Sistemas multiagente
Teoría de Juegos

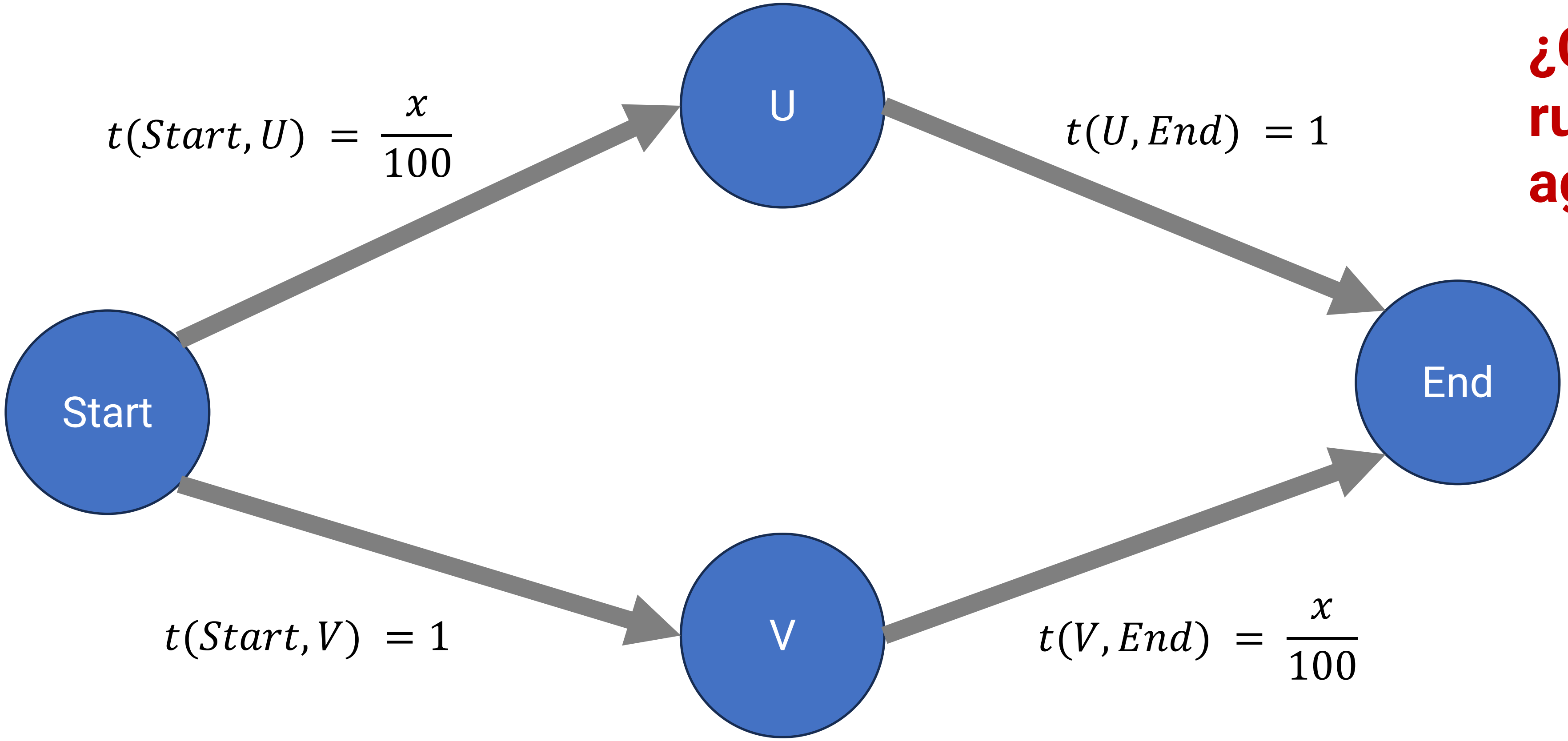
Sistemas Inteligentes Distribuidos

Sergio Alvarez
