June 5, 2020
14:45 ~ 16:15 (online seminar), Keio University

Mechanism Design with Blockchain Enforcement

Hitoshi Matsushima
University of Tokyo

Shunya Noda
University of British Columbia

1. Introduction

Enforcement of Real-World Business Agreement

Agreed Action Profile $\alpha = (\alpha_1, ..., \alpha_n)$,
Work hard, send commodity, pay money, ...

Actual Action Profile $a = (a_1, ..., a_n)$
Skipped work, broke promise, pay nothing, ...

Player i is innocent $a_i = \alpha_i$: $\omega_i = 0$

Player i is guilty $a_i \neq \alpha_i$: $\omega_i = 1$

How can we penalize guilty players?

Legal Enforcement: Authorized and Trusted Court

Business Agreement

Agreed Action Profile

$\alpha = (\alpha_1, ..., \alpha_n)$

Actual Action Profile

$a = (a_1, ..., a_n)$

Player i is innocent $a_i = \alpha_i$

Player i is guilty $a_i \neq \alpha_i$

Authorized and Trusted Court

Trial i

Costly Verification to Third Parties
Mandatory Penalties
Privacy Infringement
Elimination of Illegal Activities

We proposed a new method of business enforcement: Blockchain (Self-) Enforcement

Matsushima (May 2019):

"Blockchain Disables Real-World Governance"
CARF-F-459, U-Tokyo.
Partial Implementation, Fear of Cartelization
journalistic a bit.

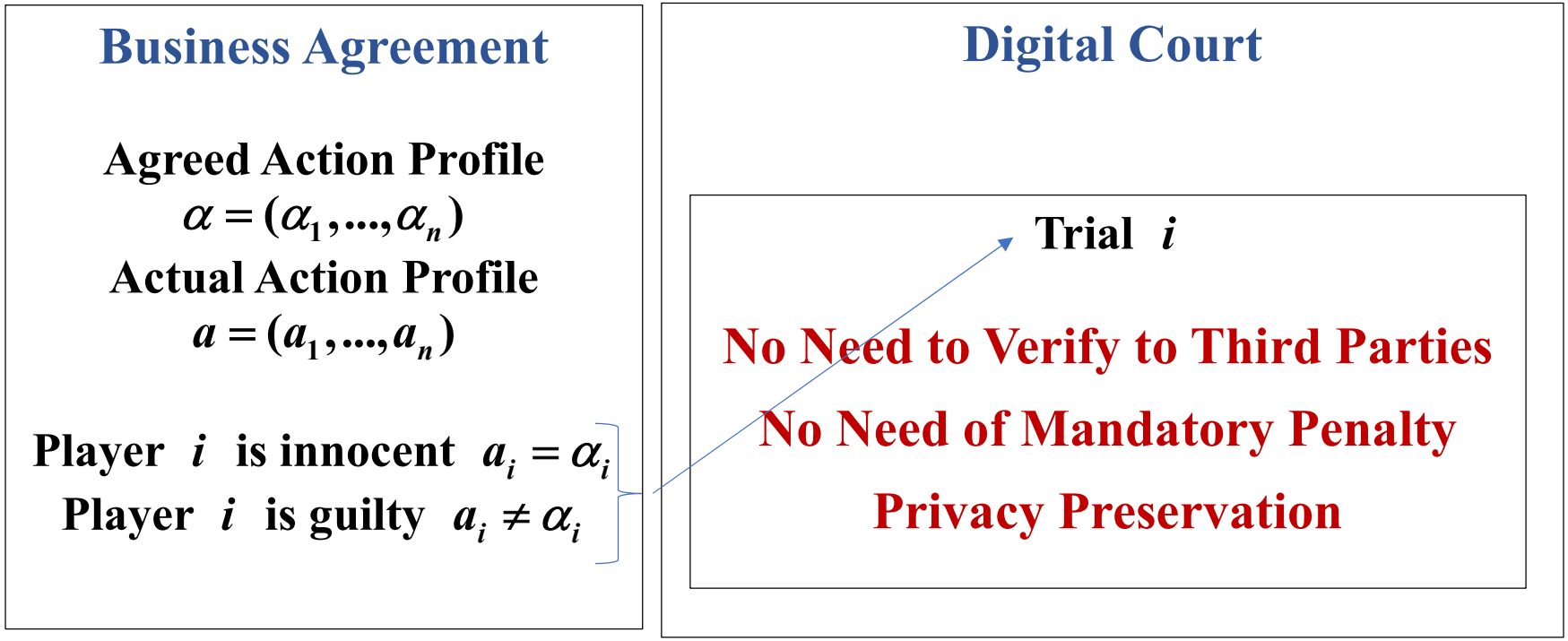
Matsushima and Noda (2020):

