2023年7月10日（月曜日）
16：00～17：30
SBI金融経済研究所セミナー（オンライン）

二つのデジタル・ディストピア

松島斉

東京大学大学院経済学研究科教授

デジタル社会

SNS、DX、Big Data、Social Graph、AI .....

信用スコア（見える化）：　Uber
食べログ
セサミクレジット（アリババ）
ジェイスコア（ソフトバンク、みずほ）
ESG スコア

ブロックチェーン（分散型台帳、非改ざん性）：
デジタル通貨
データ・システム管理
スマートコントラクト

今日のデジタル社会における不正行為

ネット犯罪：　　SNS

サイバー攻撃：　　DX、AI

近未来のデジタル社会における不正行為

Scientific Fictions：

二つのデジタル・ディストピア

信用スコア：　　国家を不正行為に駆り立てる

ブロックチェーン：　　一般市民を不正行為に駆り立てる

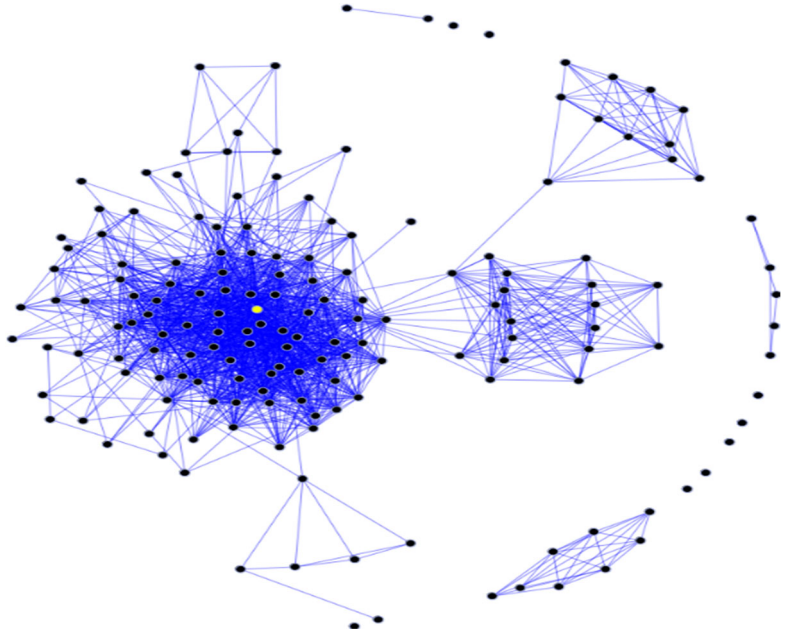
I. デジタル・ディストピアⅠ

Tirole “Digital Dystopia” 2021

「信用スコア」を悪用することによって
国家が不正行為をする

社会信用システム（中国２０１４～）

Big Data ＋ Social Graph ⇒ 個々人の信用スコアの産出



国家が信用スコアを不正利用する
(Digital Leninism)

国家は個人の信用スコアに以下の情報を混ぜ込む：

政治的態度：　　体制派か反体制派か

人間関係（Social Graph）：　　付き合いのある人のスコアを加味

高いスコアを目指そうとするインセンティブ

→　多くの市民を体制支持（模範的市民）に駆り立てる

→　高いスコアの市民は低いスコアの市民とかかわらなくなる

→　社会的絆の分断（村八分、コミュニティーの崩壊）
階層化（経済格差）

高いスコア：　体制派：　富裕層
低いスコア：　反体制派：　貧困層、Gettoization

SF 小説：　ハクスリー「すばらしい新世界」（１９３２）
オーウェル「１９８４年」（１９４９）

ディストピアＩの対策

- 信用スコアに余計な情報を入れさせないようにする
- 信用スコアにどのような情報がどのように使われているかをきちんと公開させる
- 複数のスコアが用意されて比較できるようにする
- 余計な情報が入っている疑いのあるスコアは無視するようにする

（いったいどうやって？？）

2. デジタル・ディストピア II

Matsushima “Blockchain Disables Real-World Governance” 2019

Matsushima and Noda “Mechanism Design with Blockchain Enforcement” 2020

ブロックチェーン（スマートコントラクト）を利用して
一般市民が不正行為をする

ブロックチェーン

松島斉：SBI 金融経済研究所所報 Vol. 3
「デジタル通貨とスマートコントラクト：ゲーム理論家の視点から」（2023 年 3 月）

デジタル通貨：　交換手段としての独立性：
　　法的通貨の裏付けなくても価値化
　　個人情報からの独立性（コインが本物か否か）：
　　cf. アカウント型（本人か否か）
　　非改ざん性

