From Human to Machine Networks: A Framework for Integrating Consensus Approaches

Yixuan Fan, Huanyu Wu, Zhongxu Dong, Zongyao Li and Lei Zhang, *Senior Member, IEEE*

***Abstract*—Consensus commonly refers to a group reaching an agreement. The need for consensus spans from human groups to computer networks. However, the study of consensus typically focuses on distinct contexts, such as blockchain consensus, multi-agent systems (MAS) consensus, and human consensus. In this paper, we first develop a unified consensus framework applicable across specialized domains, from human to machine consensus. This framework includes three indispensable and fundamental components: participants as carriers, communication as the bridge, and state features describing the consensus transition. These core components enable us to assess consensus networks from essential needs and develop or adapt appropriate consensus mechanisms for diverse scenarios. To explore the essence of the consensus, we also define the consensus process as eliminating or reducing cognitive differences among participants. We believe the paper can inspire consensus research across various fields. Example strategies in consensus mechanism design to overcome consensus obstacles in the proposed consensus framework are provided in the end.**

***Index Terms*—Consensus, Distributed fault-tolerance consensus, Blockchain, Human consensus, Human-machine collaboration**

I. Introduction

The word “consensus” is explained as “Agreement in opinion, feeling, or purpose among a group of people, especially.” in the Oxford English dictionary [1]. The broad discussion on consensus in human society is an ancient topic. Initially, from the early functionalist perspective that emerged in the late 19th and early 20th centuries, consensus was viewed as a manifestation of social cohesion and collective consciousness [2]. By the mid-20th century, the focus shifted to social constructivism, which viewed consensus as constantly constructed and reconstructed through social interaction and communication [3]. Further developments in the mid and late 20th centuries led to the discussions on the methodologies of forming consensus [4]. The exploration of consensus from a social sciences perspective has always been developing and changing, continually influenced by the evolving structures of societal networks.

With the rise of computer science in the 1940s, the research of consensus has spanned widely from social science to technology. The concept of consensus soon began to be explored within distributed networks due to the need for distributed computing nodes to reach a unified understanding of the state, sequence, or outcome of certain operations [5]. Since the 1960s, aerospace control systems have used replicated processors for error detection, posing a challenge in achieving consistent decisions across processes and initiating the research of Distributed Fault-Tolerance (DFT) consensus [6]. This was significantly advanced by Leslie Lamport’s 1982 “Byzantine Generals Problem,” which established a framework for achieving consensus in systems despite the presence of malicious nodes [7]. By 2000, Internet companies began to use distributed servers, which required advanced consensus algorithms to synchronize data across databases, greatly advancing consensus technology. The emergence of Bitcoin in 2008 triggered a decade of intense blockchain development. As its core technology, the consensus algorithm has made great progress and broadened the research scope in the field of consensus technology. Additionally, multi-agent systems (MAS) represents another well-known form of consensus, focusing on coordinated behavior and decision-making among autonomous agents. Over time, the integration of artificial intelligence (AI) has largely driven the development of MAS, enabling these systems to handle more complex tasks and optimize performance [8].

A. Related Work

The forms of consensus vary widely, depending on the participants, methods, objects, etc. For a long time, experts in various fields have conducted consensus research within their professional domains. Studies on the fundamental characteristics of consensus, such as [9] and [10], primarily focus on human consensus from a social science perspective, without considering machine consensus in computer science and engineering. [9] investigates consensus emergence in biological and social networks, analyzing how individuals reach collective agreement with only local information and limited communication, highlighting key factors such as group size, preference distribution, and communication strategies. Similarly, [10] examines consensus in cultural and knowledge-sharing contexts, particularly in human society. On the other hand, consensus research in engineering and computer science predominantly focuses on specific consensus mechanisms, leading to the development of classic algorithms such as Paxos and Practical Byzantine Fault Tolerance (PBFT) in the domain of distributed fault-tolerant (DFT) consensus, as well as Proof of Work (PoW) and Proof of Stake (PoS) in blockchain technology [5]. Additionally, comprehensive surveys on machine consensus, including consensus in MAS and distributed systems, can be found in works such as [8] and [5].

B. Challenges and Motivations

While these mechanisms and field-specific discussions effectively address the needs of their respective areas, the lack

Yixuan Fan, Huanyu Wu, Zhongxu Dong, Zongyao Li and Lei Zhang (the corresponding author) are with the James Watt School of Engineering, University of Glasgow, Glasgow, G12 8QQ, UK; E-mail: {y.fan.3; h.wu.3; z.dong.2; l.zongyao.1}@research.gla.ac.uk; Lei.Zhang@glasgow.ac.uk.

of an overall understanding of consensus may fail to meet the continuously emerging new consensus requirements driven by societal and technological progress. For example, new challenges such as advanced nodes and dynamic consensus networks, have been introduced to distributed autonomous systems when vehicle-to-vehicle networks are implemented, complicating the consensus process [11]. Traditional paradigms like DFT consensus serves as a fundamental framework in various advanced-node consensus networks [11], but they originally developed for ordinary nodes performing basic operations such as reading, writing, and executing, may struggle to meet the complex and adaptive demands of intelligent systems. Instead, adopting methods inspired by the human consensus paradigm to explore the consensus of highly intelligent advanced nodes may be a promising direction.