"Mechanism Design with Blockchain Enforcement"
CARF-F-474, U-Tokyo (First Version)
Unique Implementation, Behavioral Mechanism Design
more academic.

松島斉『千思万考：ブロックチェーンとゲーム理論』三菱経済研究所「経済の進路」連載全6回 2019年4月～9月

Blockchain Enforcement: Digital Court

(No Third Party, Self-Enforcement by Untrusted)



What is Blockchain?

- **Blockchain is a decentralized ledger technology.**
- **Blockchain works as an electronic payment system, managing ownership and transaction data in cryptocurrency in a tamper-proof manner.**
  - **Tradition electronic payment: Centralized and trusted record keeper backs a record to fiat money.**
  - **Blockchain electronic payment: Record is a transaction data of ownership transfer in cryptocurrency. Record keepers are decentralized and could be even untrusted, who do not back a record (cryptocurrency) to fiat money.**

Smart Contract

- **Cryptocurrency is programmable on blockchain.**
  → **Blockchain allows users to write input-contingent transfer rule as Smart Contract.**

- **Smart Contract is tamper-proof. Many steps of execution are automated.**
  → **Smart contract can be used as commitment device.**

Smart contract has limitations:

- **Only cryptocurrency transfers are available.**
- **All transactions on blockchain are publicly observable.**
  → **Parties (players) should not input details.**
- **Parties must input by themselves.**
  → **We need incentive design in smart contract (Oracle Problem).**

Players deploy Smart Contract as commitment device.

Smart Contract

$q = (q_i)_{i \in N}$, $q_i : M \to (-\infty, T_i)$

Players deploy q to blockchain.

$\to$ Each player i deposits T_i in cryptocurrency.

$\to$ They input messages $m = (m_i)$.

$\to$ Each player i automatically pays or burns $q_i(m)$.

Use smart contracts as Self-Enforcement

Digital Court

Trial i

$q^i = (q^i_j)_{j \in N}$, $q^i_j : M^i \to (-\infty, T^i_j)$

Each player $j \in N$ **inputs opinion** $m^i_j \in [0,1]$ **about whether player** i **is innocent** ($\omega_i = 0$) **or guilty** ($\omega_i = 1$)

Each player j **automatically pays or burns** $q^i_j(m^i)$.

Sentence $s^i = s^i(m^i) \in \{0,1\}$

Consider design of digital court strategically to incentivize players to input correctly!

Advantages of Digital Court

1. **Digital court needs no verification to third party:**
   We can save judicial expenses.

2. **Digital court preserves privacy:**
   Smart contract is publicly observable but includes no detail.

3. **Digital court can make any real-world agreement self-enforcing:**
   Untrusted, no long-term relationship

4. **Digital court needs no authority or monopolistic power of control:**
   We use programable money as automated, tamper-proof device.

5. **Our design of digital court is just simple:**
   Low commissions.
   It can be implemented with currently available technology, such as Ethereum (second in the industry).

2. Press Release: “A Digital Court for a Digital Age”

April 2020, U-Tokyo

HOME | UTokyo FOCUS

Features | Articles | Events | Press releases | Jobs | Find stories

PRESS RELEASES

Enter terms | Search

A digital court for a digital age

Researchers devise a way to perform legal functions with blockchain technology

Research news

Division for Strategic Public Relations | Graduate School of Economics / Faculty of Economics

Like 132 | Tweet

April 6, 2020

In a move to save time, money and effort, economics researchers utilized existing blockchain methodologies to create what they call a digital court. This would provide enforcement of contracts wherever a traditional legal court would currently settle disputes. Examples of areas which could make use of this would be auctions, business contracts and sales. As it is based on existing technology, it could be implemented right now.

