Tech and the City: Axialization, Institutionalization and Disruption*

Bud Mishra

Courant Institute, New York University, New York, USA mishra@nyu.edu https://cs.nyu.edu/mishra/

Abstract. "Something, something big, was happening in multiple places along the axis running from China to the Mediterranean [and beyond] in the first millenium BCE..." That axial age spurred by urbanization and with guidance from axial sages invented modernity (logic and institutions, rationality and strategic interactions and education and participatory civic society.) How should the smart cities of the future reaxialize to withstand deceptive interference, isolation against pandemics, cellularized instituions, etc.?

Keywords: Causality and Probabilistic Temporal Logic \cdot Deception and Signaling Games \cdot Safety, Verifiers, Liveness and Recommenders.

1 Introduction

Ruminating on and off about relocating elsewhere, I have been wondering since 2016 what our new (smart) home cities could resemble. Athens or Sparta? Kapilavatsu or Rajagriha? Qufu or Jinan? These are the home towns of the Axial age (600-350 BCE) philosophers: Socrates, Buddha and Confucious. As Karl Jaspers [1] recounted in his description of the Axial Age: "Confucius and Lao-Tse were living in China; India produced the Upanishads and Buddha; Greece witnessed the appearance of Homer, of the philosophers. Everything implied by these names developed during these few centuries almost simultaneously in China, India and the West."

It has been speculated that the Axial Period coincided with the rise of Mega Cities, whose survival depended on human reciprocal altruism, empathy and logic - but were plagued by tyranny, suffering and demagoguery. My talk at $SC360^{\circ}$ in Braga (Portugal) focused on these ingredients for smart cities: **Deductive Logic** of the west, **Samsaric Game Theory** of India and **Junzi Data Science** of China – which shape our Privacy and Trust, (Cyber) Security, Smart/Safe Households via Cellularization, Hastily Formed Networks and Identity Management.

There are two recent developments along these lines, namely, (1) Perhaps Jaspers' Axial Age theory is largely incorrect (Seshat History of the Axial

^{*} Supported by and Army Research Office Grant: W911NF-18-1-0427.

Age?). The standard assumptions of Axialization – its specific temporal boundary, its specific geographic boundary, its critical transformation in culture, ideology and religion, its spread by diffusion, and its identifiable "sages" rooted in the major largely imperial polities of Eurasia – have begun to be questioned. (2) Even though the Axialization narrative may be wrong in details, the process lifting tribal humans into "modernity" is largely captured by Jaspers' observation[3], and very relevant to the emerging cyber-physical societies, to which the smart cities and their "digital modernity" must adapt. The emerging digital-physical technologies for IoHT (Internet of Humans and Things) seem to be spiraling in to an amalgamation of Logic (Smart Contracts and Verification), Game Theory (Signaling Games with Costly Proof-of-work Signaling to maintain consistent consensus Digital Ledgers) and Statistical Inference aware of Privacy and Security (Explainable AI to maintain Recommenders and Verifiers). Thus the digital civil polities (e.g., governance) will be central to the mega-cities of the future as much as energy efficiency, supply chains, pollution control, environmental concerns, driver-less transportation, and electronic commerce. We propose a design around the technologies of Probabilistic Temporal Logic (e.g., PCTL), Model Checking, Signaling Games, Digital Ledgers and Reinforcement Learning and Bandit Problems.

2 Logic: Models and Model Checking

In logic, temporal logic refers to a system of rules and symbolism for representing, and reasoning about, propositions qualified relative to various "modes" of time (for example, "I am always shopping," "I will eventually Uber to a shop," or "I will be shopping until I am dropping of exhaustion;" time is treated in terms of its topology). Given a Kripke Structure model, there are efficient decision processes to "check" if a certain temporal property (expressed in the logic) holds true. How can smart urban societies deploy dynamically emerging (social) contracts and their rigorous logical verification via data and technology recording interactions over time?

Thus not just *relations* – neither genetic (e.g., possibly, leading to nepotism) nor reciprocation-based trust (e.g., possibly, leading to tyranny and demagoguery) – provide a good foundation for the establishment of a complex cosmopolitan urban society and the evolution of social contracts needed to glue together the strategic, shrewd and selfish members of a complex system. Such an evolving dynamic structure must introspect its inner working and understand the causes and effects in order to create robust and stable social contracts, constitutions and laws, which must be recommended, executed, recorded and verified using the best technologies available to a smart urban society and its governing body.