A typical example of a cross-field consensus need can be seen in the recent guidance provided by Industry 5.0 and Society 5.0, which emphasize enhanced human-machine collaboration rather than pursuing solely machine-based systems with advanced features [12]. A paradigm of hybrid consensus that integrates human and machine consensus is a promising trend in manufacturing and also in everyday societal interactions [12] [13]. This complex interaction requires not only the algorithmic design of machine operation in consensus but also a grasp of the psychological and sociological factors that influence human behavior in consensus [13].

As a broad concept, consensus holds the same core characteristic, i.e., a network of participants developing an apparent consistency or agreement. Commonalities and key elements are evident, even though consensus research spans different disciplines. Integrating discussions from various consensus disciplines can lead to several significant benefits. First, different consensus research can learn from each other. For example, some typical MAS consensus algorithms draw inspiration from consensus mechanisms observed in animal behavior [8]. Second, by clarifying and understanding the features of the core elements of consensus, researchers can better estimate the limits and applicability of various consensus strategies. For instance, the fault-tolerance threshold, which is well-discussed in DFT consensus, may also be applicable to general consensus research by following the quorum intersection rules. Third, adopting a comprehensive perspective on consensus helps develop consensus strategies that are flexible and not confined to specific disciplines, enabling them to address new consensus challenges across various fields. However, such integrated discussions are currently lacking.

C. Research Questions and Contributions

Recognizing the need for consensus discussions that cross disciplinary boundaries, we propose the first discussion point in this paper: **Can we identify the indispensable fundamental elements that primarily affect the consensus process across both human and machine systems in a unified way?** To explore this question, we propose a broadly adaptable consensus framework by identifying three key elements: participants, state, and communication. Participants are characterized by cognition and honesty. The transition from the initial state to the consensus state describes the consistency objective of the consensus. Communication methods both enable and limit the realization of the consensus process.

To truly grasp the depth of the consensus process, we raise a second question: **What are the primary objectives of a consensus process, and what are the primary factors affecting the achievement of consensus?** Based on the proposed consensus framework, we suggest that the essence of the consensus process is to overcome cognitive differences and dishonesty among participants. In distributed computing networks, this corresponds to data discrepancies and malicious behavior among nodes. Established methods in existing consensus systems, such as digital signatures [5], Byzantine fault tolerance protocols [7], and iterative information exchange [8], are strategies designed to address these two primary obstacles. The terms "algorithm", "protocol" and "mechanism" are used interchangeably throughout this paper to describe the underlying processes of achieving consensus. The main contributions can be summarized as follows.

- We extract the characteristics of human consensus (HC) and machine consensus (MC) and introduce a new form of consensus emerging from human-machine collaboration, called hybrid consensus (HBC). This interdisciplinary comparison transcends the boundaries of different areas of consensus and provides a comprehensive perspective on consensus.
- We propose the first unified consensus framework across consensus from human to machine with three common indispensable fundamental components that fit consensus analysis in various fields: participants as the carriers of consensus, communication as the bridge, and state descriptions marking the transformation from chaos to consistent cognition. These three fundamental elements are prerequisites for studying any type of consensus.
- Based on this consensus framework, we propose a definition of the consensus process. Fundamentally, the consensus process aims to eliminate or reduce cognitive differences among members regarding the consensus object. Cryptography techniques, different broadcast primitives, and various consensus protocols are typical strategies to overcome these primary obstacles. This understanding of consensus as an entropy-reduction process can inspire the design of more effective consensus mechanisms.
- Finally, examples of consensus mechanism design strategies are demonstrated based on the proposed consensus framework and the primary goals of the consensus process.

II. Consensus from Society to Technology

Before building a unified consensus framework, we analyze various consensus fields to compare well-known consensus. The classification of consensus types in this paper depends on the participants involved. We categorize consensus into Human Consensus (HC), characterized by human participants; Machine Consensus (MC), marked by computational entities such as computer nodes, agents, or intelligent machines; and also the promising new emerging consensus form, Hybrid

TABLE I: Comparison of consensus from technology to society

Consensus Types	Examples of Consensus Forms	Main Participants	Consensus Object	Mechanism Examples
Human Consensus	Voting, Discussion	Humans	Defined/Implicit Proposals	Majority Vote, Delphi Method
Machine Consensus	Distributed fault-tolerance consensus Blockchain consensus Multi-agent system consensus	Nodes (Processes) Nodes (Miners) Autonomous agents	Client requests Transactions Actions, Decisions	PBFT, Raft PoW, PoS Average consensus, Voting
Hybrid Consensus	Healthcare decision support system, Autonomous driving network	Intelligent agents and Humans	Cooperation tasks, Shared data	Not clear

Consensus (HBC), which involves interactions between humans and machines.

