Informational Robustness in Mechanism Design

Benjamin Brooks
UChicago

Northwestern Summer School on Robustness

August 25, 2026

What is mechanism design?

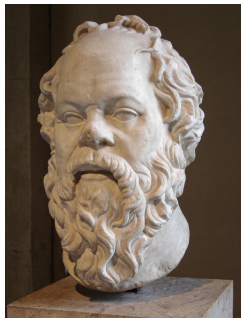
- A mathematical discipline studying processes for **social choice**, especially as it relates to
  - Optimization of welfare
  - Decentralized information
  - Individual incentives
- Choice can involve allocation of resources, production, and regulation of behavior more generally

Agenda for the lecture

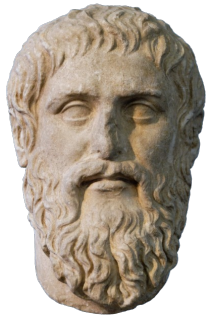
- This is a lecture on informational robustness in mechanism design
- But to understand informational robustness, one needs to know how we got here
- We will first review the history of mechanism design
  (Disclaimer: I am not a philosopher or a historian)
- Then I will segue to modern theories of robust mechanism design
  (Special focus on my work with Songzi Du and coauthors)

A Brief History of Mechanism Design: The Ancients

- Socrates/Plato (∼2500 ybp): In the *Republic*, *Meno*, *Statesman*, debating the correct form of government, how leaders should behave, the meaning and nature of justice
- Brilliant idea: Just pick great guardians to make decisions! Make sure they don't drink alcohol



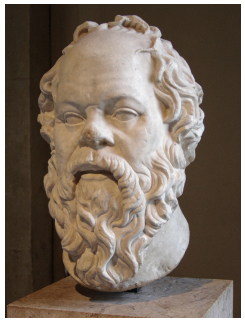
Socrates



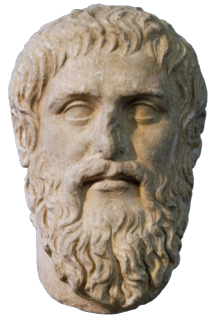
Plato

A Brief History of Mechanism Design: The Ancients

- Socrates/Plato (∼2500 ybp): In the *Republic*, *Meno*, *Statesman*, debating the correct form of government, how leaders should behave, the meaning and nature of justice
- Brilliant idea: Just pick great guardians to make decisions! Make sure they don't drink alcohol
- Juvenal (∼2000 ybp): Don't trust the guardians! They need to be guarded as well!



Socrates



Plato

Juvenal

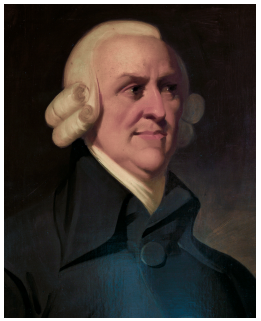
Constitutional convention of 1787 (219 ybp)

- A great debate, and intentional reflection, on the eternal question of how to select leaders, and how leaders should make decisions
- Federalism, regular elections, separation of powers, staggered terms, bill of rights
- Incorporated ideas from earlier political thought, e.g., magna carta, structure of colonial governments
- They anticipated how participants in the government would behave, sometimes correctly, sometimes incorrectly...

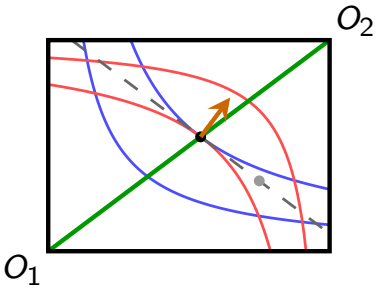
Christy, *Scene at the Signing of the Constitution* (1940)

Capitalism and Socialism (250–80 ybp)

- Adam Smith (d. 1790), Pareto, Walras, Edgeworth: "Invisible hand" of self interest, coordinated via prices, early forms of welfare theorems, efficiency of markets

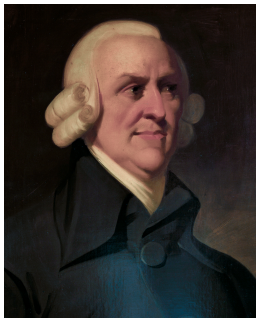


Adam Smith

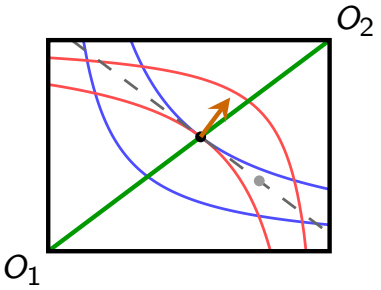


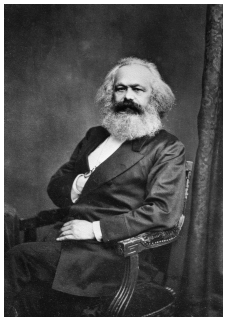
Capitalism and Socialism (250–80 ybp)

- Adam Smith (d. 1790), Pareto, Walras, Edgeworth: "Invisible hand" of self interest, coordinated via prices, early forms of welfare theorems, efficiency of markets
- Karl Marx (d. 1883): Critique of capitalism, inequality, class conflict
- Russian Revolution (1918): Command and control economy, centralized allocation of capital and consumption and production decisions, with the goal of achieving greater welfare



Adam Smith





Karl Marx

The Economic Calculation Problem (1920s and '30s)

- von Mises, Hayek: Allocating resources requires lots of information about endowments, preferences, production sets, etc.
- How can central planners collect and process that information? Markets seem to only require limited communication/decentralized processing
- Lange, Lerner: Advocated a hybrid system with markets for production/consumption decisions, centralized allocation of capital

von Mises

Hayek

Lange

Early Modern Era (1950s and '60s)

- Samuelson, Arrow, Debreu, Lange: Neoclassical general equilibrium and welfare economics; modern welfare theorems for the production and exchange of private goods
- At around the same time, Arrow developed social choice, which considered the design of efficacious institutions, while abstracting from the details of the market economy
- These theories take as a premise that the outcome, e.g., prices/equilibrium/social choice depends directly on preferences of members of society
- In both lines of research, it was quickly discovered that there are limits to what society can achieve:
  - Externalities lead to market inefficiencies
  - IIA and Pareto efficiency imply dictatorship

Samuelson

Arrow

Debreu

Samuelson on selfish manipulation of systems

I must emphasize this: taxing... can not at all solve the computational problem in the decentralized manner [for public goods]... Of course, utopian voting and signalling schemes can be imagined... But there is still this fundamental technical difference going to the heart of the whole problem of social economy: by departing from his indoctrinated rules, any one person can hope to snatch some selfish benefit in a way not possible under the self-policing competitive pricing of private good.

— Samuelson (1954), "The Pure Theory of Public Expenditure"

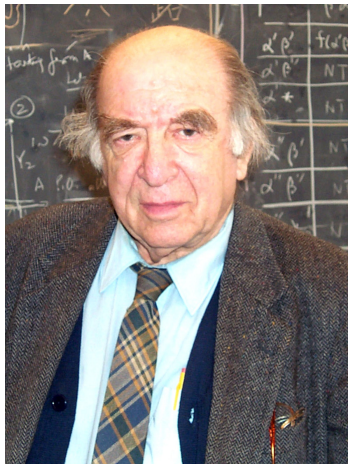
Vickrey's hot take

Once we decide to relax the independence postulate, there are many ways in which a social welfare function can be constructed...
All such schemes will, of course, be open to the objection that the social choice may be affected by changes in the range of alternatives presented to the individuals...
There is another objection to such welfare functions, however, which is that they are vulnerable to strategy. By this is meant that individuals may be able to gain by reporting a preference differing from that which they actually hold.

— Vickrey (1960), "Utility, Strategy, and Social Decision Rules"

Towards a Theory of MD

- Hurwicz (1917–2008): An economic system is a **mechanism** for incorporating decentralized information into collective decision making
- He introduced formal models of how decentralized information could be utilized by the mechanism
- Key ideas:
  - **Information structure**: A formal description of decentralized information, e.g., each agent knows their own preferences
  - **Informational efficiency**: A mechanism should collect only as much information as it needs
  - **Incentive compatibility**: It must be in each agent's self interest to communicate their information to the mechanism
  - **Participation constraints**: Agents may be able to achieve something on their own, which must be respected by the mechanism



Hurwicz

Descriptive vs Prescriptive Theories

- Hurwicz viewed MD as a tool for investigating questions both normative and positive
  - Descriptive/Positive: Use MD to understand what institutions have arisen in the world (presuming that they are the optimal solution to a design problem)
  - Normative/Prescriptive: Use MD to derive advantageous designs for mechanisms, which may or may not have been used before

Early applications

- Hurwicz was very interested in the implementation of competitive equilibrium, as a continuation of the socialist calculation debate
- Many of his papers model the mechanism as an iterative process that, at each step, elicits some information from agents; the dynamic process requires relatively little information at each step, similar to the market
- Hurwicz (1972): Impossible to implement competitive equilibrium in a two-agent Edgeworth economy; agents would have an incentive to behave as if they have different preferences in order to manipulate prices
- He implicitly assumes:
  - Complete information among the agents
    (i.e., they know one another's preferences, the planner does not)
  - Participation constraint: Each agent has the option to consume their own endowment
- These assumptions seem natural, relative to the baseline of complete information competitive equilibrium

Divergent Threads in MD

- Research on MD exploded in the 1970s and '80s, proposing various ways to generalize and operationalize the theory, and different approaches along the following dimensions:
  - The designer's objective
  - The form of decentralized information and preferences
  - The requirements for incentive compatibility
- I will organize this discussion according to the following classifications:
  - Dominant strategy implementation
  - Complete information/full implementation
  - Bayesian implementation

(Disclaimer: This ignores huge swaths of the MD literature)

