

McLaren B.M., 2006. Computational models of ethical reasoning: Challenges, initial steps, and future directions. IEEE intelligent systems, *21*(4), pp.29-37.

Moor J.H., 2006. The nature, importance, and difficulty of machine ethics. *IEEE intelligent systems*, *21*(4), pp.18-21.

Saptawijaya A. and Pereira L.M., 2012, March. Moral reasoning under uncertainty. In International Conference on Logic for Programming Artificial Intelligence and Reasoning (pp. 212-227). Springer Berlin Heidelberg.

Trueblood, J.S. and Busemeyer, J.R., 2012. A quantum probability model of causal reasoning.

Wallach W. and Allen C., 2009. Moral machines: Teaching robots right from wrong. Oxford University Press.

Yilmaz L., Franco-Watkins, A. and Kroecker, T.S., 2016. Coherence-driven reflective equilibrium model of ethical decision-making. In Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2016 IEEE International Multi-Disciplinary Conference on (pp. 42-48). IEEE.

Yilmaz, L., Franco-Watkins, A. and Kroecker, T.S., 2017. Computational models of ethical decision-making: A coherence-driven reflective Equilibrium Model. Cognitive Systems Research. http://doi.org/10.1016/j.cogsys.2017.02.005

AUTHORS BIOGRAPHY

Levent Yilmaz is Professor of Computer Science and Software Engineering at Auburn University with a courtesy appointment in Industrial and Systems Engineering. He holds M.S. and Ph.D. degrees in Computer Science from Virginia Tech. His research interests are in agent-directed simulation, cognitive autonomous systems, and model-driven science and engineering for complex adaptive systems. He is the former Editor-in-Chief of Simulation: Transactions of the Society for Modeling and Simulation International and the founding organizer and general chair of the Agent--Directed Simulation Conference series. He published over 200 contributions in journals, book chapters, and conference proceedings and edited or co-authored several books on Modeling and Simulation. He received the Distinguished Service and Distinguished Professional Achievement Awards from the Society for Modeling and Simulation International, as well as the inaugural Auburn University External Consulting Award. He also served as the member of the Board of Directors and Vice President for Publications of the SCS. His email address is yilmaz@auburn.edu.