1) Human Consensus (HC): HC is typically studied as a process that leads a group of people within a social network to reach an agreement on various issues, ranging from social norms to political affairs. While early discussions of consensus in social sciences focused on theoretical understandings of group dynamics and social cohesion, it was not until the mid and late-20th centuries that specific methodologies for forming consensus among people were proposed. Notable among these mechanisms are the Delphi Method, Nominal Group Technique (NGT), and Round-Robin Discussion [4]. These methods are commonly used in fields such as public policy development, organizational management, and social research, where gathering diverse opinions and reaching a collective agreement are essential. While those mechanisms are primarily designed for structured environments, the HC also manifests in various voting systems and in the consensus state of culture, social norms, and ethical standards within social networks. TABLE I outlines the main characteristics of HC. Moreover, consensus is not exclusive to humans in the biological realm. It is also prevalent in animal behaviors such as migration, foraging choices, and the selection of group territories and defense strategies among birds, insects, fish, and mammals. In this paper, we focus on humans as consensus examples of living organisms.

2) Machine Consensus (MC): MC refers to all forms of consensus exclusively involving machines, such as computer nodes, agents, robots, and other automated participants in a system that relies on a communication network to achieve consensus. Generally, there are three well-known MC domains: DFT consensus, Blockchain consensus, and MAS Consensus. DFT consensus plays a crucial role in distributed computing systems. Typically, Byzantine Fault Tolerance (BFT) algorithms are designed to achieve consensus on a valid value within a system that tolerates a limited number of malicious nodes. Crash Fault Tolerance (CFT) algorithms, on the other hand, only assume a limited number of nodes that may crash but no malicious behaviour. Blockchain technology, which emerged later, adopts similar principles of consistency from DFT but introduces innovations in achieving consensus. This is particularly evident through a consistency log maintained as a “chain” in a widely accessible network. The third type, MAS consensus, adopts a distinct approach to consensus compared to the other two. Operating under the assumption of autonomous agents, its key algorithms, such as average consensus and interactive consistency, draw parallels to consensus behaviors observed in biological populations [8], which enable MAS systems to reach reliable consensus outcomes through cooperative decision-making among agents. TABLE I outlines the main characteristics and comparisons of MC.

3) Hybrid Consensus (HBC): While DFT and blockchain consensus mechanisms prioritize consistency and fault tolerance, MAS consensus focuses on adaptive decision-making and collaborative problem-solving, reflecting the dynamic interaction and collective intelligence of autonomous agents. This makes MAS consensus research extend to higher intelligent agent consensus with the recent rise of AI technology. The autonomy and intelligence level of machines holds immense potential to operate in an unframed and versatile goal mode, similar to human behavior. The intelligent upgrade of nodes and the increasing demand in the environment have revealed the fusion goal of integrating fault tolerance and preferences, which is commonly recognized in HC. On the other hand, as highly intelligent machines become integrated into human life and take on more flexible work, the interaction and cooperation between high-intelligence machines and humans are inevitable [13]. Humans and machines generally have relatively obvious independence in HBC scenarios. The logic of consensus can be specified in advance by algorithms or autonomously negotiated in a relatively open environment. The following is an example of an existing HBC. In an L3 autonomous driving system, when an unexpected obstacle is encountered during driving, humans and autonomous driving car agents negotiate decisions. The vehicle system and the human driver each retain independent decision-making perspectives. The vehicle agent AI will provide optimization solutions based on the perception data, and humans will select or fine-tune these solutions based on situational awareness and ethical considerations that go beyond pure data analysis. Ultimately, the consensus is not simply reached by letting people or machines take over completely but by cooperating within a safe and controllable range. Cases like this will blur the boundary between consensus paradigms of machines and humans. Considering both technological and societal dimensions in consensus is very necessary. To build a dialogue of consensus from society to technology, we have established a unified consensus system in the following sections to condense the common important elements and characteristics of different forms of consensus. Examples and their features of HBC are also listed in TABLE I.

III. A Generalized Consensus Framework

Despite the diverse forms of consensus spanning from society to technology, we observe commonalities in the consensus process and conclude that the essential constraints can be concluded into three key elements: participants, state, and communication, which are the primary factors affecting the achievement of consensus. Based on these elements, we propose a generalized consensus framework illustrated in Figure 1, laying the groundwork for a generalized consensus process. This framework explains that consensus is a transformation process from a chaotic initial state to a consensus state of significant agreement among participants regarding a consensus object, facilitated by an implicit or explicit consensus mechanism that depends on information exchange through communication networks. Essentially, the consensus process relies on participants utilizing the communication network to interact effectively and exchange information. The participants' characteristics (cognition and integrity), the communication network conditions, and the consensus state requirements largely determine the feasibility and method of achieving consensus. Therefore, clarifying these three elements is essential before establishing or selecting any specific consensus mechanism. Here, we explain these three elements and their impact on the consensus system in detail.

A. Participants

Participants are the entities involved in the consensus process, acting as carriers of consensus. They can be human individuals, computational nodes, autonomous agents, or hybrid entities. The cognitive features of participants, along with their honesty, significantly influence the intricacy of the consensus process. The following subsections will explore them in detail and compare them across different types of consensus, as summarized in TABLE II.