The mathematical/logical underpinnings of Probabilistic Causation are easily expressible in the logic below, which also allows efficient model checking in general. Thus enumerating complex *prima facie* causes from data or probabilistic state transition models becomes feasible. Thus, starting with a discrete time Markov chain (DTMC) – a directed graph with a set of states, S, it is endowed (via labeling functions) with the atomic propositions true within them.

It is then possible to make the labeling probabilistic, so that one may express that with some non-negligible probability, one's (e.g., Athen's) "optimistic views of democrcacy" may be false and may not suffice to harmonize other neighboring cities (e.g., Sparta) into a better governance – thus, "truthiness" of such a statement could have been used to avoid long and devastating wars. A city's laws, logic and $\lambda o \gamma o \zeta$ could create its cosmos, ordering it and giving it form and meaning!

The states of the underlying (Kripke structure) model are related pairwise by the transition probability. We also have an initial state from which we can begin a path (trajectory) through the system. Each state has at least one transition to itself or another state in S with a non-zero probability. A general framework for causality analysis is provided by model checking algorithms in PCTL (Probabilistic Computational Tree Logic) and has been explored in details elsewhere [9]. We have shown there how Suppes' prima-facie causality can be formulated in PCTL, and also suggested developing an efficient, albeit simplified, approach to testing contracts using Kripke-models and SBCN (with pair-wise causality represented as edges in a graphical model) – originally introduced elsewhere. See [7, 4].

Definition Probabilistic Computational Tree Logic, PCTL The types of formulas that can be expressed in PCTL are path formulas and state formulas. State formulas express properties that must hold within a state, determined by how it is labeled with certain atomic propositions, while path formulas refer to sequences of states along which a formula must hold.

- 1. All atomic propositions are state formulas.
- 2. If f and g are state formulas, so are $\neg f$ and $f \land g$.
- 3. If f and g are state formulas, and t is a nonnegative integer or ∞ , then $fU^{\leq t}g$ is a path formula.
- 4. If f is a path formula and $0 \le p \le 1$, then $[f]_{>p}$ is a state formula.

The syntax and the logic builds on standard propositional Boolean logic, but extends with various modes: the key operator is the metric "until" operator: $fU^{\leq t}g$: here, use of "until" means that one formula f must hold at every state along the path until a state where the second formula g becomes true, which must happen in less than or equal to t time units. Finally, we can add probabilities to these "until"-like path formulas to make state formulas.

Path quantifiers analogous to those in CTL may be defined by: $Af \equiv [f]_{\geq 1}$ [Inevitably f]; $Ef \equiv [f]_{>0}$ [Possibly f]; $Gf \equiv fU^{\leq \infty}$ false [Globally f], and $Ff \equiv \text{true } U^{\leq \infty}f$ [Eventually f]. Formal semantics of the PCTL formulæ may be associated in a natural manner. One can then say event f "probabilistically causes" g, iff f is temporally prior to g and f raises the probability of g

$$f \mapsto_{>p}^{\leq t} g \iff AG[f \to F_{>p}^{\leq t}g],$$

for some suitable hyper-parameters p > 0 (for probability raising) probability and t > 0 duration (for temporal priority). Additional criteria (e.g., regularization) are then needed to separate spurious causality from the genuine ones . SBCN, thus, provides a vastly simplified, and yet practical, approach to causality, especially when explicit time is not recorded in the data. There are efficient algorithms to ensure that smart contracts in PCTL facilitate a future, thus keeping the city's activities *alive*, and to ascertain *safe* behavior in the past by model checking specifications in PCTL.

3 Games: Signaling and Deception

Urban societies of the future will be structured around anonymous citizens interacting rationally (e.g, Dharma) and strategically (e.g., Karma) to improve their utilities, even though other individuals' "types," identity and personal information must be allowed to remain private. Deceptive behavior in the cities could be rampant and must be tamed. Safe house-holds in the city will protect the citizens by a "cellularization-process," which must include (in a cell) humans, pets and things associated by familial relations. Multi-cellular neighborhoods may emerge, experiment, persist and extinguish by "hastily formed networks," and more permanent Intra- and Internets – some explorable and some dark! Game theory – be it evolutionary, or epistemological – provides a forum in which such dynamics may be studied and moulded.