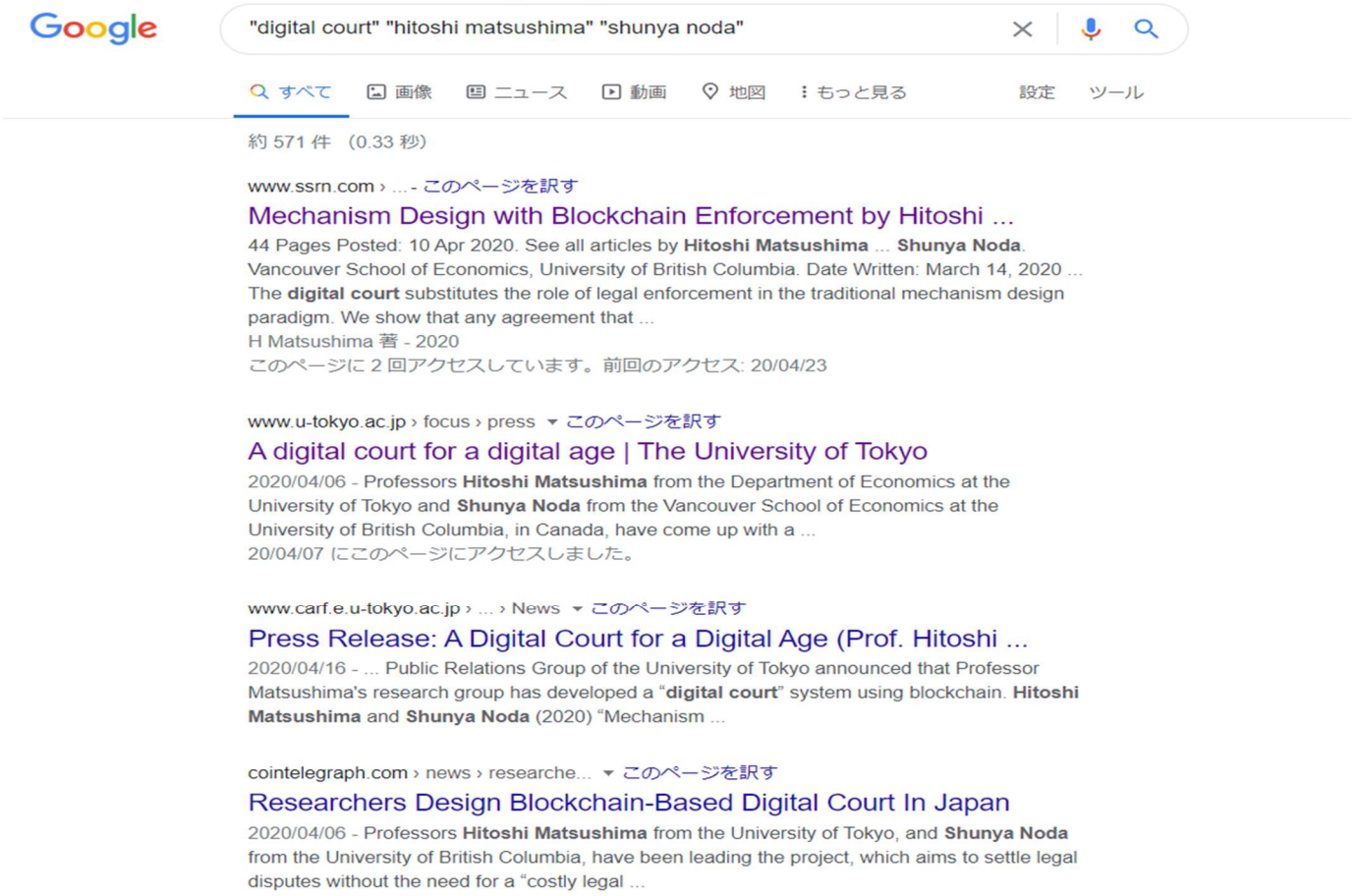
The digital court could open up commercial opportunities to those who cannot access traditional legal services. Image CC-0