1) Cognitive Ability: Cognitive ability is an anthropomorphic expression of the consensus participants' ability to receive, understand, evaluate and transmit information and to make decisions based on received information in the consensus process. It is also strongly influenced by participants' openness to external assistance and sensitivity to environmental impacts. For a machine system, cognitive ability can refer to the capability to receive and process information, such as signal acquisition (e.g., sensor inputs) and data evaluation (computational analysis). For instance, each node aggregates and evaluates data from multiple sources, determines potential consensus targets, and adjusts its decision logic based on detection and inference processes. The cognitive abilities of participants significantly influence the complexity of the consensus process by introducing more intricate decision-making factors. For example, in human voting-based consensus, participants can employ tactical voting to conceal their true preferences in order to achieve a collective voting outcome that is closer to their own inclinations.

Here, we further categorize MC based on the cognitive abilities of the participants into two types: Ordinary Machine Consensus (OMC) and Advanced Machine Consensus (AMC). OMC includes consensus systems with low cognitive abilities, such as DFT and blockchain consensus algorithms, which typically offer simpler functionality focused on read-write consistency. AMC includes consensus systems with relatively high cognitive abilities, such as MAS consensus systems, particularly those involving intelligent agents. The cognitive abilities of agents in AMC rely on their level of intelligence but are generally higher than OMC due to their advanced functions, such as AI-enhanced sensing and complex decision-making. Generally, humans, as participants, show high cognitive abilities. The cognitive abilities of different consensus types are listed in TABLE II.

2) Cognitive Differences: Cognitive abilities are relative to individual participants' measurement standards. Since consensus must be processed within clusters, the diversity in how participants process, respond to, and interpret information serve as a critical collective metric because eliminating cognitive disparities towards consensus objects within these clusters is the primary goal of consensus. We name this metric cognitive differences. Cognitive differences are influenced not only by the levels of participants' pre-existing knowledge but also by their distinct data-processing methods, characteristics, and operational modes for handling incoming information. Low cognitive differences are exemplified by DFT consensus systems, where the majority of nodes are non-faulty and exhibit highly consistent, programmed reactions to the consensus process. Conversely, nodes with different behaviors, such as crash nodes and Byzantine nodes, typically represent a minority and show cognitive inconsistencies. In cases of low cognitive differences, the priority of the consensus process is to reach an agreement by minimizing differences among participants regarding the consensus object and tolerating the minority of inconsistent participants. An illustrative example of high cognitive differences is a project meeting where participants have different priorities, with some focusing on profit and others on environmental considerations. These significant cognitive differences can pose substantial challenges to reaching a consensus. In such cases, consensus groups usually require more centralized decision-making methods, such as majority voting that relies on a central platform, to achieve consensus.

TABLE II presents the cognitive consistency comparisons of MC, HC and HBC. In HC, humans often exhibit high cognitive differences due to varied information processing and reasoning. HBC similarly tends toward high differences because humans and machines may prioritize decisions differently. Note that in scenarios involving clearly defined consensus choices, specific instances of HC may sometimes demonstrate moderate cognitive differences. Cognitive similarities in consensus can stem from shared consensus backgrounds within the group, such as culture, shared life experiences, etc. As for OMC, the cognitive differences of DFT consensus have been discussed. In most PoX-type blockchain algorithms, the standards for verifying a transaction between nodes are highly consistent due to the verification methods being standardized, showing low cognitive differences. AMC featured participants with higher levels of intelligence (such as intelligent agents) than those in MC but generally lower than humans in HC. In such systems, cognitive differences depend on the uniformity of participants' design features. For instance, if the same or similar algorithms are applied to identical agents, then their cognitive differences are low. However, due to the potential difference in agent systems and the varying designs of their operational mechanisms, diversities in cognition about the