DLT：　ガバナンス：　PoW、PoS
　　独占の排除
　　環境問題：　PoW → PoS

スマートコントラクト

Ethereum

Programmable Currency (ETH)：

入力条件付きデジタル通貨移動が自動化

→ Smart Contract ＝ Credible Commitment

プライバシー問題：　スマートコントラクトの中身は第三者に筒抜け

オラクル問題：　現実空間からブロックチェーン（仮想空間）に「正直に」入力するインセンティブ設計が必要

デジタル法廷（Digital Court）

Matsushima and Noda　“Mechanism Design with Blockchain Enforcement” 2020

松島＋野田によるスマートコントラクトの利用方法

スマートコントラクトを利用することによって
「現実空間における約束」を守らせる（Enforcement）ための装置

プライバシー問題、オラクル問題も克服

例：A さんと B さんの約束の遂行

約束「A は行動 α をおこなう」
しかし実際の行動は a だった

$a = \alpha$:　　A is innocent
$a \neq \alpha$:　　A is guilty

例えば　　α　　この黒布を A が B に送る約束
　　a　　A は実際には「あの」黒布を B に送った

「$a \neq \alpha$（約束を破る）」をどのように阻止すればいいか?
約束違反すると罰金を払わせるようにしたらいい
はたしてどうやって？

約束違反を阻止するための三つのアプローチ

1．Legal Enforcement

2．Reputation

3．デジタル法廷

Legal Enforcement

違反が実際に発生した場合「事後的に」法律で裁いてもらう

「$a \neq \alpha$（A is guilty）」の時裁判所に訴え出る：

Cost Compensation

Criminal Punishment

実はあまり有効でない：　司法コスト大、プライバシー問題

立証不可能性：　「約束は α しかし実際は a」を他者に立証しないといけない

「この黒」と「あの黒」の違いを他者にどう説明すればいいか？（当事者同士は認識）

Reputation

Ａと Ｂ は長期的関係にあることを仮定：

当事者だけでインセンティブを解決：

約束違反 → 友好関係の停止というペナルティー

Multiplicity （フォーク定理）：

長期的関係には多様な可能性（協調、非協調など）

後付けの説明にすぎない

３．デジタル法廷

裁判所に約束の具体的内容を立証する必要がない

長期的関係を前提としない（一期一会）

スマートコントラクトの活用

プライバシー問題の解決

オラクル問題の解決：　　スマートコントラクトに内在する問題の解決
インセンティブと一意性の保証

デジタル法廷の作り方：5 Steps

「事前に」スマートコントラクトをブロックチェーンに設置することによって「Aが（実社会で）約束違反した場合に金銭的ペナルティーを科す」ための、改ざん不可能なコミットメントが成立する

約束違反か否かが判明した後、ＡとＢが各々情報入力し、スマートコントラクトにしたがって自動的にペナルティーが遂行される

∴　ペナルティーを避けるため実社会で約束を守るインセンティブが生まれる

正直に情報入力するのか（オラクル問題）
実社会の約束内容が第三者にどの程度漏洩するか（プライバシー問題）
がデジタル法廷の優位性のカギ

Step 1:　　Ａと Ｂは「現実社会における約束 α 」に合意する

Step 2: スマートコントラクト $(T_i, M_i, \gamma_i)_{i\in\{A,B\}}$ を作成し
ブロックチェーンに設置する（コミットメント）

デポジット　$T_A > 1$、$T_B = 1$、デジタル通貨

$M_i = [0,1]$　「Aが約束を守ったか否か」について
AとBが Step4 にて入力（$m_i \in M_i$）
Input ‘0’ implies A is innocent
Input ‘1’ implies A is guilty
曖昧（$0 < m_i < 1$）も認める

Punishment Rule:　$\gamma_i : M \to [0, T_i]$
デポジット $\gamma_i(m_1, m_2) \in [0, T_i]$ が戻らない（burned）

Step 3: 実社会においてAが行動 a を決定する

行動 a は二人にとって周知の事実（common knowledge）になると仮定する

Step 4: 「Aが約束を守ったか否か（$a = \alpha$ or $a \neq \alpha$？)」の確認後

A は 入力 $m_A \in [0,1]$, B は 入力 $m_B \in [0,1]$ を決定

同時に各自適当なデータ β_i も決定

まずハッシュ値 $h_i = h_i(m_i, \beta_i)$ を公開する

(m_i, β_i) は未公開とする

お互いにハッシュ値を確認後に (m_i, β_i) を公開する

$h_i = h_i(m_i, \beta_i)$ が成立することを確認する

(確認できない場合デポジットが全額返ってこない)