Blockchain technology has great potential to impact many areas of life, commerce in particular. Put simply it is a way to ensure that information can be recorded in such a way that it cannot be manipulated afterwards. Blockchain is what is known as a distributed ledger, that is, there is no central authority, it is peer-to-peer, and its most famous application at this time is the online currency bitcoin. However, people find other uses for it.

Professors Hitoshi Matsushima from the Department of Economics at the University of Tokyo and Shunya Noda from the Vancouver School of Economics at the University of British Columbia, in Canada, have come up with a mechanism which uses blockchain to settle legal disputes without the need for an otherwise costly legal process. This is an extension to existing ideas for smart contracts which exist without central administration, but which until now have not found an application in the more general field of legal enforcement.

Media reacted immediately:

“Academics proposed a business model”



Ethereum

Ethereum News | Ethereum

Updated: April 10, 2020 3:08 pm EDT

Ethereum Founder Vitalik Buterin Shared Highlights On Decentralized Courts

By Steve Anderrson | April 10, 2020 3:08 pm EDT | 0

Share

Must Read

ETH/BTC |
May 23, 2020 3:31 pm EDT
Ethereum (ETH) Price Analysis: ETH Managing to Maintain Sustainability On $200

Analysis |
May 23, 2020 3:15 pm EDT
Ethereum Classic (ETC) Price Analysis: ETC Price Reclaims $6.50 Mark Opening

- Vitalik Buterin tweeted regarding the paper from Hitoshi Matsushima and Shunya Noda on blockchain-based smart contracts for arbitration.
- A mechanism to settle legal disputes without the involvement of a legal procedure utilizing the blockchain technology has been brought up by Professor Hitoshi Matsushima.

Vitalik Buterin, co-founder of Ethereum on April 10, 2020, tweeted regarding the paper from Hitoshi Matsushima and Shunya Noda on blockchain-based smart contracts for arbitration, also with the provision of the link of the same.

HOT NEWS

Ethereum (ETH) Price Analysis: ETH Managing to Maintain Sustainability On $200

Ethereum Classic (ETC) Price Analysis: ETC Price Reclaims $6.50 Mark Opening

Bitcoin Adoption Continues to Increase in the African Region

Tron Founder Faces Accusations of Criminal Conspiracy Amid Steem Hardfork

Kleros: Decentralized Court by Ethereum

(but after Matsushima, May 2019)

Kleros is substantially different from Digital Court.

Kleros needs verification to third parties.
Fear of infringing privacy

3. Model

Real-World (Business) Agreement

Agreed Action Profile $\alpha = (\alpha_1, ..., \alpha_n)$

Actual Action Profile $a = (a_1, ..., a_n)$

Actual action profile is observable with each other but may not be verifiable to third parties.

Player i **is innocent** $a_i = \alpha_i$**:** $\omega_i = 0$

Player i **is guilty** $a_i \neq \alpha_i$**:** $\omega_i = 1$

Irrespective of $\omega_i \in \{0,1\}$**, each player** $i \in N$ **becomes a defendant and has a separate trial on blockchain.**

All players take part in every trial as jurors.

Trial $i \in N$

All players deploy smart contract as transfer rule profile $q^i = (q^i_j)_{j \in N}$, where

$$q^i_j : M^i \to R, \quad M^i = \underset{j \in N}{\times} M^i_j, \quad \sum_{j \in N} q^i_j(m^i) \geq 0.$$

Each player $j \in N$ deposits $T^i_j \equiv \max_{m^i} q^i_j(m^i) \geq 0$ in cryptocurrency.

After all players selecting actual actions a offline, each player j inputs message $m^i_j \in M^i_j \subseteq [0,1]$ about whether defendant i is innocent or guilty.

Each player j automatically pays or burns $q^i_j(m^i)$.