Game theory involves the study of the strategies followed by individuals, and organizations, in situations of conflict and cooperation. A Nash equilibrium refers to a certain mixture of strategies where a unilateral change in strategy by one player will not bring any benefit to it [2]. Maynard Smith pioneered the use of game theory in evolutionary biology, developing the concept of the evolutionarily stable strategy (ESS). An ESS is a form of Nash equilibrium in a population where a mutant with a variant strategy cannot successfully invade. Replicator dynamics addresses the dynamics of fitter players (which possess superior utility) that preferentially replicate within a population [cccc]. An important contribution in these types of evolutionary games was the recognition that there is no need for epistemologically aware agents given that the players could be non-human Bots which may use black-box AIs, unable to consciously adopt strategies.

Signalling game theory is a branch of game theory that was developed concurrently in both economics and organismal evolutionary biology in the 1970s, and it involves the sending of signals, honest or deceptive, from a sender to a (possibly, many) receiver(s) [10]. Information asymmetry occurs when the sender possesses information about its type, that is not available to the receiver, thus the sender can choose whether or not to reveal its true type to the receiver. In comparison to organismal evolutionary biology, cyber-physical evolution has made lesser use of concepts from game theory, but with a growing number of contributions (e.g, cyber security), for example [6]. In addition, microbial ecology has made use of evolutionary game theory to explain cooperative interactions where metabolites are public goods shared between microbes. Not unlike in the Buddhist axialization (evolving a Samsara created by a deceptive Mara/Maya), one can build technologies founded upon signalling game theory that has great explanatory power for a range of social processes, by pinpointing the 'strategies' of humans and things in their interactions with other humans and things. In doing so, we also wish to highlight commonalities between signalling behaviour at the device, organismal, human and social levels.

To understand whether in a city – smart or otherwise – undesirable outcomes may arise in the form of deception, we may call upon the theory of informationasymmetric signaling games, which unify many of the adversarial use cases under a single framework. In particular one may be interested in situations when adversarial actions may be viewed mathematically as rational (i.e., utility-optimizing agents possessing common knowledge of rationality). The simplest model of signaling games involves two players (e.g, a Uber driver and a passenger). They are asymmetric in information (driver may not be told passenger's destination until they agree to engage in a ride). They are called S, sender (informed passenger), and R, receiver (uninformed driver). A key notion in this game is that of type, a random variable whose support is given by T (known to sender S). Also, we use $\pi_T(\cdot)$ to denote probability distribution over T as a prior belief of R about the sender's type (e.g., a Uber driver may guess a possible destination based on the source and time of the journey). A round of game proceeds as follows:

- Player S learns $t \in T$;
- S sends to R a signal $s \in M$; and
- -R takes an action $a \in A$.

Their payoff/utility functions are known and depend on the type, signal, and action:

$$u^i: T \times M \times A \to R$$
, where $i \in \{S, R\}$.

In this structure, the players' behavior strategies can be described by the following two sets of probability distributions: (1) $\mu(\cdot|t)$, $t \in T$, on M and (2) $\alpha(\cdot|s)$, $s \in M$, on A. For S, the sender strategy μ is a probability distribution on signals given types; namely, $\mu(s|t)$ describes the probability that S with type t sends signal s. For R, the receiver strategy α is a probability distribution on actions given signals; namely, $\alpha(a|s)$ describes the probability that R takes action a following signal s. A pair of strategies μ and α is in Nash equilibrium if (and only if) they are mutually best responses (i.e., if each maximizes the expected utility given the other):

$$\sum_{t \in T, s \in M, a \in A} u^{S}(t, s, a) \pi_{T}(t) \mu^{*}(s|t) \alpha(a|s)$$

$$\geq \sum_{t \in T, s \in M, a \in A} u^{S}(t, s, a) \pi_{T}(t) \mu(s|t) \alpha(a|s); \qquad (1)$$

and

$$\sum_{t \in T, s \in M, a \in A} u^R(t, s, a) \pi_T(t) \mu(s|t) \alpha^*(a|s)$$

$$\geq \sum_{t \in T, s \in M, a \in A} u^R(t, s, a) \pi_T(t) \mu(s|t) \alpha(a|s);$$
(2)

for any μ, α . It is straightforward to show that such a strategy profile (α^*, μ^*) exists. We conjecture that the natural models for sender-receiver utility functions could be based on functions that combine information rates with distortion, as in rate distortion theory (RDT). For instance, assume that there are certain natural connections between the types and actions, as modeled by the functions f_S and f_R for the sender and receiver respectively:

$$f_S: T \to A; \quad f_R: A \to T.$$
 (3)