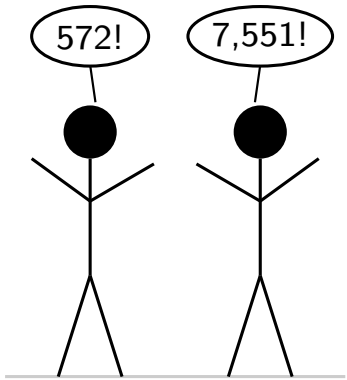
Dominant Strategy Implementation

- Gibbard (1973) and Satterthwaite (1973, 1975) built on the Arrovian framework, responding to Vickrey's critique
- IC is operationalized as: no agent is better off ex post by "misreporting" their preferences, i.e., dominant strategy implementation
- Key finding: any dominant strategy and Pareto efficient mechanism is a **dictatorship**
- Dominant strategy plays an analogous role to IIA, where the social choice is the top ranked alternative
- This presumes that having agents simply "report" their prefs to the mechanism is WLOG, later developed as a **revelation principle** by Dasgupta, Hammond, and Maskin (1979)

Augustus of Prima Porta

Full Information MD

- Another approach due to Maskin, embracing the classical idea that agents know one another's preferences:
  - The desired social outcomes and agents' preferences depend a **state of the world** θ (which could be the preference profile)
  - Agents have **common knowledge** of θ (so information is not really "decentralized" but it is unknown to the mechanism)
  - IC is operationalized as **full implementation**: The set of desired outcomes coincides with outcomes that arise in Nash equilibrium
- Revelation mechanisms no longer suffice; "monotonicity" is a necessary (but not sufficient) condition for full implementation
- Various sufficient conditions, but the positive constructive results involve "integer games," so that the mechanism restricted to subsets of the action space in which the agents disagree do not have equilibria



Adding more structure to preferences

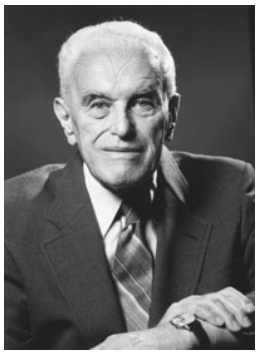
- These lines of research assumed minimal structure on preferences
- Possible to do more if agents are expected utility maximizers
- Vickrey-Clarke-Groves: Can implement efficient allocations in dominant strategies (though they do not use that language)
- Abreu and Matsushima (1992): under complete information, can *virtually* implement any outcome that is a NE of the revelation game, using a sequence of finite mechanisms that have unique rationalizable actions

Modeling information

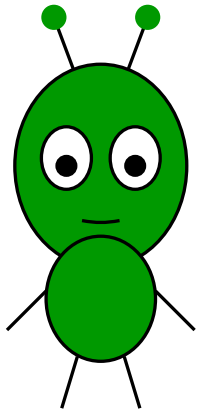
- In the Arrovian tradition, it must have seemed self-evident to model the decentralized information as the preference profile (as in DSI)
- This is implicit in the Economic Calculation Debate: that economic agents know their preferences and resources and can communicate them to the system, and it was also presumed in other work, e.g., Vickrey's (1961) positive theory of first- and second-price auctions
- But this is a rather limited view of the possible forms that private information could take
- John Harsanyi's (1967) articles opened up the vast world of **type spaces** for flexible modeling of information about preferences and others' information
- Myerson (who was a student of Arrow's at the same time as Maskin) incorporated Harsanyi's model of distributed information into MD as the "information structure"

Harsanyi's model of information

- Model decentralized info as an **information structure**:
  - Agents $i = 1, \ldots, N$
  - Payoff relevant **state of the world** $\theta \in \Theta$
  - Agent i has **signals** $s_i \in S_i$
  - Interim beliefs $\pi_i(\theta, s_{-i}|s_i)$
- Often (though not always) model a **common prior** $\pi(\theta, s)$ from which interim beliefs are derived



John Harsanyi



A Martian

Mechanisms and equilibrium

- The mechanism can control an **outcome** $\omega \in \Omega$
- A mechanism consists of:
  - Actions $a_i \in A_i$
  - Outcome function $m(\omega|a)$ (probability of ω given action profile a)
- Unless stated otherwise, take S_i and A_i to be finite
- Preferences: $u_i(\omega, \theta)$ (agent i) and $w(\theta, \omega)$ (designer)
- All together, $(N, \Theta, \Omega, S, \pi, A, m)$ describe a **Bayesian game**
- Agent i's behavioral strategy $b_i(a_i|s_i)$
- Solution concept: Bayes Nash equilibrium

Bayesian mechanism design

- Harsanyi takes the game—parametrized by (Ω, A, m)—as given
- Myerson takes the information structure as fixed, but treats the mechanism as something to be optimized
- His approach is distinguish by:
  - Designer's objective: Maximize expectation of w, subject to IC
  - IC is operationalized as BNE
  - Designer can *pick the equilibrium*
    (aka **partial implementation**, in contrast to full implementation)
  - Drawing inspiration from bargaining and cooperative games, he reformulates participation constraints as **interim individual rationality**:
    Conditional on s_i, agent i's expected payoff must be at least the expectation of a lower bound $\underline{u}_i(\theta)$

Myerson's findings

- General results in the 1979 paper on the bargaining problem:
  - **Revelation principle**: WLOG to use IC direct mechanisms, for which $A_i = S_i$ and $b_i(a_i|s_i) = 1$
  - Reduction of the designer's problem to a **linear program**
- Specific applications:
  - 1981: Optimal auctions
    - Independent private values: virtual values, revenue equivalence, optimality of first/second-price (w/ symmetry, regularity)
    - Correlated private values: "very strange auctions" become optimal
  - 1982 (w/ Baron): Regulating a monopolist
  - 1983 (w/ Satterthwaite): Bilateral trade, impossibility of efficient trade, second-best mechanism, optimality of double auction (in uniform/uniform special case)

Challenges

- All three approaches achieved much but are dogged by persistent challenges
- DSI: relies on private values, or with ex post implementation, need to know what incentives are "ex post" relative to; also it may be that not many SCFs can be implemented, more on that later
- FI: Dependent on complete information, uses exotic "integer games"
- BI: The mechanism is highly dependent on the type space... which is the right one? For many type spaces, e.g., CPV, optimal mechanisms seem overfitted and implausible
  - Full surplus extraction paradox: Correlation is generic, FSE should be everywhere, but we don't see it anywhere
  - Even with ex post IR/IC, side bets are a robust theoretical prediction
- IMHO, all three approaches haven't yet delivered on the grand promise of MD: To suggest new and more efficient designs of systems for resource allocation, bargaining, governance, etc
- Suggests that further innovations in modeling are needed

Hurwicz on Modeling the Designer's Preferences

Now the normative problem involves selecting a mechanism with only partial knowledge of the environment in which it will have to operate. Such knowledge might, for instance, be expressed (in the Bayesian spirit) as a probability measure on the space of conceivable environments. For the sake of simplicity, however, we shall suppose at this point that the "organizer" only has an a priori admissible class of environments E_0, without any probabilistic information within this class... If a probability measure on E_0 were available to the organizer, he might maximize with respect to [the mechanism] the corresponding expected value... In the absence of such 'prior' probabilities, the organizer might adopt some principle of decision making under uncertainty, say maximin, over E_0.

— Hurwicz (1972), "On Informationally Decentralized Systems"

How to model the designer's preferences?

- Hurwicz explicitly suggested modeling the designer as having Bayesian or maxmin preferences
- Bayesian was quickly adopted in the '70s and '80s, but maxmin seems have been largely ignored until the 2000's
- The argument for maxmin is, implicitly, that the designer does not "know" a prior with which to take an expectation with respect to
- Overfitting the mechanism to a particular information structure exposes the mechanism to underperformance in the event of that the information structure is misspecified
- This is not a super satisfying justification, and indeed, I think the profession is still struggling to justify maxmin as a design criterion, even though it seems to have been a productive assumption
- In a few lectures, Lars will discuss in more detail axiomatic foundations; not "knowing" a prior is a failure of Savage's axioms and in particular the sure-thing principle

Why maxmin?

- My own view: Information structure, common prior, and equilibrium are not meant to be taken literally; they are loose metaphors for forces towards common understanding of the world and optimization of individual welfare
- Information structures/equilibria are a disciplined way of thinking through strategic scenarios that will not unravel due to individual incentives or common knowledge
- But which is the right scenario to focus on? There are so many, and they generate strikingly different conclusions about optimal designs
- Moreover, if we try to build the mechanism to be optimal for all environments, then the informational load will be debilitating
- We are looking for environments for which the optimal mechanisms
  - Elicit relatively little information (informational efficiency)
  - Captures general principles about incentives and private information that are broadly applicable
- The "min" in maxmin systematizes the search for such environments

What kind of robustness?

- A key question in maxmin MD is what is the min over:
  1. Information (the Harsanyi types (S, π))
  2. Fundamentals (the state θ parametrizing preferences)
  3. Equilibrium
- Strands of the literature explore all of these, in particular
  - Large literatures in CS and econ study robustness to the distribution of preferences in (2) but not (1) or (3) (often in private value models)
  - Also there is a large literature on collusion/coalitional deviations which is more focused on (3) and less so (1) and (2)
- I am primarily interested in (1) and (3) but will also consider (2)
- I will review some antecedents in maxmin mechanism design and then describe my own work with coauthors

Wilson's critique

Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one player's probability assessment about another's preferences or information.
I foresee the progress of game theory as depending on successive reductions in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality.

— Wilson (1987), "Game-Theoretic Analysis of Trading Processes"

Robust = dominant strategies?