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/MNET.2025.3559378

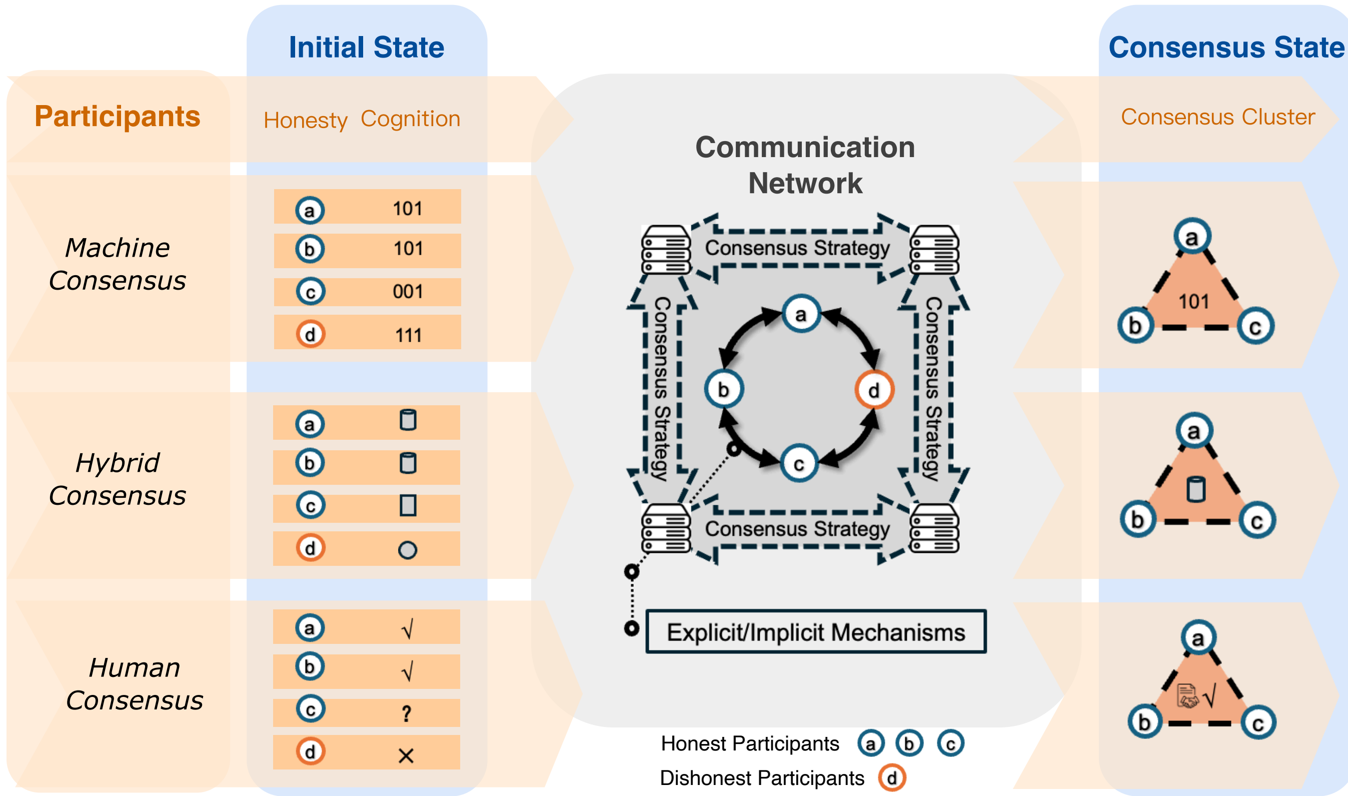

Fig. 1: Consensus framework

TABLE II: Features of Participants

Consensus Type	Honesty Assumption	Cognitive Ability	Cognitive Differences
Human Consensus	Unpredictable	Very high	Moderate to very high
Ordinary Machine Consensus	Byzantine nodes: low Crashed nodes: Absolute honest Good nodes: Absolute honest PoX: Conditional (constrained by incentive mechanism)	Low	BFT: low CFT: low (Only crashed/Byzantine nodes have different cognition) PoX: generally low (the standard to verify a valid transaction among different nodes is highly consistent between nodes)
Advanced Machine Consensus	Undetermined	Moderate to high	Moderate to high
Hybrid Consensus	Undetermined	Moderate to high	Moderate to high

consensus object can arise. Therefore, overall, we consider the cognitive differences in AMC to be moderate to high. An analysis of the design of mechanisms to overcome different cognitive differences assumptions in several consensus cases can be found in Sec. IV-1.

3) Honesty: The consensus process involves information exchange among participants, where not only cognitive differences but also the integrity of participants significantly impact the outcome. Given that honesty assumptions vary widely across different scenarios, we specifically address honesty within the context of explicit mechanism examples. Conclusive comparisons are listed in TABLE II, while detail analysis are as follows. The BFT algorithm tolerates a limited number of dishonest nodes by allowing some malicious behavior. In contrast, the CFT algorithm assumes nodes may fail (crash) but do not act maliciously, implying overall honesty, even among failed nodes. In PoX blockchain consensus mechanisms, nodes are generally considered unreliable. However, these algorithms use incentives to encourage honesty and limit dishonest behavior, making their honesty conditional. A discussion on how different honesty assumptions in BFT, CFT, and blockchain consensus are addressed in Sec. IV-2. In HC and HBC systems involving humans, honesty is unpredictable due to strategic and complex thinking, such as tactical voting. In AMC and HBC, the assumption of intelligent agents may be defined case by case and waiting for further research. Therefore, we do not assume their honesty in TABLE II.

B. State

States represent the level of agreement in the consensus process, marking the transition from an initial state of disparity to a consensus state of agreement. These states are fundamental and universally applicable across all forms of consensus. The initial state can range from a set of clear preferences (e.g., agree/disagree, choosing among multiple options) to being entirely unclear and lacking a definite direction (e.g., in iterative or convergent consensus). For consensus state, different consensus systems have varying criteria around the consistency requirement. Common consensus states may include agreement on a single value, multiple valid values, or even a range of values. In addition to the state of the consensus object, the degree of consistency can also vary. The agreement could involve all participants, only non-faulty participants, a majority of participants, or even just key participants. Time requirements are also considered as a dimension of the consensus state, including whether the consensus process is a one-time event or continuous and whether there are specific time series requirements. Additionally, some consensus algorithms define strong and weak consistency. Strong consistency requires all participants to have the same view simultaneously, while weak consistency allows temporary discrepancies but ensures eventual convergence [5]. Moreover, in certain cases, consensus is designed to be achieved with a very high probability [11]. In TABLE III, we illustrate common initial state and consensus state of the process.