Because of separate trials, we can focus on 'Trial 1'.

(Hence, omit subscript and superscript.)

Design I: $M_i = \{0,1\}$ for each $i \in N$

Juror i's Input: "innocent $m_i = 0$", or "guilty $m_i = 1$"

$$q_1(m) = T_1 s(m_{-1}) + \frac{\eta}{n-1}\sum_{k\neq 1}(m_1 - m_k)^2$$

$$q_i(m) = \frac{\eta}{n-1}\sum_{k\neq i}(m_i - m_k)^2 \quad \text{for all } i \neq 1,$$

where a sentence function $s : M \to \{0,1\}$ is specified as **majority rule:**

$$s(m_{-1}) = 1 \qquad \text{if } \sum_{i\neq 1} m_i > \frac{n-1}{2}$$

$$s(m_{-1}) = 0 \qquad \text{otherwise.}$$

- Defendant 1 is fined a large amount T_1 if she is convicted ($s(m_{-1}) = 1$).
- Each juror is incentivized to make her message close to the others' messages according to a simple form of private proper **scoring rule** $(m_i - m_k)^2$.

Suppose that each player j is rational (purely self-interested), i.e., minimizes $q_j(m)$ in expectation. Then, in Design I, both $m = (0,...,0)$ and $m = (1,...,1)$ are NE irrespective of $\omega \in \{0,1\}$.

Coordination Failure: Both truth telling and lying are NE in Design I.

We can generalize this impossibility:

Theorem 1: Suppose all players are rational. Then, irrespective of design of digital court, the set of all Nash equilibria is independent of whether the defendant is innocent or guilty.

If all players are rational, full (unique) implementation is impossible.

Why?

Only transfer is available online.

No relevant digital data are available online.

The fact "you didn't pay" can be picked up on line (?)

- **Any juror always prefers greater transfer irrespective of $\omega \in \{0,1\}$.**
- **We cannot use any incentive device explored in mechanism and contract design literature, such as VCG, Abreu-Matsushima, and various moral hazard schemes.**
- **We have the same impossibility result even if NE is replaced with any refinement.**

Hence, it is inevitable rational players have multiple equilibria.

We need a good explanation about
which equilibrium behavior rational players actually take and why so.

More specifically,
we must provide a good explanation about
when and why a rational player expects the other players
to behave honestly.

∴ Consider not only rational (pure self-interest) motive but also
behavioral motives

4. Behavioral Model

Incorporate behavioral aspects into mechanism design theory

Related Literatures:

Reputation Theory:	**Crazy Types**
	Gang of Four (1982)
Behavioral Mechanism Design:	**Preference for Honesty**
	Matsushima (2003)

Consider continuum message space $M_i = [0,1]$ instead of $M_i = \{0,1\}$.

Each player i announces $m_i \in [0,1]$ about how likely defendant 1 is to be guilty.

We assume that each player's type is:

Rational

Honest (behavioral) or

Adversarial (behavioral)

Rational (R) **just minimize monetary payment $q_i(m)$ (in expectation).**

Type R only considers financial gain.

Honest (H) **minimize (in expectation)**

$$\lambda_{i,H} q_i(m) + \left\{ \omega c^1_{i,H}(m_i) + (1-\omega) c^0_{i,H}(m_i) \right\}$$

where $(c^1_{i,H})' < 0,\ (c^1_{i,H})'' > 0,\ (c^0_{i,H})' > 0,\ (c^0_{i,H})'' > 0$

Type H considers both financial gain and psychological cost.
Type H prefers honest input and hates a big lie (convexity).

Adversarial (A) **minimize (in expectation)**

$$\lambda_{i,A} q_i(m) + \left\{ \omega c^1_{i,A}(m_i) + (1-\omega) c^0_{i,A}(m_i) \right\}$$

where $(c^1_{i,A})' > 0,\ (c^1_{i,A})'' > 0,\ (c^0_{i,A})' < 0,\ (c^0_{i,A})'' > 0$

Type A considers both financial gain and psychological cost.
Type A prefers dishonest input and hates an idiot honesty (convexity).