- So maybe the issue is that we should not be assuming common knowledge of the information structure between designer and agents
- Many interpret this as saying we should use a stronger implementation concept, e.g., dominant strategy implementation, so that the incentive compatibility holds independent of beliefs
- The connection between a concern about misspecification of the environment and dominant strategy implementation (or weaker cousin ex post implementation) is the subject of Bergemann and Morris (2005) and Chung and Ely (2007)

Ex post implementation

- An information structure has **know payoff types** (KPT) if:
  - $\theta = (\theta_1, \ldots, \theta_N)$
  - Each agent's signal is of the form $s_i = (\theta_i, t_i)$
  - θ_i is agent i's **payoff type**
  - t_i is their **belief type**
- Given a mechanism, an **ex post equilibrium** is one for which each agent has an optimal strategy that only depends on θ_i, and is IC conditional on θ
- (NB Without KPT, we would have to ask: ex post w.r.t. what?)
- To be contrasted with Bayes Nash equilibrium, where optimal strategies generally depend on beliefs, and optimality is interim

Bergemann and Morris (2005)

- Designer wants to (weakly) implement a social choice correspondence $F : \Theta \to 2^{\Omega}$
- If an SCF can be implemented ex post, then it can be implemented in Bayes Nash equilibrium for all KPT information structures
- Is the converse true? In general no; BM construct examples where F is implementable in all KPT information structures but not ex post implementable
- They show that they are equivalent under the hypotheses they call "separability":
  - $\Omega = \Omega_0 \times \Omega_1 \times \cdots \times \Omega_N$
    (Common and individual components)
  - $F(\theta) = \{f_0(\theta)\} \times F_1(\theta) \times \cdots \times F_N(\theta_N)$ (No flexibility on common, no joint restrictions on individual)
  - $u_i(\omega, \theta) = v_i(\omega_0, \omega_i, \theta)$
    (Agents care about common and own components)
- The quasilinear auction setup, with allocation as ω_0 and transfer as ω_i, and F being the efficient allocation, is one example that fits separability

Chung and Ely (2007)

- Auctions with correlated private values
- Myerson (1981) (and Crémer and McLean and others) assume a common prior
- In contrast, CE allow all (possibly non-common prior) information structures, and take the correlated prior to be the designer's belief
- One option for the designer is to use the optimal dominant strategy mechanism...
- Are there mechanisms such that for any specification of the agents' information, there is a Bayes Nash equilibrium with strictly higher revenue? In other words, is there another mechanism with a higher **guarantee** for revenue than the optimal DS mechanism?

CE's Answer

- Sometimes yes and sometimes no
- In the Myerson regular case, exactly one IC/IR binds for each type (local downward, and IR at the bottom)
- If the solution to the dominant strategy mechanism also has that feature, then the dominant strategy mechanism has the highest revenue guarantee
- In fact, there is a particular (non-common prior) information structure in which optimal Bayesian revenue is exactly the optimal dominant strategy revenue, wherein beliefs are proportional to the optimal Lagrange multipliers for the DSIC/IR program
- However, CE give examples where this condition fails and the guarantee is higher than the optimal DS revenue, either because more constraints bind for DSIC/IR, or if we impose the common prior

Other issues with DS/EP implementation

- BM and CE provide a qualified maxmin foundation for DS/EP implementation, but it requires exceptional structure (KPT, separability, minimal binding IC/IR)
- They also demonstrate that in general, a designer with maxmin preferences could achieve more with Bayesian mechanisms, for which outcomes depend on beliefs
- This suggests that perhaps robustness and relaxation of common knowledge assumptions does not mean DS/EP
- Another issue: Jehiel, Meyer-ter-Vehn, Moldovanu, and Zame (2006) argue that for a class of smooth two-dimensional implementation problems, EP implementable SCFs must be constant
- Suggests that we should reengage with Bayesian mechanisms/guarantees

Guarantees with Bayesian mechanisms

- This is where my research enters the story: Bergemann, Brooks, and Morris (2015, 2016, 2017, 2019); Du (2018); Brooks and Du (2021, 2023, 2024, 2025, 2026); Brooks, Du, and Feffer (2026); Brooks, Du, and Zhang (2026)
- These papers solve a maxmin problem where
  - We drop the KPT structure as a hypothesis
  - The min is over **common prior** information structures
  - And in contrast to BM and CE, who allow the designer to select the equilibrium, we will consider best-case and worst-case equilibrium selection

BD's guarantee maximization program

- Fix a prior $\mu \in \Delta(\Theta)$
  (or more generally, a set of priors... this keeps us away from uninteresting cases that the designer can rule out, e.g., everyone thinks the goods being allocated are worthless)
- The guarantee $G(A, m)$ is the minimum expectation of $w(\theta, \omega)$ over all (common-prior) information structures with marginal μ and over all Bayes Nash equilibria
- What is the supremum guarantee $G(A, m)$ over all (finite) mechanisms (A, m), and what are the mechanisms that attain it?
- Call the supremum G^*

Modeling participation

- Of course, we may want to add participation constraints on the agents, as with interim IR
- But whether or not interim IR is satisfied depends on the information structure/equilibrium
- (A, m) is **strongly individually rational** (SIR) if interim IR is satisfied in every information structure and equilibrium
- For example, if not bidding in an auction gives zero utils, then first- and second-price auctions are SIR, because there is always the option to bid zero

Participation Security

- Actually, the standard auctions satisfy a stronger property: bidding zero is a **secure** action, meaning, guarantees a payoff higher than the outside option, for all (a_{-i}, θ)
- (A, m) is **participation secure** (PS) if every agent has a secure action
- PS $\implies$ SIR but not the other way around
- Brooks and Du (2026): For every (A, m) that is SIR, there is an (A', m') that is PS and $G(A', m') \geq G(A, m)$
- So it is WLOG to optimize $G(A, m)$ over mechanisms that are PS

Dual reduction (adapted from Myerson 1997)

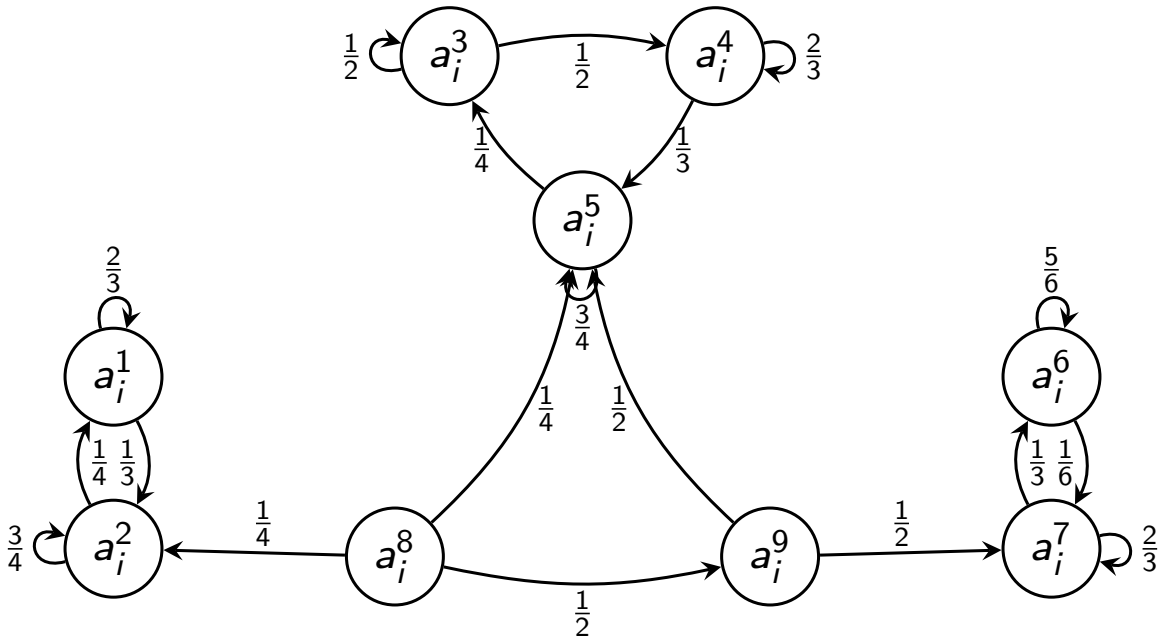
- Let $\Sigma(A)$ be distributions on $A \times \Theta$ with marginal μ
- Consider the program

$$\max_{\sigma \in \Sigma(A), \{\nu_i : A_i \to \mathbb{R}\}} \sum_{i, a_i} \nu_i(a_i)$$

$$\text{s.t.} \sum_{a_{-i}, \theta, \omega} u_i(\omega, \theta) \left[m(\omega | a_i, a_{-i}) - m(\omega | a_i', a_{-i}) \right] \sigma(a_i, a_{-i}, \theta) \geq \nu_i(a_i) \ \forall i, a_i, a_i'.$$