ハッシュ値は

$m_A \in [0,1]$ と $m_B \in [0,1]$ の「同時決定」を保証する役割をなす

Step 5: $\gamma_i(m)$ are burned：

A には　$T_A - \gamma_A(m)$　が返還される

B には　$T_B - \gamma_B(m) = 1 - \gamma_B(m)$　が返還される

Punishment Rules の具体的な設計方法

$$\gamma_A(m) = (m_A - m_B)^2 + (T_A - 1)\chi_{\{m_B > 1/2\}}(m_B)$$

$$\gamma_B(m) = (m_A - m_B)^2$$

デジタル法廷は A と B の「非協力ゲーム」

$(m_A - m_B)^2$ の値を小さくしたい
（相手と同じ入力値を選びたい）

デジタル法廷において求められている入力行動
＝ 正直入力

Aが約束を守る　→　正直入力　$(m_A, m_B) = (0,0)$

Aが約束違反する　→　正直入力　$(m_A, m_B) = (1,1)$

インセンティブ：経済主体の三つの行動動機

利己的（Selfish）：　相手と同じ入力をしたい

正直（Honest）：　相手よりも（少し）正直に入力したい

不正直（Adversarial）：　相手よりも不正直に入力したい

４．オラクル問題

不可能性定理：

「二人が利己的である」が周知の事実であると仮定する。この時、約束を守るか否かに関係なく、任意の $m_A = m_B \in [0,1]$ がナッシュ均衡になる．

正直か否かに関係なく
「相手と同じ入力」が全部均衡になってしまう

オラクル問題の解決

実証・実験結果：Abeler et al. (2019)

一般に、金銭が絡むと不正直（Adversarial）動機が消える

可能性定理：

Matsushima　Epistemological Implementation of Social Choice Functions 2022

「二人が利己的である」が周知の事実でないと仮定する。この時

Aが約束を守る　→　正直入力　$(m_A, m_B) = (0,0)$

Aが約束違反する　→　正直入力　$(m_A, m_B) = (1,1)$

が唯一のナッシュ均衡になる.

５．プライバシー問題の解決

ブロックチェーン自体はプライバシーにやさしい技術ではない
入力内容が第三者（record keepers）に公開されてしまうから

しかしデジタル法廷においては入力内容（０から１までの実数値）からは
セマンティクスを理解できない：

入力０：　「A が正しい黒布を送った」
　「A が殺人依頼を実行した」

当事者にはどっちか明らかだが第三者にはわからない

∴　デジタル法廷を使った行動はデジタル情報化されにくい
　cf. 信用スコア

6．ディストピアとしてのデジタル法廷

ユートピア：　　信頼関係にない人とビジネスできる
ex．クラウドファンディング

ディストピア：　　信頼関係にない人と談合が成立することによって、
「競争原理」が機能しなくなる

市場の崩壊：

オークションの入札者同士は談合して低価格で落札できる

→　誰も出品しない

組織の崩壊：

従業員が談合してサボる

→　評価基準を独立になるように設定しなければならない

→　固定給のみ

７．ディストピア II の対策

＊　デジタル法廷と実社会の法廷の違い

実社会の法廷では約束違反の疑いが生じた際に「事後的に」手続き（審査）がスタートする：

二つのチェック機能：　（１）約束違反が起きたかどうか
（２）約束内容そのものが不正かどうか

デジタル法廷では約束違反の疑いが生じる前に「事前的に」手続き（コミットメント）がスタートする：

ディストピアの原因の一端は
チェック機能（２）に対応する措置がないことにある

対策１（Ex-Ante Verification）：

スマートコントラクトに、約束自体が違法行為でないことを立証する情報を載せることの義務付け

cf. Ex-Post Verification (Legal Enforcement)

対策２（Ex-Ante Endorsement）：

スマートコントラクトの設置前に「信用力のある機関（銀行など）」にお墨付きをもらうことの義務付け

自生的解決の可能性について：

- 約束内容に関係なくデジタル法廷を利用する人は例外者であり、Adversarial である恐れがある

  → 誰もデジタル法廷を利用しようとしない（現状？）

- 約束内容が不正な場合のみ Adversarial である恐れがある

  → デジタル法廷が不正利用されるケースはまれである

  → みなが良い目的のために（のみ）デジタル法廷を積極的に利用する（ユートピア）

以上