Remark: **We can envision more diverse behavioral types such as**

always announce "innocent 0"
always announce "guilty 1"
heterogeneity of honest types
heterogeneity of adversarial types
…….

We can generalize our model without substantial changes on this line.

Important Features of our behavioral model are:

A behavioral motive is state-contingent.
A behavioral type sticks to a specific pattern such as 'be honest' and 'be adversarial'.

Incomplete Information

The other players do not know which type:

Player i is	**honest (H)**	**with prob.**	$\delta_{i,H}$
	adversarial (A)	**with prob.**	$\delta_{i,A}$
	rational (R)	**with prob.**	$1-\delta_{i,H}-\delta_{i,A}$

We slightly modify Design I:

Design II: **Modify Design I by replacing $M_i = \{0,1\}$ with continuum message space:**

$$M_i = [0,1] \text{ for each } i \in N$$

Theorem 2 (Uniqueness): Suppose

$$\delta_{i,H} + \delta_{i,A} > 0 \text{ for some } i \in N.$$

Then, in Design II, irrespective of $\omega \in \{0,1\}$, there exists unique BNE, m^{ω}, as unique iteratively undominated strategy profile (dominance solvable), where we denote

$$m^{\omega} = (m_i^{\omega})$$
$$m_i^{\omega} = (m_i^{\omega}(R), m_i^{\omega}(H), m_i^{\omega}(A)) \in [0,1]^3.$$

Proof of Theorem 2 depends on:

- **Behavioral types are less elastic than rational (or stick to patterns)**
  - → **BNE is expressed by a fix point of some contractive mapping**
  - → **Uniqueness of fixed point (Edelstein's Fix Point Theorem, 62)**

- **Continuum of message spaces and continuity of preferences:**
  - → **Existence of fixed point**

- **Supermodular game: Convexity of psychological cost, proper scoring rule**
  - → **Uniqueness of BNE implies Dominance Solvable.**

Properties of Unique BNE

* $\left|\omega - m_i^{\omega}(X)\right|$ **is decreasing in** $\delta_{j,H}$ **and increasing in** $\delta_{j,A}$ **for all** i, $j \neq i$, **and** $X \in \{R, H, A\}$.

  ∴ **The more (less) honest a rational player behaves, the more likely the other players are to be honest (adversarial).**

* **Turning over A and H, we have** $1 - m^{\omega}$ **instead of** m^{ω}.

  ∴ **We have symmetry between A and H.**

*** (Very Important!) If**

$$\delta_{i,A} = 0 \text{ for all } i \in N, \text{ and}$$

$$\delta_{i,H} > 0 \text{ for some } i \in N,$$

then, we have

$$m_i^{\omega}(R) = \omega \text{ for all } i \in N.$$

∴ At least a single player could be honest (but a tiny prob). All players are never adversarial.

→ Rational types input full honesty.

Why? Tail-Chasing:
Any rational player attempts to announce more honestly than the average of the other rational players because of possible honest types.

→ Tail-chasing competition reaches full honesty.

From these properties we can say:

- **Whenever players are more likely to be honest type than adversarial type, correct judgement is supported by unique BNE.**

- **On the other hand, whenever players are less likely to be honest type than adversarial type, incorrect judgement is supported by unique BNE.**

Examples

Suppose $\lambda_{i,H} = \lambda_{i,A} = 0$, i.e., behavioral types never consider financial gains.

We can calculate unique BNE as reduced forms:

$$m_i^{\omega}(H) = \omega, \quad m_i^{\omega}(A) = 1 - \omega,$$

$$m_i^1(R) = \frac{\sum_{j \in N} m_j^1(R) - \delta_{i,H}}{n - \delta_{i,H} - \delta_{i,A}}$$

$$m_i^0(R) = 1 - \frac{\sum_{j \in N} \{1 - m_j^0(R)\} - \delta_{i,H}}{n - \delta_{i,H} - \delta_{i,A}}, \text{ where}$$

$$\sum_{j \in N} m_j^1(R) = \sum_{j \in N} \{1 - m_j^0(R)\} = \frac{\sum_{j \in N} \frac{\delta_{j,H}}{n - \delta_{j,H} - \delta_{j,A}}}{\sum_{j \in N} \frac{1}{n - \delta_{j,H} - \delta_{j,A}} - 1}.$$