Then the utility functions for each consist of two weighted-additive terms, one measuring the mutual information with respect to the signals and the other measuring the undesirable distortion, where the weights are suitably chosen Lagrange constants

$$u^{S} = I(T, M) - \lambda_{S} d^{S}(f_{S}(t), a), \quad \&$$

$$u^{R} = I(A, M) - \lambda_{R} d^{R}(t, f_{R}(a)), \quad (4)$$

where I denotes information and d^R, d^S denote measures of distortion.

This definition also captures the notion of *deception* as follows. Note that the distribution of signals received by R is given by the probability distribution π_M , where

$$\pi_M(s) = \sum_{t \in T} \pi_T(t) \mu(s|t), \tag{5}$$

and the distribution of actions produced by R is given by the probability distribution π_A , where

$$\pi_A(a) = \sum_{s \in M} \pi_M(s) \alpha(a|s).$$
(6)

Clearly π_T and π_A are probability distributions on T and A respectively. If $\hat{\pi}_T$ is the probability distribution on T induced by π_A under the function f_R , then

$$\hat{\pi}_T(\cdot) := \pi_A(f_R^{-1}(\cdot)). \tag{7}$$

A natural choice of measure for deception is given by the relative entropy between the probability distributions π_T and $\hat{\pi}_T$:

Deception := Rel. Entropy
$$(\hat{\pi}_T | \pi_T)$$

= $\sum_{t \in T} \hat{\pi}_T(t) \log_2 \frac{\hat{\pi}_T(t)}{\pi_T(t)}$. (8)

This definition describes deception from the point of view of the receiver. To get the notion of deception from the point of view of the sender, one needs to play the game for several rounds. The equation implies that deception can be both defined as the sending of misleading information, or the witholding of information, both in order to manipulate the receiver. The Shapley value describes the distribution of utility to different players in a cooperative game. In a signaling game where deception occurs the value is skewed towards the sender.

4 Institutions: Recommenders and Verifiers

It is inevitable that predictive learning systems (ML, Machine Learning and AI, Artificial Intelligence and Formal Methods) will play an important role in the urban societies of the future. The inferences obtained by these systems will manifest themselves in social networks induced by "trust" relations, where trust may be measured by a "correlation of encounter." In other words, if one selects to rationally (but, possibly, information asymmetrically) and strategically interact with another individual, how likely would it be to choose to interact with the same individual repeatedly – in other words how trustworthy is the other individual thus encountered. There have been growing interest in distributed permissionless and trustless systems supported by distributed ledgers (e.g., Kripke Structures) and non-strategic verifiers (e.g., miners with costly signaling supported by proof-of-work), the technology still remains in its infancy. Not unlike Confucian "scholars," and "state officials," we may envision machine learning to produce a system of "recommenders," and "verifiers," - duals yin and yang serving negative and positive aspects of aspirational and traditional values - said differently, the evolutionary (replicator) dynamics set forth by variations and selections. Recommenders and verifiers are non-strategic, perform costly signaling to display trustworthiness and augment intelligence of the cities' humans and things, who are nonetheless strategic; they thus, rationally optimize their individual utilities. Agents (e.g., humans and things) then "virtualize" themselves by selecting a tribe of suitable recommenders and verifiers, while maintaining their privacy and strengthening their trust relations; possible distributed algorithms and policies, for this purpose, may be built upon adversarial bandits (with interpretations).