The consensus process, which converges from an initial state of disorder to a state of unanimity, can be viewed as a process of entropy reduction. The message transmission and decision-making in the consensus process all contribute to this entropy reduction. This state transition process can occur through explicit methods, such as thoughtful negotiations or structured decision-making procedures. Often, these explicit processes are accompanied by well-defined consensus mechanisms, such as voting, DFT algorithms, or blockchain consensus algorithms. Alternatively, the process can occur through implicit methods, which evolve gradually from a shared culture or a tacit understanding of group behavior without clear coordination.

TABLE III: Initial state and consensus state

Category	Details
Initial state	Binary choice / Multiple choice Open proposal
Consensus state	Strong consistency / Weak consistency Single value / Multivalue / Range All the good nodes / All the nodes reach consistency Consensus in majority / Critical members Single consensus / Continuous consensus Time series requirement Deterministic / Probabilistic consensus

C. Communication

According to the state transition idea of consensus, information exchange is essential for establishing agreement among participants. Only through communication can participants interact and eliminate cognitive differences. Communication here refers to the means and channels through which participants exchange information during the consensus process. The environment of the communication network is subject to objective scene restrictions and can also be crafted during the consensus process to create more favorable channels for consensus. The mode of communication can significantly impact the feasibility, speed, consistency, and even the outcome of consensus formation.

1) Communication Across Different Consensus Types: In the process of HC, communication often relies on language interaction or the use of informational platforms. Some of these communication methods are straightforward and transparent, such as real-time messaging through person-to-person oral communication or modern digital channels like texting or online chatting. However, communication can also occur through more complex and layered means, such as public media, culture expressions and arts. They convey messages for broader consensus in subtler and often more profound ways.

Furthermore, the integration of humans and intelligent machines in HBC introduces additional communication challenges. In addition to establishing a strong communication network infrastructure, this hybrid interaction also requires human-machine interaction technology to assist communication. For example, natural language processing enables the comprehension and generation of human language, facilitating seamless textual interactions. Speech recognition and synthesis technologies support verbal communication, allowing for natural and intuitive exchanges. Additionally, tactile feedback mechanisms provide a tangible dimension to interactions, enhancing the physical and sensory experience [12].

TABLE IV: Communication assumptions overview

Category	Assumptions
Timing	Synchronous/Asynchronous
Network Structure	Centralized/Distributed; Fully/Partially connected
Communication Mode	HC: Verbal/Non-verbal communication MC: Signal/Data transmission

2) Communication Assumptions: Here, we list some assumptions of communication in TABLE IV. Under the timing assumption, synchronous communication among machines bears resemblance to live workshop discussion, where all the participants knows when the discussion starts and ends by the announcement of the chair. This entails the transmission of messages within predefined temporal constraints universally acknowledged by all participants. Put differently, each participant has a synchronized global time, enabling discernment that if a message remains undelivered within a specified time frame, it has not been dispatched at all. Conversely, in asynchronous scenarios, communication parallels distant individuals employing foot messengers, where the transmission may be delayed. Consequently, recipients are unable to distinguish whether the delay is attributable to the messenger's departure or the absence of a messenger.

The network structure of communication can both impose limitations and aid in forming consensus. For example, a fully connected topology facilitates the transfer of messages among participants, while a poorly connected structure may cause message blockages or singular channels for some participants, reducing the reachability of consensus messages. On the other hand, if the communication network architecture is highly centralized, it results in consensus messages being predominantly controlled by a central entity. For example, in wireless communication, a centralized approach entails nodes transmitting their information to a central control station. This central station is responsible for making final decisions and sending instructions back for execution. Examples can also be found in HC, such as traditional news dissemination, which relies on public media. Although the centralized method often has higher consensus efficiency, the overall consensus outcome more significantly relies on the central entity, and this influence can be either positive or negative.

While the communication network builds a backbone, the communication mode decides how communication happens. Examples are retransmission mode and non-retransmission mode according to the Quality of Service (QoS) based on the inherent communication framework. Another example is granted communication requires explicit permission for entities to interact within a system, enhancing security, while non-granted communication allows open interaction without prior authorization, prioritizing ease of access and convenience.

D. The Consensus Process

Based on the proposed consensus three key elements, we propose a perspective on the definition of the consensus process in Remark 1.

Remark 1. *The consensus process involves overcoming cognitive differences and dishonesty among participants, transitioning their cognition of the consensus object from chaos to significant agreement. Fundamentally, it aims to eliminate or reduce cognitive differences among members regarding the consensus object.*

Remark 1 is motivated by the fundamental need for consensus, which arises when participants have differing or unclear understandings of the consensus object. As discussed in Sec. III-A2, participants with significant cognitive differences face greater challenges in achieving consensus. Furthermore, participant dishonesty can negatively impact the dissemination of messages that facilitate consensus, thereby impeding the consensus process and introducing uncertainty. To answer the raised question in Sec I, we proposed the primary objective of a consensus process is to eliminate or reduce cognitive differences and dishonesty among participants, guiding their cognition of the consensus object from disorder to significant agreement. This definition is exploratory and may reveal its incompleteness in future research. Nevertheless, we hope that this viewpoint can inspire research on consensus across various domains.