Example 1 (Neutrality across Players): **Assume**

$$\delta_{i,H} = \delta_H \quad \text{and} \quad \delta_{i,A} = \delta_A \quad \text{for all} \quad i \in N.$$

We have

$$m_i^1(R) = \frac{\delta_H}{\delta_H + \delta_A} \quad \text{and} \quad m_i^0(R) = \frac{\delta_A}{\delta_H + \delta_A}.$$

If $\delta_H > \delta_A$, unique BNE yields correct judgement.
If $\delta_H < \delta_A$, unique BNE yields incorrect judgement.

Example 2: **Assume that each player is either potential honesty or potential liar, that is,**

$$\max[\delta_{i,H},\delta_{i,A}]=\delta \quad \text{and} \quad \min[\delta_{i,H},\delta_{i,A}]=0.$$

We have

$$m_i^1(R)=\frac{\tilde{n}-\delta}{n-\delta} \quad \text{and} \quad m_i^0(R)=1-\frac{\tilde{n}-\delta}{n-\delta} \quad \text{for} \quad \tilde{n} \quad \text{players,}$$

$$m_i^1(R)=\frac{\tilde{n}}{n-\delta} \quad \text{and} \quad m_i^0(R)=1-\frac{\tilde{n}}{n-\delta} \quad \text{for} \quad n-\tilde{n} \quad \text{players.}$$

With $\tilde{n}\geq 1/2$ **(potential honesties), unique BNE yields correct judgement.**

With $\tilde{n}<1/2$ **(potential liars), unique BNE yields incorrect judgement.**

∗ **(State-Contingent Belief): In Example 2, suppose that**

$$\text{Player } i \text{ is innocent } (a_i = \alpha_i) \rightarrow \delta_{i,H} = \delta \text{ and } \delta_{i,A} = 0$$
$$\text{Player } i \text{ is guilty } (a_i \neq \alpha_i) \rightarrow \delta_{i,H} = 0 \text{ and } \delta_{i,A} = \delta$$

Then, if more than half are guilty, not guilty but innocent is penalized in digital court.

Alternative Interpretation:

Player i is guilty, but not on purpose.
→ **The other players still believe $\delta_{i,H} = \delta$ and $\delta_{i,A} = 0$.**

5. Legal Purpose, Illegal Purpose, Logistics

Illegal Purposes with Real-World Logistics:

Illegal Drug smuggling is an example.

Prevent drug crime by observing real-world logistics.

Illegal Purpose without Real-World Logistics:

It is hard to crack down illegal behavior without real-world logistics:

Serious tension between privacy preservation and elimination of illegal activities.

Cartelization is an example.

Purpose-Contingent Beliefs:

Fear of adversarial type relieves sound business from illegal cartels.

Optimistic View:

Business is legal $\rightarrow$ $\delta_H > \delta_A$

Business is illegal $\rightarrow$ $\delta_H < \delta_A$

Pessimistic View:

Even illegal business satisfies $\delta_H > \delta_A$.

$\rightarrow$ **Whether legal or blockchain enforcement, any sound business chance fails due to fear of blockchain-based Cartelization.**

$\rightarrow$ **Regulator may have to ban blockchain use.**

6. False Charges

With adversarial type ($\delta_{i,A} > 0$), even innocent may be fined a large amount with a positive probability.

Innocent should not be fined a large amount.
Honest should not be fined a large amount.

→ We need alternative design of digital court
(maybe more complicated, and even ad hoc).