One may be inspired by the Chinese Axial Sage, Kong Qiu (a name confusingly translated as Confucius by Matteo Ricci [1552-1610], a Jesuit Minister). Confucius assumed (1) humans' (and things') ability to develop "morally," (2) using self-cultivation (by rationally modeling utility sought), (3) thus perfecting the world (state, city, or cell), (4) for which purpose, examples of "sages" (data, recommenders and verifiers) may be used. Thus ultimately, cellularization, and supporting institutions, for developing and evaluating recommenders and verifiers would be AI's main contributions to the cities of the future.

Returning back to the framework of signalling games, one notes that the Nash equilibria of these signaling games fall into few classes: (1) desirable separating equilibrium, albeit conventional or (2) uninteresting pooling equilibrium [or combinations there of, in partial pooling equilibrium]

- Separating Equilibrium: Each type t sends a different signal M_t .

$$f^S: t \mapsto a[M_t]$$

- Pooling Equilibrium: All types t send a single signal s^* with almost surely.

Thus in order for the Internet of the future to be relevant to physical smart urban societies, it (e.g., hyper-visor on a cloud) must be aware of the (partially)

observed data and meta-data involved in signaling on the Internet and the underlying inter-twined sender-receiver games. For example, highly relevant to the Internets' revenues and returns-on-investment are signals involved in Google Search queries (with the users' state of ignorance remaining private), key-words (private to Google) and advertisement selected by auction in an Ad-exchange (private to product developers) – utilities respectively being: page relevance for the user, return on investment for the advertisers, and customer satisfaction and retention for the publishers. A key proposal to tame deception in these systems would be to control them by non-strategic Recommenders and Verifiers.

- Recommenders ensure Liveness: $\forall_A \exists_T \exists_S U_S(T, M, A) \ge \theta^*$;
- Verifiers ensure Safety: $\forall_T \exists_A \exists_R U_R(T, M, A) \ge \theta^*$.

thus acting as correlating devices helping the entire system to evolve towards good separating equilibria, albeit conventional.

5 Cities & the Techs: Internet of Humans and Things

The technologies must be built around various Cyber-physical Games important to urban societies: e.g., Cellularization, Kripke-ledgers, model-checking and ZK (Zero Knowledge). In summary, following protocols may be devised to play a role to organize safe households in the cities.

- 1. Game Protocols.1
 - -S =Sender (Informed) $\mapsto R =$ Receiver (Uninformed)
 - The game may reach a Nash equilibrium that permits deception, but in this scheme it is tamed by Checkers. The checkers verify
 - Local (Propositional Logic Properties) [using CRYPTO]
 - Global (Modal Logic Properties) [*KRIPKE-LEDGER*]
 - The system uses asymmetric cryptography, described earlier:
 - Public/Verification Key: VR_S and VR_R
 - Private/Signature Key: SG_S and SG_R
 - Keys are linked via *COMPUTATIONALLY-ONE-WAY-FUNCTIONS*: e.g., McEliece scheme.
- 2. Game Protocols.2
 - (a) S generates SG_S and VR_S , and publishes only VR_S .
 - (b) S detects (i) type/state $t \in T$, (ii) message $m \in M$ and (iii) a time stamp τ .
 - (c) S sends an augmented message

$$C \equiv (VR_S, VR_R, m, \#\langle t, m \rangle, \tau) \|_{SG_S}$$

- 3. Game Protocols.3
 - (a) R ensures that S sent the message $C \parallel_{VR_S} \mapsto VR_S, VR_R$, etc.
 - (b) Check Local properties ... e.g., m is consistent with t: $\mathbf{F}(t, m)$.
 - (c) Check Global properties [using model checking and ZK SNARKs] ... e.g., $t_{\tau_1}, t_{\tau_2}, \cdots$ satisfy some modal properties: $\mathbf{G}(t, m)$.