IV. Consensus Strategies

Here, we show typical examples of strategies in established consensus systems to overcome these primary obstacles.

1) Strategy Examples of Overcoming Cognitive Differences: Identifying cognitive differences among participants in a consensus system can help in selecting more suitable consensus strategies. For example, in systems where participants share a clear, unified task orientation and operate within a limited set of actions, consensus can typically be achieved by synchronizing trusted information. For instance, in DFT and blockchain consensus mechanisms to ensure that a consensus value, based on single input consensus, is either accepted or rejected by a sufficient number of participants. For example, in CFT, a request is processed depending on whether it receives adequate support from nodes. In blockchain, the longest chain rule decides which chain is confirmed. However, in scenarios involving an open consensus object that participants might hold significantly different values, participants often need to change information iteratively to converge either to a single value or to values within an acceptable margin. In systems with greater cognitive difference, a more open and multidimensional decision space may require deliberate convergence mechanisms over multiple rounds. For instance, an consensus approach in [14] uses AI mediation to iteratively generate "group statements," helping participants converge on shared perspectives while incorporating both majority and minority views.

2) Strategy Examples of Overcoming Dishonesty: Honesty assessment of nodes in a consensus system is a crucial factor in selecting and establishing an appropriate consensus method. For example, in a DFT consensus system, if all nodes are assumed to be non-malicious (no dishonest nodes), a CFT algorithm can be used, which each quorum has $f+1$ ($f=\frac{n}{2}$) such that at least one node is correct and it is responsible for reducing the cognitive inconsistency. However, if the system is expected to include malicious (Byzantine) nodes, a BFT consensus is necessary, requiring $2f+1$ nodes ($f=\frac{n}{3}$) so that the number of honest nodes ($f+1$) is always at least one more than the potential Byzantine nodes (up to f). The quorum threshold design in protocols ensures that there are enough honest nodes to outweigh any dishonest ones (Byzantine nodes), allowing participants to identify and rely on information from honest nodes to resolve cognitive differences. This example illustrates how the honesty conditions of the system can increase the complexity of reaching a consensus, requiring different strategies.

We provide some other examples of blockchain consensus to display their strategies to overcome dishonesty. In terms of mitigating the dishonesty of participants, public key cryptography could be used to authenticate the identity of participants and constrain the ability to act as someone else. Reliable broadcast [15] can force a sender to tell the same story to everyone. Typically, blockchain consensus generally uses digital signature (public key cryptography) to guarantee that every participant cannot pretend to be another node.

In token-integrated blockchain, token-based consensus might be used, including PoS and PoW. In these consensus, token are used as the incentives for reaching consensus. In PoW, miners compete to solve a complex math problem for the reward token, the winner's block will be validated and accepted. In PoS, the more token a participant holds, the more likely it could be selected to validate a block. Because a stake holder tend to increase its money rather than losing value, including its stake and the price of its stake in the whole system, stake-holders are more likely to maintain the system honestly.

V. Discussion: How AI Will Affect Consensus

AI brings advanced intelligence to machine groups, introducing a new form of machine consensus among high-intelligence agents (AMC). While traditional machine consensus primarily focuses on information synchronization and verification, AMC significantly increases the demands of decision-making. Each intelligent agent may exhibit distinct learning strategies and risk preferences, amplifying differences within the machine group and complicating coordination. As a result, consensus mechanisms must extend beyond state consistency and data synchronization, potentially requiring complex game-theoretic approaches to balance diverse interests and allocate resources effectively.

Simultaneously, the advancement of AI technology has also intensified the need for human-machine consensus (HBC). As machines gain sufficient intelligence to collaborate with humans, additional considerations arise beyond fundamental technical aspects, including interaction mechanisms, decision-making priorities, and role allocation. Key challenges include designing efficient human-machine collaboration frameworks, ensuring a balanced distribution of authority in critical scenarios, and defining responsibilities and boundaries among multiple decision-making entities.

To facilitate cooperative consensus, the establishment of trust is crucial. A fundamental challenge is ensuring that AI's decisions are reliable, transparent, and comprehensible to all participants, as AI agents assume increasingly central roles in decision-making. Enhancing explainability in AI decision-making is therefore a critical research direction. Additionally, trust depends on the predictability and consistency of both human and AI participants in adhering to established agreements. Mechanisms such as incentive and punishment systems, along with reputation and credit scoring models based on game-theoretic principles, can help regulate participant behavior and mitigate risks associated with malicious or uncooperative actions. Finally, secure and verifiable technologies, such as blockchain, can enhance data integrity and transaction traceability, reducing concerns about data manipulation and information asymmetry, thereby strengthening mutual trust among multiple stakeholders.