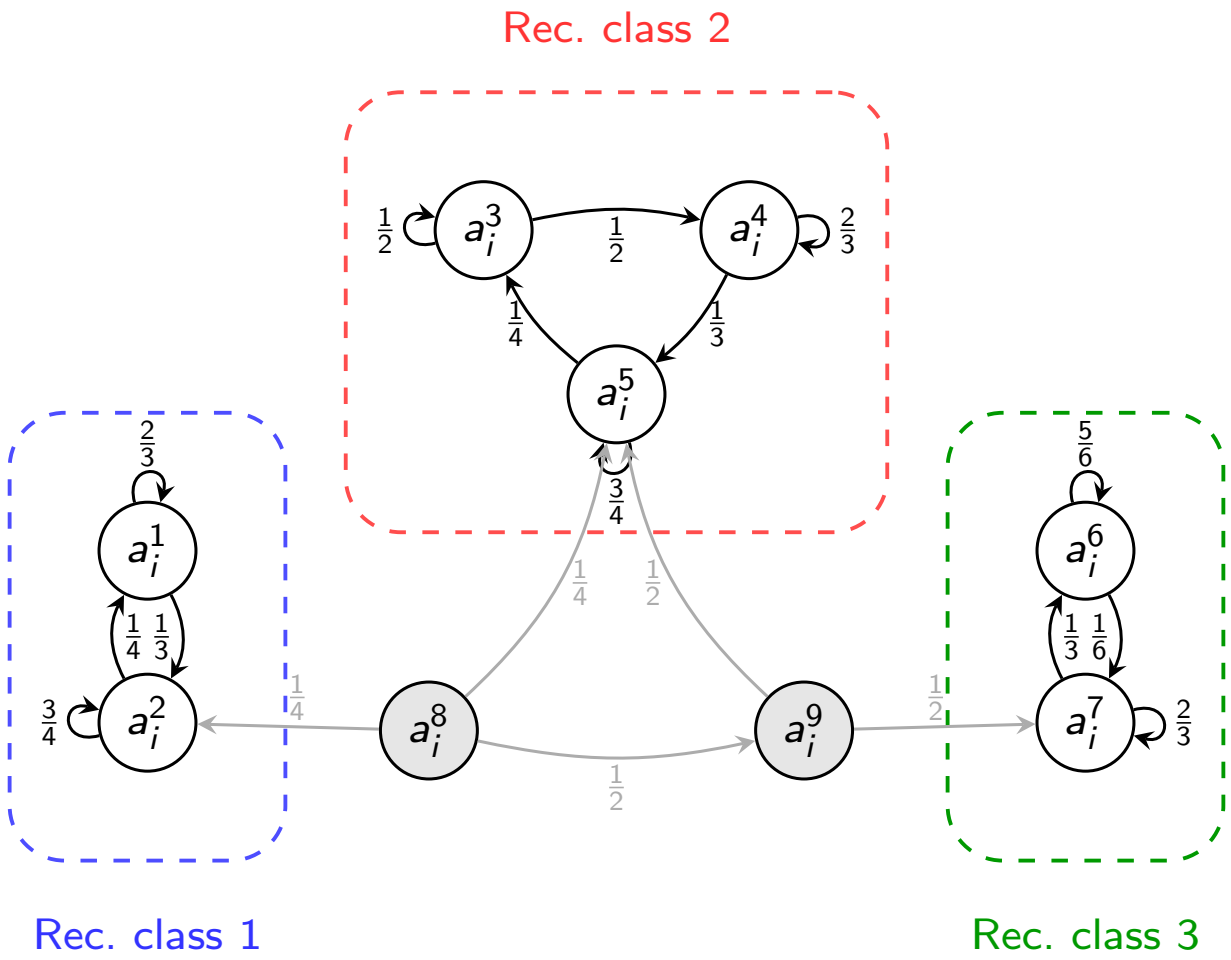
- This LP has a saddle point; in fact the optimal solution is zero, and is attained by any **Bayes correlated equilibrium**: A direct recommendation information structure, which is feasible with $\nu_i(a_i) = 0$ for all i and a_i
- Let $\alpha_i(a_i' | a_i) \geq 0$ be optimal Lagrange multipliers on the obedience constraints
- Dual feasibility requires that for all i, a_i, $\sum_{a_i'} \alpha_i(a_i' | a_i) = 1$

The action network



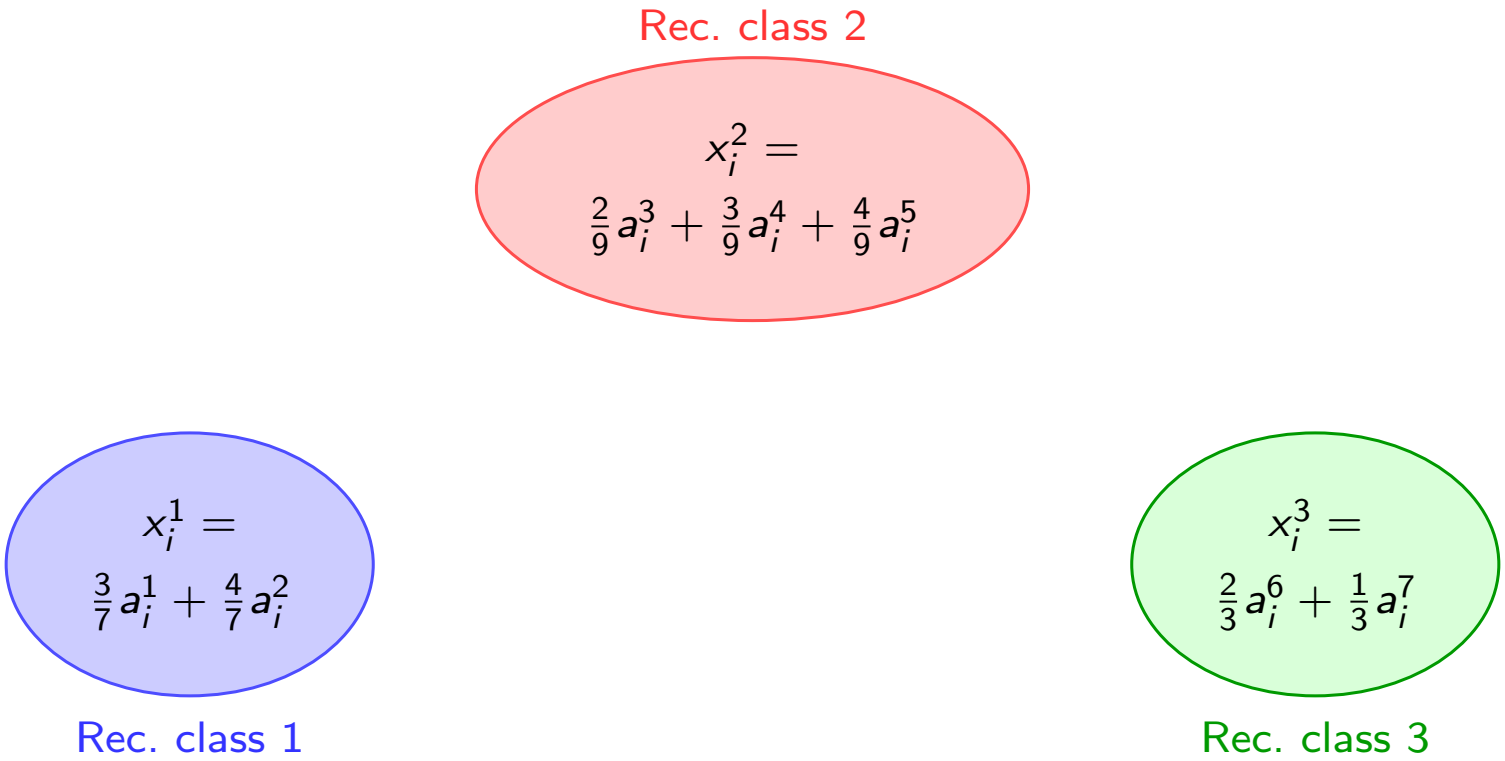
Edge weights are the transition probabilities $\alpha_i(a_i'|a_i)$

Recurrent classes



Edge weights are the transition probabilities $\alpha_i(a_i'|a_i)$

Dual reduction: merging recurrent classes



Each recurrent class is merged into a single mixed action, equal to its invariant measure

Properties and Implications of Dual Reduction

- This dual reduction has a remarkable property:
  If b is a BNE of the dual reduction (A', m'), then it is also a BNE of the original mechanism, wherein agents **voluntarily** randomize as specified in the reduced actions
- This means that dual reduction reduces the set of BNE outcomes, information structure by information structure
- Hence, if the original mechanism is SIR, then so is its reduction
- By iteratively reducing, we arrive at an SIR mechanism that is **elementary**: $\alpha_i(a_i'|a_i) = \mathbb{I}_{a_i'=a_i}$, so there is a BCE with all obedience constraints strict
- $\implies$ any interim belief can be realized in equilibrium, so that no matter the interim belief about (θ, a_{-i}), interim IR is satisfied
- Minimax $\implies$ each agent has a secure (possibly mixed) action

More fun with dual reductions

- Now, suppose $a_i^0 \in A_i$ is secure
- Then $G(A, m)$ is the solution to

$$\min_{\sigma \in \Sigma(A)} \sum_{a,\theta,\omega} w(\omega, \theta) m(\omega|a) \sigma(a, \theta)$$