Design II': Each player inputs multiple messages to a trial:

$$m_i = (m_i(1),...,m_i(Z)) \in [0,1]^Z$$

$$q_1(m) = \frac{T_1}{Z}\sum_{z=1}^{Z} s(m_{-1}(z))$$

$$+\frac{\eta(1)}{n-1}\sum_{k\neq 1}\left\{m_1(1)-m_k(1)\right\}^2 + \sum_{z\neq 1}\eta(z)\left\{(m_1(z)-s(m_{-1}(z-1))\right\}^2$$

$$q_i(m) = \frac{\eta(1)}{n-1}\sum_{k\neq i}\left\{m_i(1)-m_k(1)\right\}^2 + \sum_{z\neq 1}\eta(z)\left\{(m_i(z)-s(m_{-1}(z-1))\right\}^2$$

for all $i \neq 1$

Incentive in 1-st Input: Scoring Rule $\left\{m_i(1)-m_k(1)\right\}^2$

Incentive in z**-th Input:** Punishment on distance from the sentence based on $(z-1)$-th $\left\{(m_i(z)-s(m_{-1}(z-1))\right\}^2$

We have Z **sentences:** Each sentence punishes defendant by T_1/Z.

Behavioral Model (Modified):

Honest **minimize (in expectation)**

$$\lambda_{i,H} q_i(m) + \sum_{z=1}^{Z} \lambda_{i,H}(z) \left\{ \omega c_{i,H}^1(m_i(z)) + (1-\omega) c_{i,H}^2(m_i(z)) \right\}$$

Adversarial **minimize (in expectation)**

$$\lambda_{i,A} q_i(m) + \sum_{z=1}^{Z} \lambda_{i,A}(z) \left\{ \omega c_{i,A}^1(m_i(z)) + (1-\omega) c_{i,A}^2(m_i(z)) \right\}$$

where $\sum_{z=1}^{Z} \lambda_{i,H}(z) = \sum_{z=1}^{Z} \lambda_{i,A}(z) = 1.$

$\sum_{z=1}^{Z} \lambda_{i,A}(z) = 1$ **hints implicitly about the possibility that both** $\lambda_{i,A}(1)$ **and** $\lambda_{i,A}(2)$ **are sufficiently small when** Z **is large.**

We assume $M_i(1) = [0,1]$

$M_i(z) = \{0,1\}$ **for all** $z \neq 1$

$\eta(z) > \lambda_{i,A}(z)\left\{c^0_{i,A}(0) - c^0_{i,A}(1)\right\}$ **for all** $i \in N$.

With this assumption, Design II' is dominance solvable, where any rational player inputs correctly from 2 to Z:

$$m_i^{\omega}(z;R) = m_i^{\omega}(z;H) = m_i^{\omega}(z;A) = \omega \quad \text{for all} \quad z \neq 1.$$

If both $\lambda_{i,A}(1)$ **and** $\lambda_{i,A}(2)$ **are small** ($Z \to \infty$), **we can set** $\eta(1) + \eta(2)$ **close to zero. Hence, Design II' solves false charge problem:**

Innocent defendant is never fined more than a small amount T_1/Z.

Honest juror is never fined greater than a small amount $\eta(1) + \eta(2)$.

7. Further Remarks

7.1. Coalition in Digital Court

A guilty defendant asks you:

"Please vote for innocence. I will give you $100."

Do you accept this request?

If you are trusted third party	**No**
If you are victim (Digital Court)	**No?** **Reciprocal Retaliation**
If you are untrusted third party (Kleros)	**Yes**

7.2. Deposit Savings

A player may have to deposit a large amount in advance:
cf. Auction: Bidders deposit before or after a win?

Can we save deposit?

→ **Consider dynamics with sequential business opportunities.**

- **We can reuse a deposit for many business purposes:**
  Matsushima (2012, JER)

- **We can add deposit little by little.**

8. Conclusion

- **We demonstrated blockchain enforcement, a new method.**
- **We introduced digital court as a commitment device in cryptocurrency. We characterized the case that a digital court makes business agreement self-enforcing.**
- **By replacing legal enforcement with blockchain enforcement, we can eliminate verification processes, saving judicial expenses and preserving privacy.**
- **This method, however, can be used in illegal applications such as cartelization. In the worst case, blockchain plucks the bud of all business opportunities.**
- **To understand appropriate policies, it is important as future research to develop systematic analysis in theory, experiment, and social implementation to unify real-world governance with incentive in digital court.**