VI. Conclusion

In summary, this paper compares and analyzes consensus mechanisms across technological and social science domains, establishing a unified consensus framework. This framework, derived from a broad perspective on consensus, incorporates three core elements: the participants as carriers, the state transitions in the consensus process, and communication serving as the connecting bridge. This consensus framework contributes to a more comprehensive understanding of consensus and proposes a systematic approach to evaluate and integrate consensus mechanisms in different contexts. Based on this framework, we conclude the nature of the consensus process as eliminating cognitive differences among participants. The consensus framework and definition proposed in this paper aim to promote consensus mechanisms to adapt to complex and dynamic real-world scenarios and transcend traditional limitations. This not only inspires further consensus research in various disciplines but also provides systematic analysis guidance for various traditional and new consensus.

References

[1] Oxford English Dictionary, "s.v. 'consensus (n.)'," https://doi.org/10.1093/OED/7567688553, Sep. 2023, accessed: April 30, 2024.

[2] J. H. Turner and A. Maryanski, Functionalism. Benjamin/Cummings Publishing Company Menlo Park, CA, 1979.

[3] C. Agius et al., "Social constructivism," Contemporary security studies, vol. 3, pp. 87–103, 2013.

[4] R. Balasubramanian and D. Agarwal, "Delphi technique–a review," International Journal of Public Health Dentistry, vol. 3, no. 2, pp. 16–26, 2012.

[5] G. Zhang, F. Pan, Y. Mao, S. Tijanic, M. Dang'ana, S. Motepalli, S. Zhang, and H.-A. Jacobsen, "Reaching consensus in the byzantine empire: A comprehensive review of bft consensus algorithms," ACM Computing Surveys, vol. 56, no. 5, pp. 1–41, 2024.

[6] A. Hopkins, "A fault-tolerant information processing concept for space vehicles," IEEE Transactions on Computers, vol. C-20, no. 11, pp. 1394–1403, 1971.

[7] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," in Concurrency: the works of leslie lamport, 2019, pp. 203–226.

[8] Y. Li and C. Tan, "A survey of the consensus for multi-agent systems," Systems Science & Control Engineering, vol. 7, no. 1, pp. 468–482, 2019.

[9] A. Baronchelli, "The emergence of consensus: a primer," Royal Society open science, vol. 5, no. 2, p. 172189, 2018.

[10] S. P. Borgatti and D. S. Halgin, "Consensus analysis," A companion to cognitive anthropology, pp. 171–190, 2011.

[11] H. Xu, Y. Fan, W. Li, and L. Zhang, "Wireless distributed consensus for connected autonomous systems," IEEE Internet of Things Journal, vol. 10, no. 9, pp. 7786–7799, 2023.

[12] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and industry 5.0—inception, conception and perception," Journal of manufacturing systems, vol. 61, pp. 530–535, 2021.

[13] K. Makovi, A. Sargsyan, W. Li, J.-F. Bonnefon, and T. Rahwan, "Trust within human-machine collectives depends on the perceived consensus about cooperative norms," Nature communications, vol. 14, no. 1, p. 3108, 2023.

[14] M. H. Tessler, M. A. Bakker, D. Jarrett, H. Sheahan, M. J. Chadwick, R. Koster, G. Evans, L. Campbell-Gillingham, T. Collins, D. C. Parkes et al., "Ai can help humans find common ground in democratic deliberation," Science, vol. 386, no. 6719, p. eadq2852, 2024.

[15] G. Bracha, "Asynchronous byzantine agreement protocols," Information and Computation, vol. 75, no. 2, pp. 130–143, 1987.

Yixuan Fan received the B.Eng. degree in communication engineering from University of Electronic Science and Technology of China (UESTC) in 2021. She is currently pursuing the Ph.D. degree with the James Watt School of Engineering, University of Glasgow. Her current research interests include distributed consensus, autonomous driving, blockchain, and decentralized autonomous organizations.

Huanyu Wu is currently pursuing a Ph.D. in Electronics and Electrical Engineering at the James Watt School of Engineering, University of Glasgow, U.K. His research interests include distributed systems, blockchain, security and reliability, and applied cryptography.

Zhongxu Dong received the B.Eng. degree in electrical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2021. He received the M.Sc.R. degree with the Agile Tomography Group, School of Engineering, the University of Edinburgh, Edinburgh, U.K. He is currently pursuing the Ph.D. degree with the Trustworthy lab, James Watt School of Engineering, the University of Glasgow.

Zongyao Li received the B.Eng. degree in electronics and electrical engineering from University of Electronic Science and Technology of China (UESTC) in 2020. He is currently pursuing the Ph.D. degree with the James Watt School of Engineering, University of Glasgow. His current research interests include distributed consensus, wireless communication system, connected and autonomous system.

Lei Zhang (Senior Member, IEEE) is a Professor of trustworthy systems with the University of Glasgow, U.K. His research spans wireless communications, distributed systems, blockchain, IoT, and autonomous systems. He holds 20 patents filed in over 30 countries/regions and has published over 200 peer-reviewed papers and three books. He is the Founding Chair of the IEEE Special Interest Group on Wireless Blockchain Networks and has delivered tutorials at leading IEEE conferences, including ICC, Globecom, PIMRC, VTC, ICBC, and EUSIPCO.