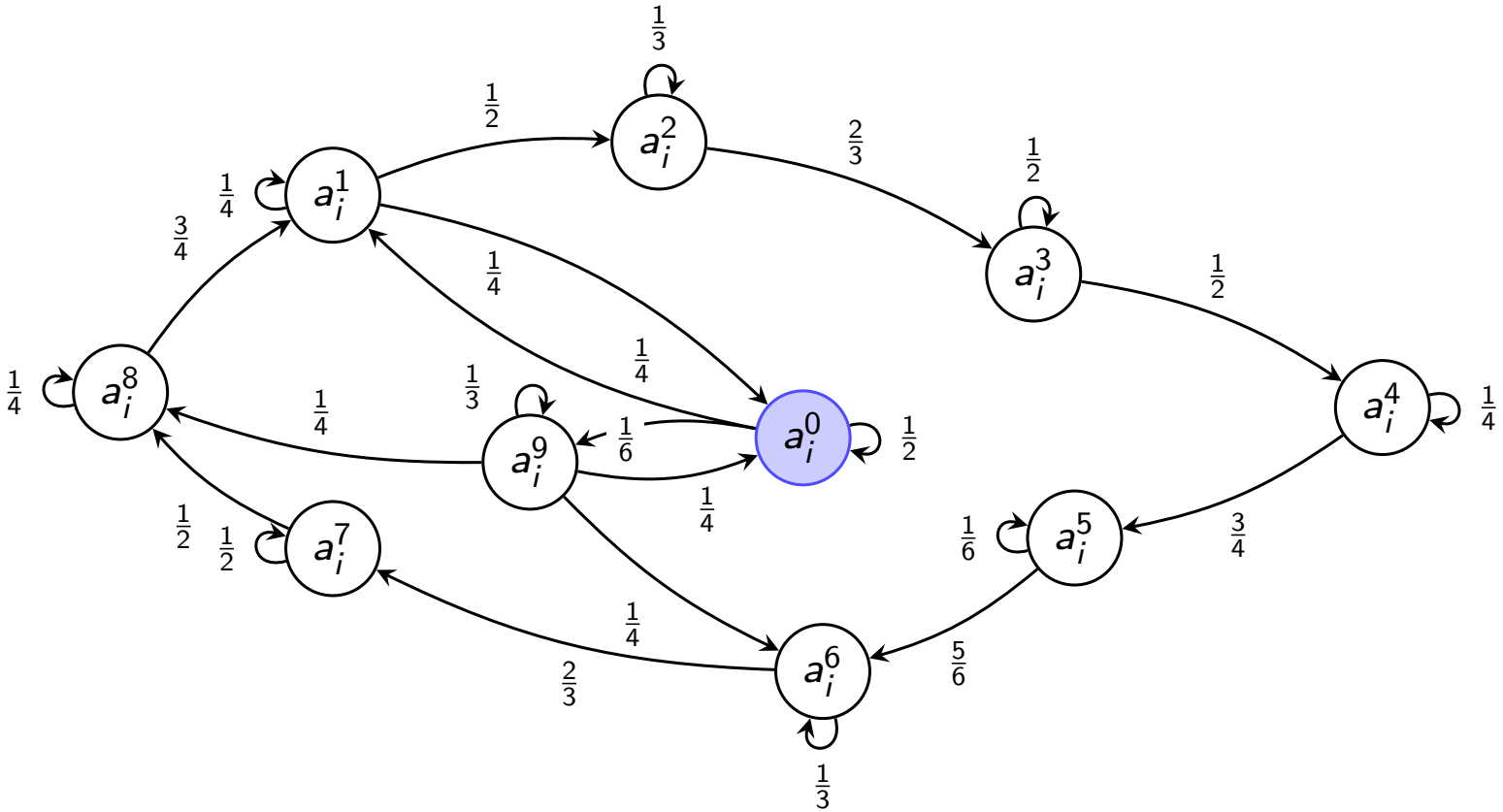
$$\text{s.t.} \sum_{a_{-i},\theta,\omega} u_i(\omega, \theta) \left[m(\omega|a_i, a_{-i}) - m(\omega|a_i', a_{-i}) \right] \sigma(a_i, a_{-i}, \theta) \geq 0 \ \forall i, a_i, a_i'$$

- Again this program has a saddle point, and we let $\alpha_i(a_i'|a_i)$ be the optimal Lagrange multipliers
- NB $\alpha_i(a_i|a_i)$ is free, so we can always choose it to be strictly positive so that for some $C > 0$,

$$\sum_{a_i'} \alpha_i(a_i'|a_i) = C \ \forall i, a_i$$

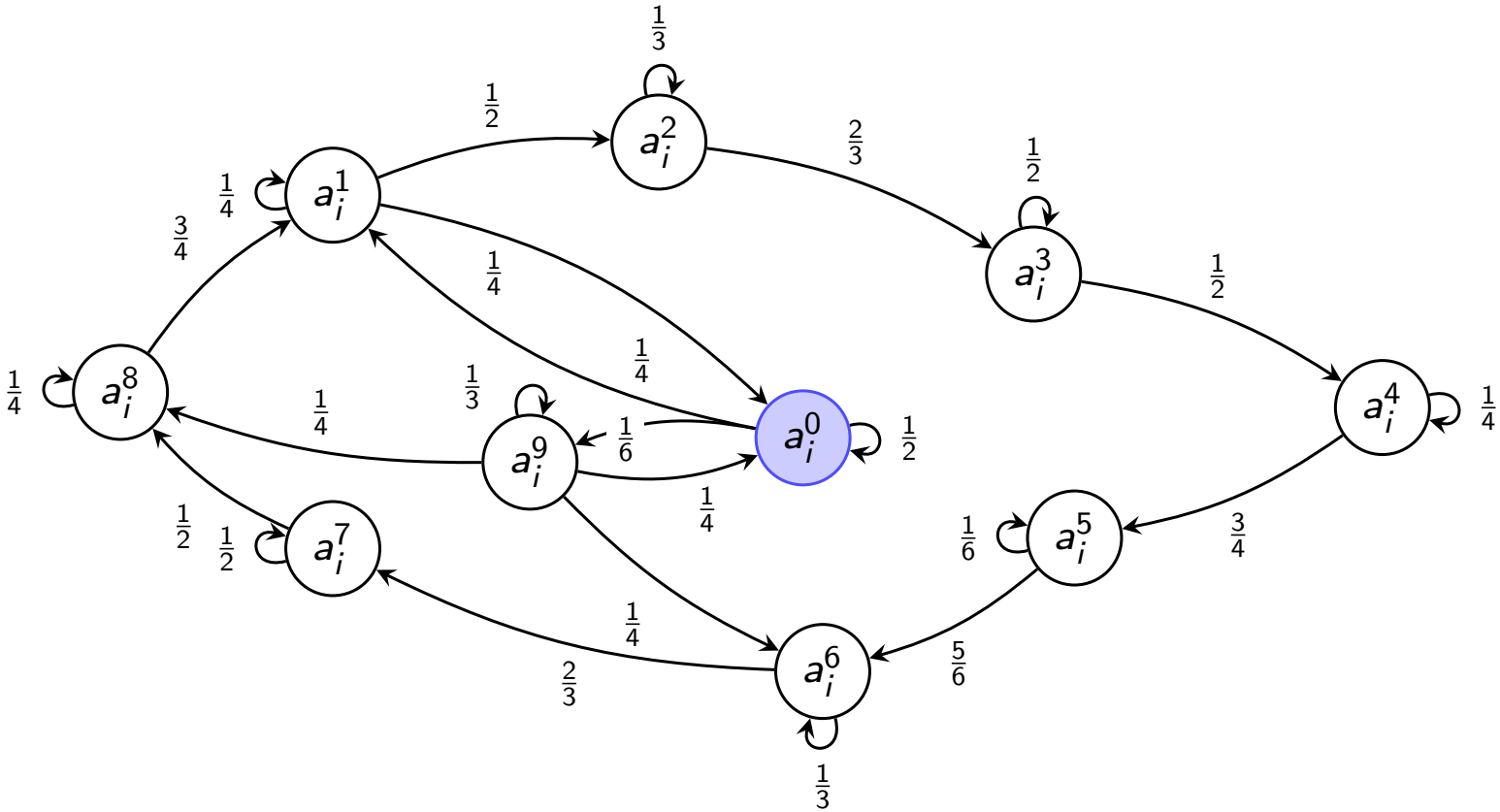
- α_i / C is a "most tempting" deviation

The "most tempting" deviation



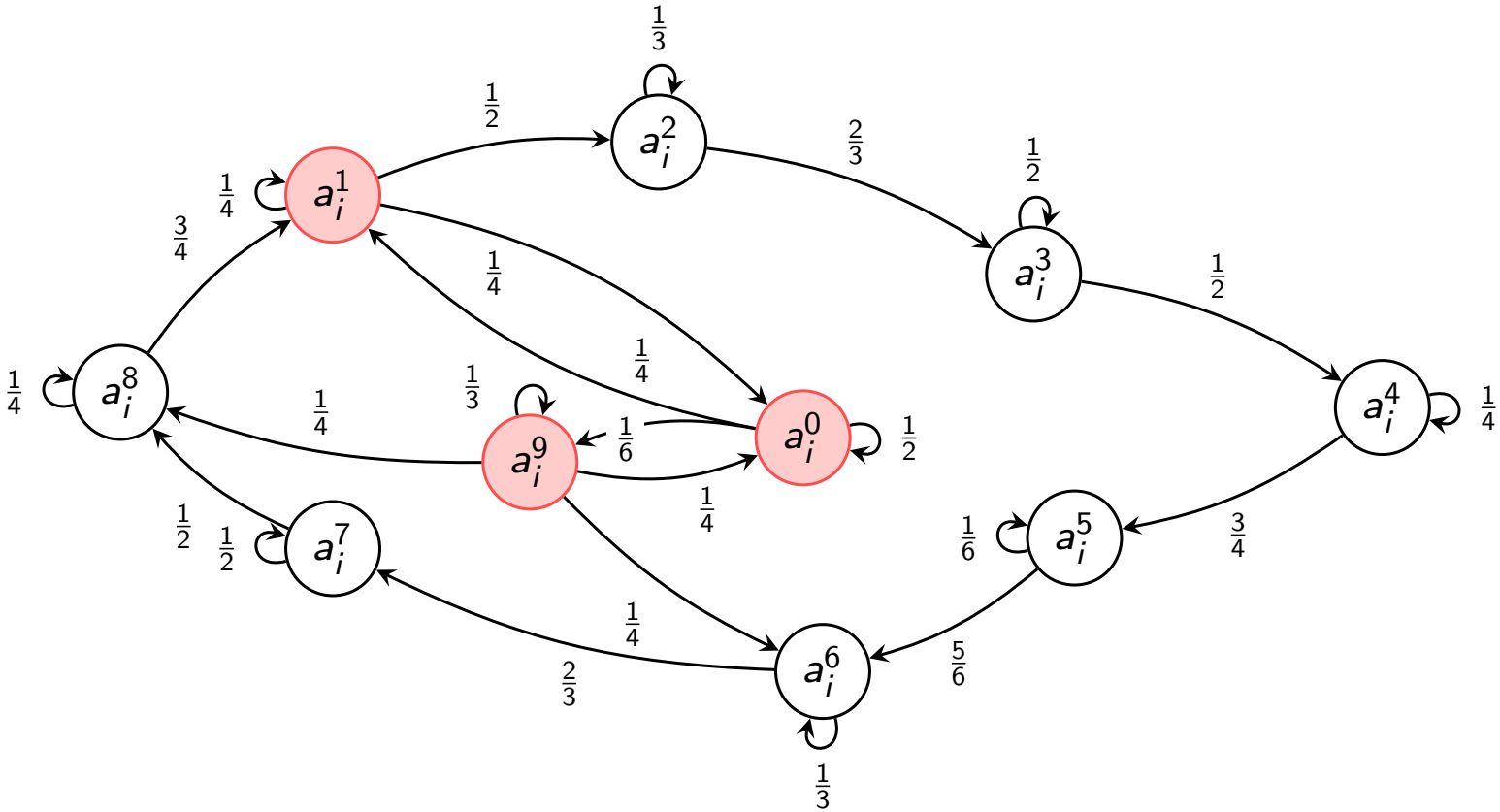
Edge weights are the transition probabilities $\alpha_i(a_i'|a_i)/C$