Tech and the City: Axialization, Institutionalization and Disruption

- (d) R performs an action a consistent with m:
 - -S gets utility $U_S(t, m, a)$
 - R gets utility $U_R(t, m, a)$
- 4. Game Protocols.4
 - (a) To check the global properties and to be strategic, the players need access to the records of t's, $\#\langle t, m \rangle$'s and $U^{S,R}(t, m, a)$'s over time.
 - (b) For this purpose one creates a *KRIPKE-LEDGER*: a distributed database that maintains a continuously-growing list of data records hardened against tampering and revision.
 - (c) KRIPKE-LEDGER is maintained by *MINERS* who are subject to costlysignaling via proof-of-work or proof-of-soeace related to certain intractable computational problems.
- 5. SPVs, Coalitions and Intermediaries
 - (a) An intermediary (e.g., a House Holder) may be interested in only a particular group of players/humans/things (senders and receivers).
 - (b) The intermediary must not reveal membership information.
 - (c) The intermediary checks certain local and global properties about the players and publishes the results.
 - (d) The intermediary convinces a member that he is truthful. (Using zk-SNARK's for Propositional Modal Logic).
 - (e) Trivial Corollary: The players can make smart contracts with one another: Futures, Derivatives, Bonds subject to Positive and Negative Covenants, etc.

5.1 Example: A Library

One may cellularize a subset of households to create a safe and secure ways of sharing data, files (e.g., books and music), computation and things (e.g., childrens' toys) where families can get together for community and civic activities without leading to security problems or lack of fairness (e.g., tragedy of the commons)¹. The process begins with many households joining to create a hastily formed (ephemeral) network and may use Bare Metal as a Service; in other words, user (e.g., each household) gets a physical machine, can install (open source) firmware, hypervisor, OS, etc. All communications are performed encrypted. It may require that providers controlling the network can only deny service, but not snoop. Furthermore, results of computation may be hidden, and computation obfuscated (further incorporating Differential Privacy, Multi-Party Computing, Erasure Coding – data makes sense only when k out of n pieces come, etc.) The participants (humans and things) interact subject to enforceable smart contracts.

¹ Jointly with Larry Rudolph, VP TwoSigma and MIT

6 Conclusion

Smart Mega Cities of the future require novel philosophical bases, computation and governance. If Axial Age Sages can be our inspiration; our urban societies will use logic, game theory and data science; model checking, credible and noncredible threats (with costly signaling) and reinforcement learning (with capabilities for intervening and interpreting regularly); central governance (monarchy and tyranny, critically analyzed by Qufu's Confucius), rank-based governance (oligarchy and aristocracy, critically analyzed by Shakya-tribe's Buddha from Lumbini) and decentralized governance (politiae and democracy, critically examined by Athenian Socrates). With new technologies (based on data sharing, crowd sourcing and gig economy) these questions have now come to forefront, but we seem to be " weeping and wailing from being united with the unloved [deception] and separated from the loved [honesty]²," not able to foresee how the Internets could be rescued from fragmentation and collapse. Are we then building a city of logos, logic and lawfulness or just another fortified favela of mis-communicated signals, waiting to collapse like a modern day Tenochtitlan?

References

- Assmann, Aleida: Jaspers' Achsenzeit, oder Schwierigkeiten mit der Zentralperspektive in der Geschichte, in Karl Jaspers: Denken zwischen Wissenschaft, Politik und Philosophie, 187–205,(1989)
- Binmore, Ken, Game Theory: A Very Short Introduction, Oxford University Press, (2007)
- Black, Antony, The "Axial Period:" What Was It and What Does It Signify?, 70(12), 23–39, The Review of Politics, (2008)
- Bonchi, Francesco, Gullo, Francesco, Mishra, Bud and Ramazzotti, Daniele, Probabilistic Causal Analysis of Social Influence, Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018, 1003–1012, (2018)
- 5. Byrom, T., Dhammapada: The Sayings of the Buddha, Shambhala pocket classics, (1993)
- Casey, William, Kellner, Ansgar, Memarmoshrefi, Parisa, Morales, Jose Andre and Mishra, Bud, Deception, Identity, and Security: the Game Theory of Sybil Attacks, *Commun. ACM*, 62(1),85–93, (2019)
- Gao, Gelin, Mishra, Bud and Ramazzotti, Daniele, Causal Data Science for Financial Stress Testing, J. Comput. Science, 26, 294–304, (2018)
- 8. Hoyer, Daniel and Reddish, Jenny, Seshat History of the Axial Age, (2019)
- Kleinberg, Samantha and Mishra, Bud, The Temporal Logic of Causal Structures, UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009, 303–312, AUAI Press, (2009)
- Skyrms, Brian, Signals: Evolution, Learning, and Information, publisher =Oxford University Press, (2010)

² Cited from the Buddhist Text Dhammapada[5].