Iterated deviation: round 0



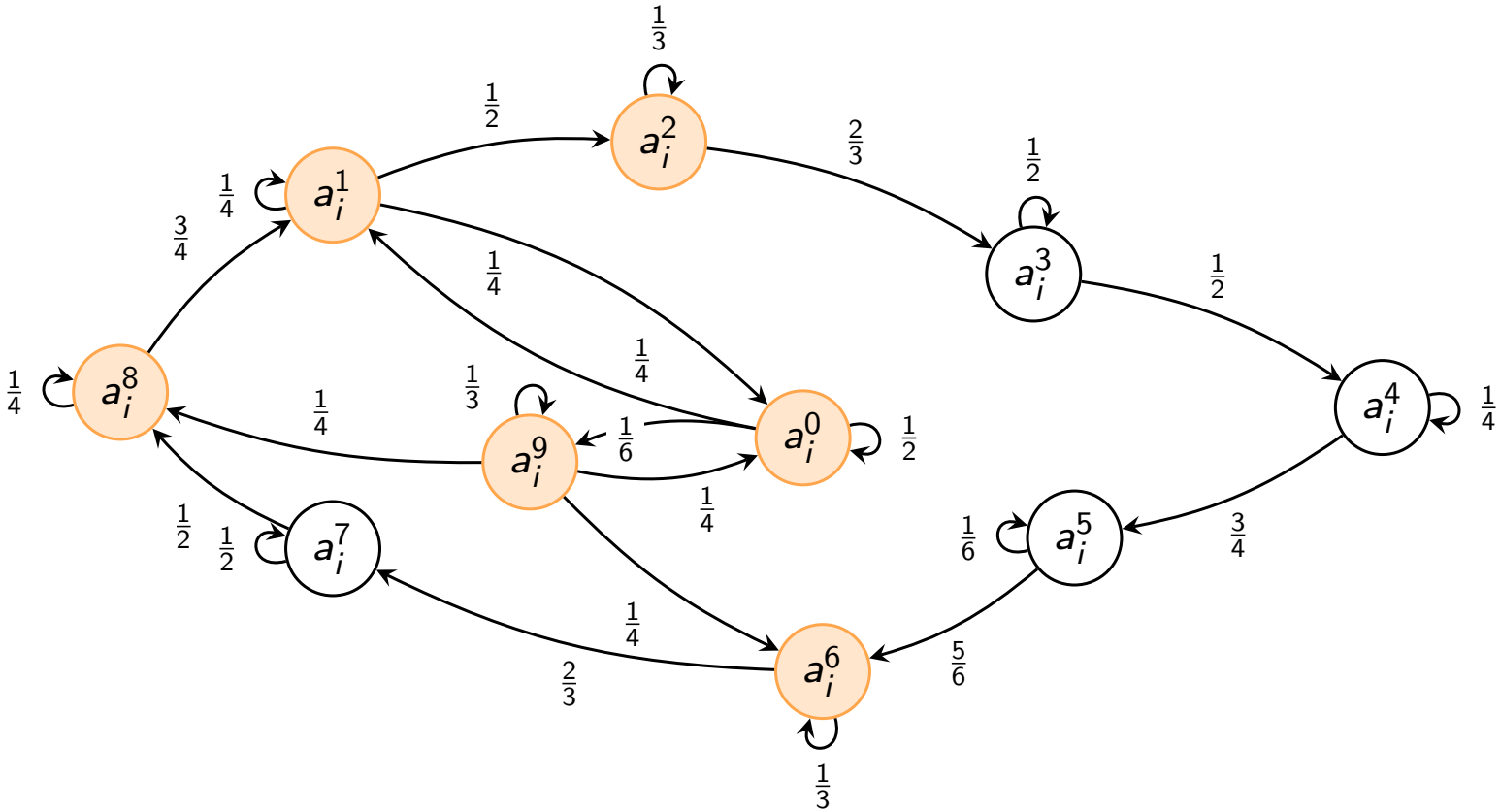
$$x_i^0 = a_i^0$$

Iterated deviation: round 1



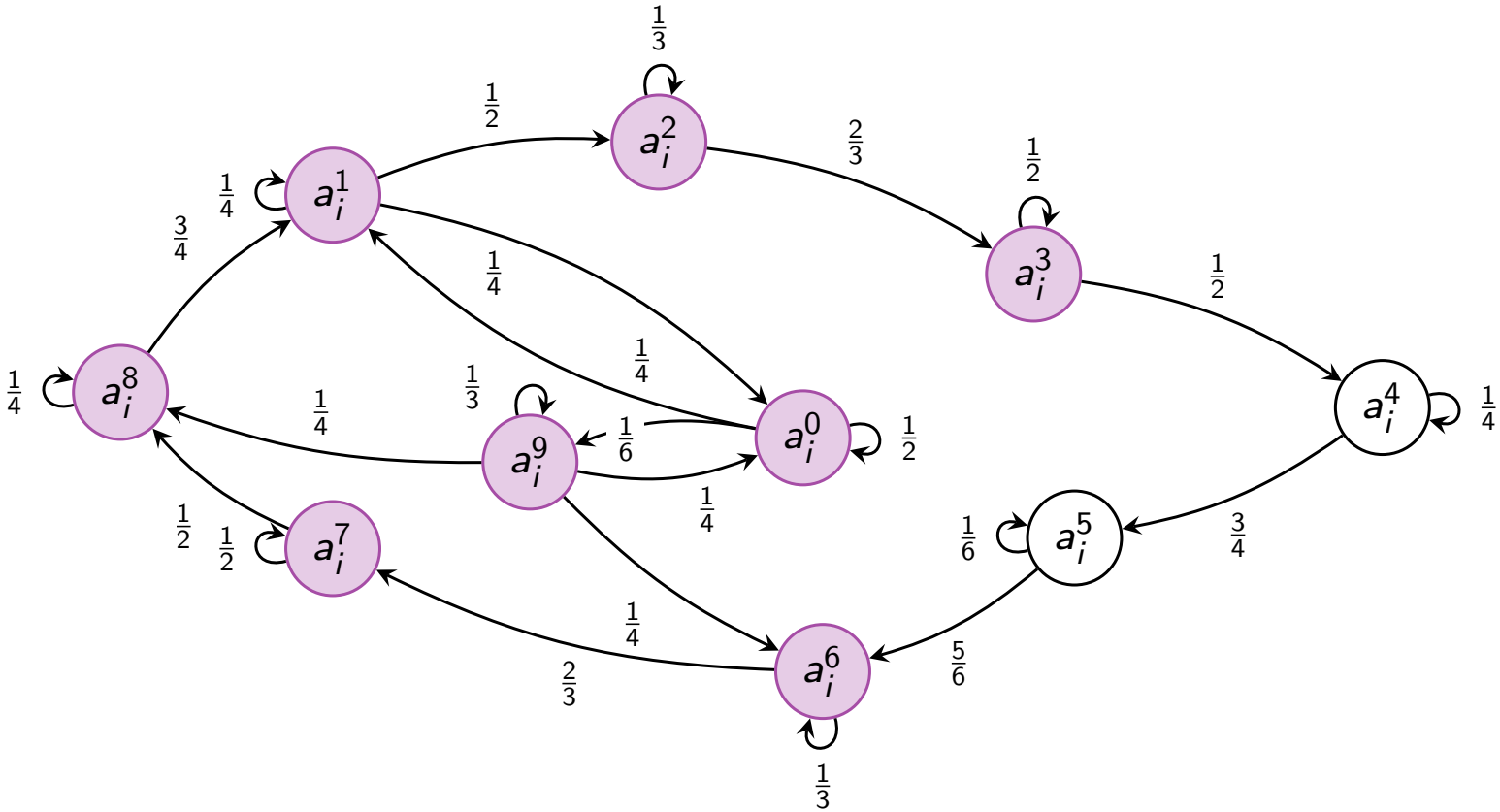
$$x_i^1 = \tfrac{1}{2}a_i^0 + \tfrac{1}{4}a_i^1 + \tfrac{1}{4}a_i^9$$

Iterated deviation: round 2



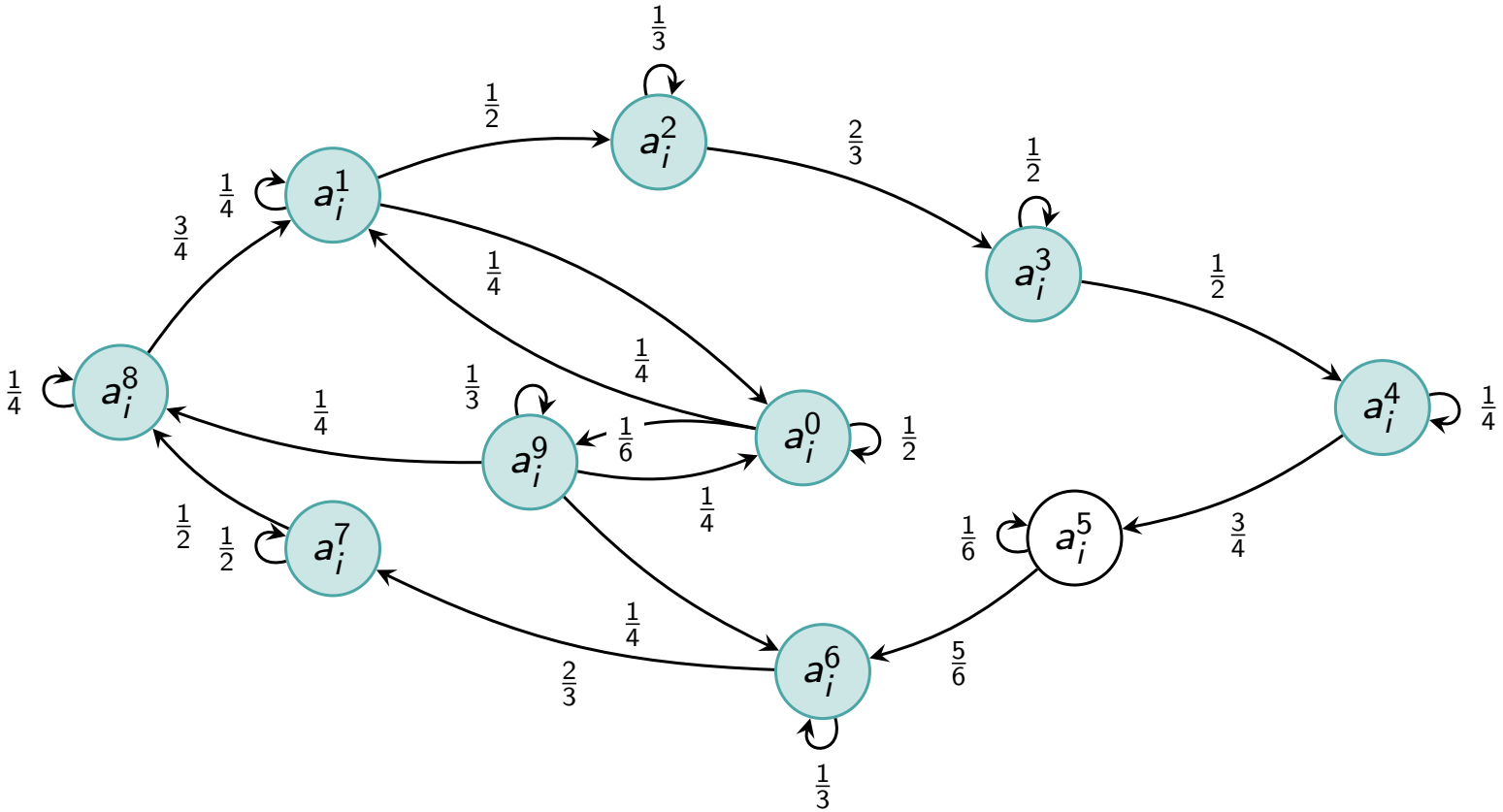
$$x_i^2 = \frac{17}{48}a_i^0 + \frac{3}{16}a_i^1 + \frac{1}{8}a_i^2 + \frac{1}{16}a_i^6 + \frac{1}{16}a_i^8 + \frac{5}{24}a_i^9$$

Iterated deviation: round 3



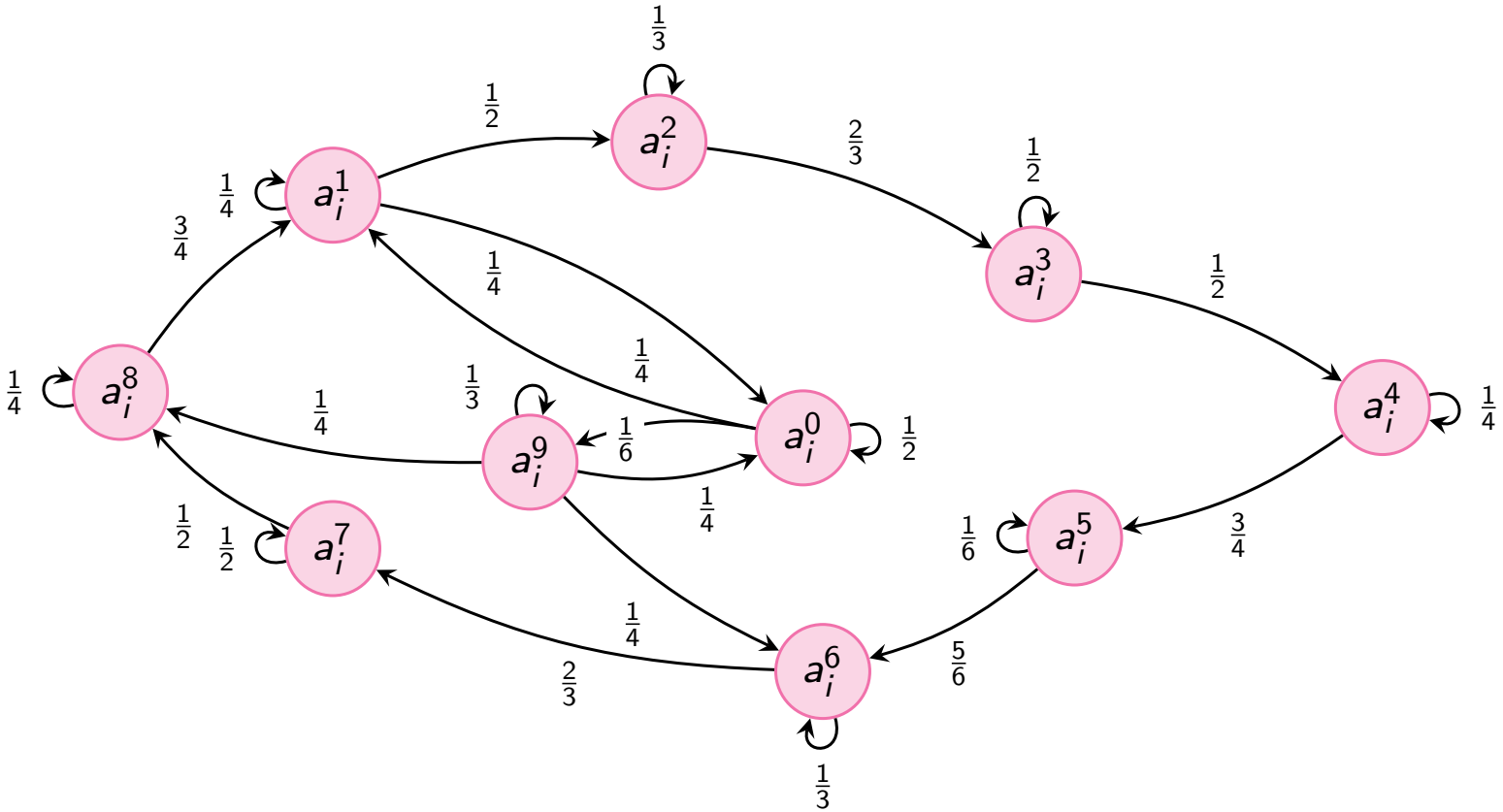
$$x_i^3 = \tfrac{149}{576}a_i^0 + \tfrac{35}{192}a_i^1 + \tfrac{13}{96}a_i^2 + \tfrac{1}{12}a_i^3 + \tfrac{7}{96}a_i^6 + \tfrac{1}{24}a_i^7 + \tfrac{13}{192}a_i^8 + \tfrac{91}{576}a_i^9$$

Iterated deviation: round 4



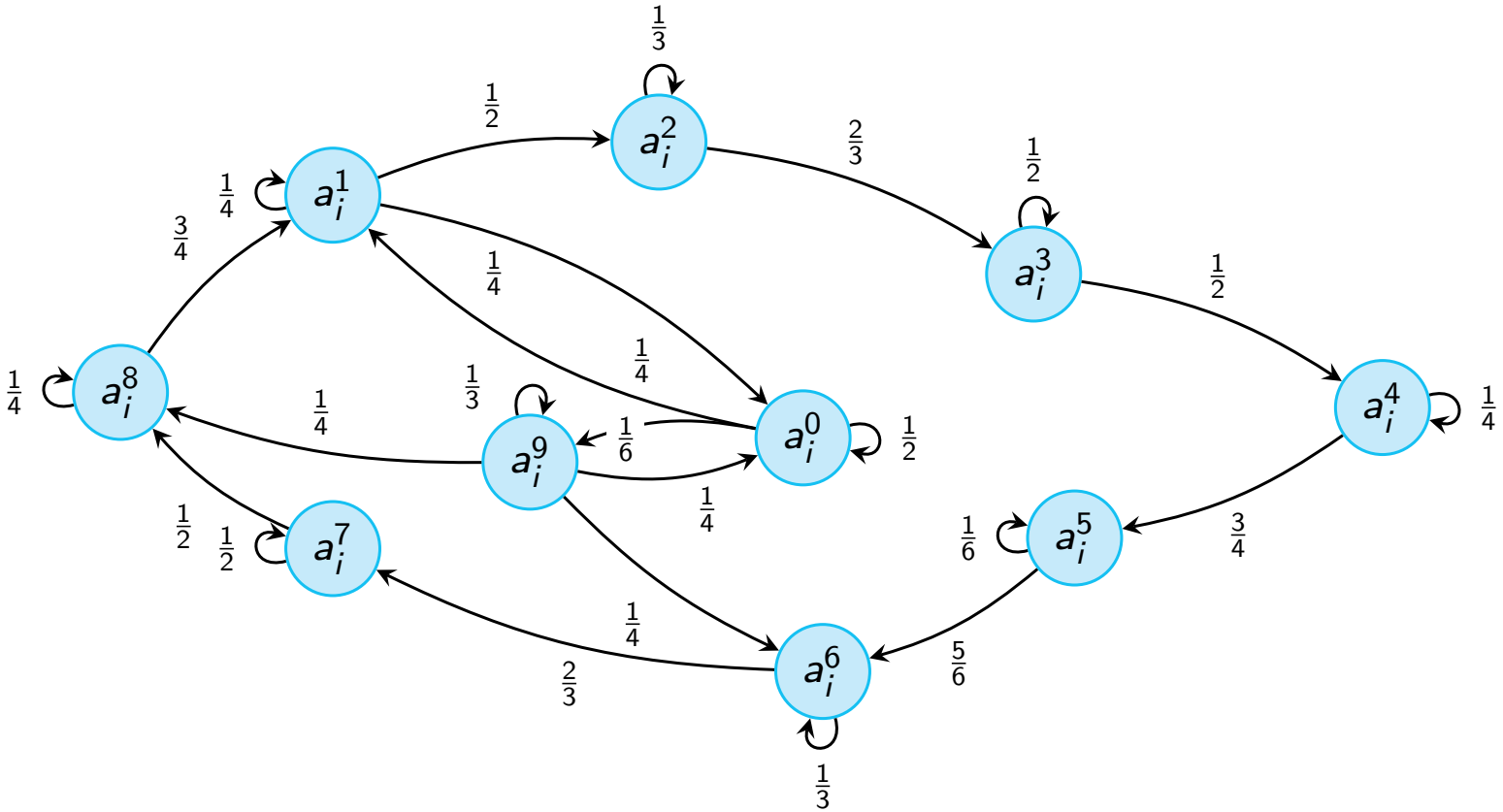
$$x_i^4 = \frac{1391}{6912}a_i^0 + \frac{371}{2304}a_i^1 + \frac{157}{1152}a_i^2 + \frac{19}{144}a_i^3 + \frac{1}{24}a_i^4 + \frac{49}{768}a_i^6 + \frac{5}{72}a_i^7 + \frac{89}{1152}a_i^8 + \frac{811}{6912}a_i^9$$

Iterated deviation: round 5



$$x_i^5 = \frac{13307}{82944}a_i^0 + \frac{2053}{13824}a_i^1 + \frac{1741}{13824}a_i^2 + \frac{271}{1728}a_i^3 + \frac{11}{144}a_i^4 + \frac{1}{32}a_i^5 + \frac{1399}{27648}a_i^6 + \frac{89}{1152}a_i^7 + \frac{2305}{27648}a_i^8 + \frac{7417}{82944}a_i^9$$

Iterated deviation: round ∞



$$x_i^\infty = \tfrac{960}{12683}a_i^0 + \tfrac{1680}{12683}a_i^1 + \tfrac{1260}{12683}a_i^2 + \tfrac{1680}{12683}a_i^3 + \tfrac{1120}{12683}a_i^4 + \tfrac{1008}{12683}a_i^5 + \tfrac{1395}{12683}a_i^6 + \tfrac{1860}{12683}a_i^7 + \tfrac{1360}{12683}a_i^8 + \tfrac{360}{12683}a_i^9$$

Properties and implications of this dual reduction

- Now define the dual reduction to be the original mechanism, but where each agent is restricted to the mixtures $\{x_i^0, x_i^1, \ldots\}$
- Brooks and Du (2025) show that this mechanism has a weakly higher guarantee than the original mechanism; roughly speaking, fewer outcomes are feasible, but the critical deviation that pins down the value of the program is still feasible
- Moreover, it is possible to verify this by using only the constraints associated with deviating from x_i^k to x_i^{k+1}, where the Lagrange multiplier is C!
- NB the reduction has countably infinitely many mixtures, but the sequence can be truncated at arbitrarily small loss to the guarantee
- TI;dr: If we started with a mechanism whose guarantee is close to G^*, then the reduction's guarantee is also close to G^*!

The bounding program

- Thus, we might as well take the actions to be ordered, and maximize the "local" lower bound
- Let $M(K)$ be the set of rules $m : \{0, \ldots, K\}^N \to \Delta(\Omega)$ for which 0 is secure for all i
- Then G^* is equal to the supremum over (C, K) of

$$\sup_{m \in M(K)} \sum_{\theta} \mu(\theta) \min_{a} \sum_{\omega} \left(w(\omega, \theta) m(\omega|a) + C \sum_{i} u_i(\omega, \theta)[m(\omega|a_i + 1, a_{-i}) - m(\omega|a)] \right)$$

The bounding program

- Thus, we might as well take the actions to be ordered, and maximize the "local" lower bound
- Let $M(K)$ be the set of rules $m : \{0, \ldots, K\}^N \to \Delta(\Omega)$ for which 0 is secure for all i
- Then G^* is equal to the supremum over (C, K) of

$$\sup_{m \in M(K)} \sum_{\theta} \mu(\theta) \min_{a} \underbrace{\sum_{\omega} \left(w(\omega, \theta) m(\omega|a) + C \sum_{i} u_i(\omega, \theta)[m(\omega|a_i + 1, a_{-i}) - m(\omega|a)] \right)}_{\equiv \textbf{strategic virtual objective at } (a,\theta)}$$

- So, G^* is the
  highest (across mechanisms)
  expected (across states)
  lowest (across actions) strategic virtual objective
- For fixed (C, K), this is a linear program!

Information structures

- There is a parallel theory of dual reductions for information design
- The potential $P(S, \pi)$ is the highest expected welfare across all mechanisms and equilibria subject to interim IR
- P^* is the infimum potential, which is an upper bound on G^*
- Dual reduction of information structures can be used to show that P^* is similarly attained with information structures in which signals are ordered, and the only binding constraints are local downward IC and IR at the bottom, with a constant multiplier
- If $P^* = G^*$, then the potential minimizers and guarantee maximizers "certify" each other as optimal

Example: Bargaining with incomplete information

- Legal arbitration, peace negotiation, legislative bargaining
- A unit surplus is to be split between agents 1 and 2
- If agreement is not reached, the pie shrinks to $\delta \in (0,1)$, and $\theta \in [0,1]$ is the fraction of surplus that would be obtained by agent 1
- No distributional assumption on θ (drop the μ constraint)
- $\omega \in \{\emptyset\} \cup [0,1]$
  - $\omega \in [0,1]$ means agreement, with agent 1 obtaining a share ω

$$u_1(\omega,\theta) = \omega,\ \ u_2(\omega,\theta) = 1-\omega,\ \ w(\omega,\theta) = 1$$

  - $\omega = \emptyset$ means disagreement, both agents receive their outside option

$$u_1(\emptyset,\theta) = \underline{u}_1(\theta) = \delta\theta,\ \ u_2(\emptyset,\theta) = \underline{u}_2(\theta) = \delta(1-\theta),\ \ w(\emptyset,\theta) = \delta$$

- (Demonstration)

Key points

- Agreement prob is of the form $q(a_1 + a_2)$, with the precise functional form equalizing the SVO across levels
- Sharing rule is proportional:

$$\omega(a) = \frac{(1-\delta)a_1 + \delta a_2}{a_1 + a_2}$$

- a_i is an expression of desire for agreement
  Intuitive tradeoff: higher $a \implies$ higher prob of agreement but smaller share
- Simple, general purpose class of mechanisms that only relies on an upper bound on δ
- If $\delta < 1/2$, $G^* = 1$, but if $\delta > 2/3$, $G^* = \delta$
- In this example, $G^* = P^*$: Low guarantee is not a result of equilibrium selection! When $\delta > 2/3$, no mechanism can obtain positive probability of agreement in **any** equilibrium

Other examples

- Brooks and Du (2021): Optimal auctions with common values, proportional auctions, can guarantee FSE with $N \to \infty$
- Brooks and Du (2024): No duality gap for interdependent-value multi-good auctions
- Brooks, Du, Feffer (2026): Asymmetric interdependent value auctions, compound proportional auctions
- Brooks and Du (2023): Proportional cost sharing mechanisms for public expenditure, proportional price trading mechanisms for bilateral trade
  (both are reparametrizations of the bargaining model)
- Brooks, Du, Zhang (2026): Market order mechanisms

Retrospective on common prior guarantee maximization

Pro

- Not overfit to particular information structures, or even
- Captures some notion of "informational efficiency"
- Simple pattern of incentive constraints is **derived** rather than assumed: most tempting deviation pulls agents away from outside option
- Guarantee holds over **all** equilibria
- Generalizable (in some respects) to lower and upper bounds on information
- It just works!

Con

- Overly pessimistic
- Not sure what to make of randomization
- Not an intuitive "language" for communicating with the mechanism

Open questions

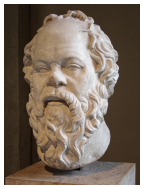
- Within guarantee maximization, there is still plenty to do to keep us busy, and open questions that AI still can't answer:
  - What are specifications that are insightful and analytically tractable?
  - How can we better interpret the actions?
  - When is there no duality gap ($G^* = P^*$)?
  - How can we incorporate lower and upper bounds on information into the theory?
    (Brooks, Du, Haberman 2025)
  - The mysterious "double revelation principle"
  - Weaker implementation concepts?
  - More explicit modeling of moral hazard?

Looking more broadly

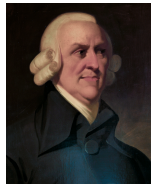
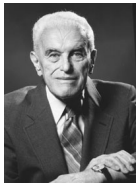
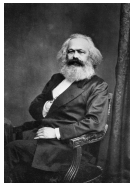
- Guarantee maximization is a beautiful and powerful technique, but hardly the definitive solution to MD's problems
- For MD to become more useful as a normative theory, and solve the bigger problems of how to design governments and economic systems, we will need new modeling approaches
- That means finding the right environments on which to optimize, that will deliver mechanisms that are not too complicated but are still effective at shaping behavior and promoting welfare
- But it also means more directly confronting issues of informational efficiency and limited computational and cognitive resources that Hurwicz was writing about long ago, and have fallen by the wayside

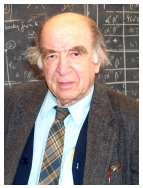
The enforcement problem

- Hurwicz titled his Nobel lecture "But Who Will Guard the Guardians?"
- For small scale mechanisms, we can appeal to external enforcement mechanisms like the contracts clause to uphold the rules
- But for designing constitutions, the question of enforcement is essential, and needs to return to the foreground
- Participation constraints are a reduced form for whatever happens if someone challenges the mechanism... but exactly what happens in that continuation equilibrium, and how is it related to the outcome of the mechanism? This needs deeper consideration
- A truly complete normative theory of a mechanism should account for the possibility of the mechanism's own demise, through either amendment or replacement by future generations






Thank you all!