Javier Vázquez

Bibliografía

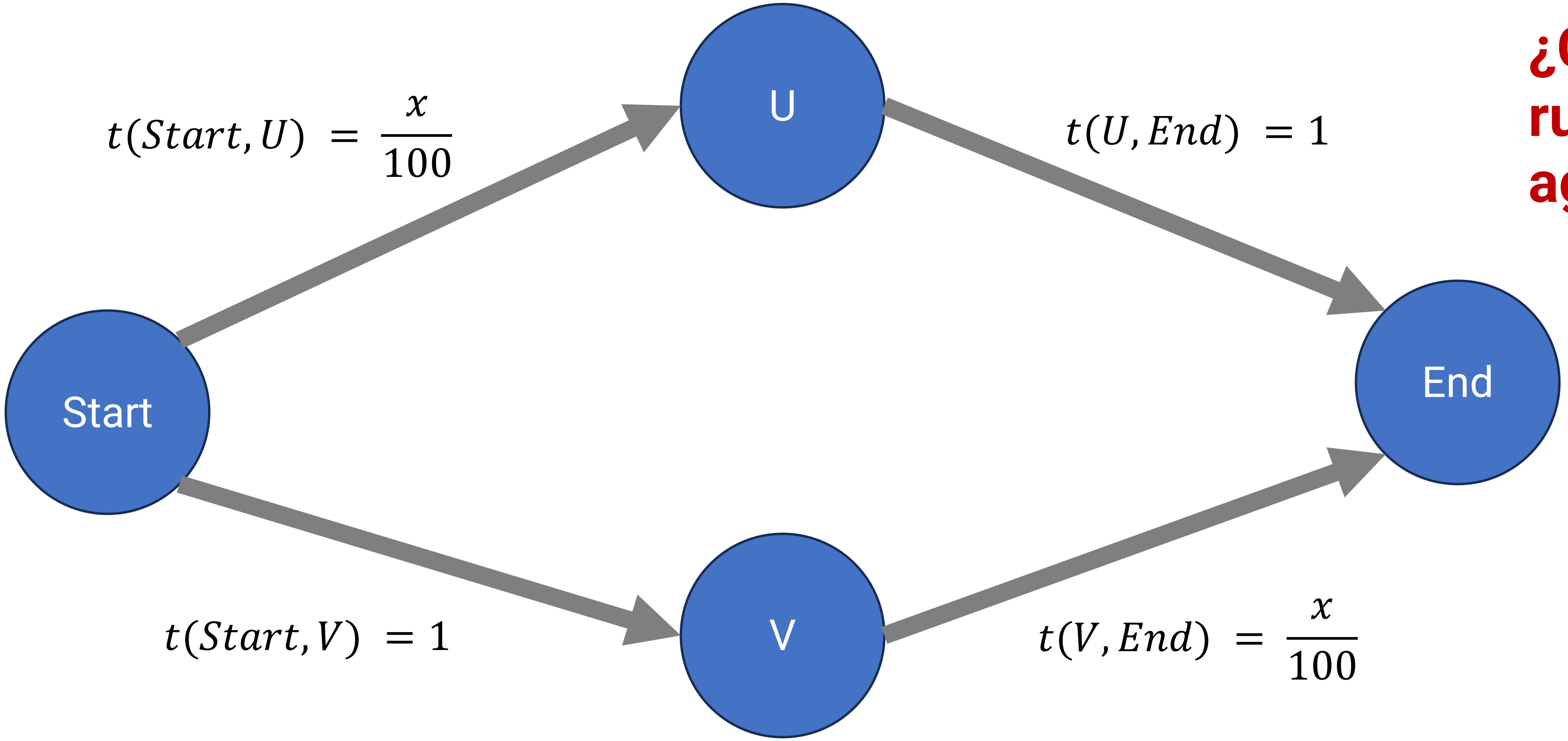
- *Artificial intelligence: a modern approach* (Russell & Norvig), cap. 6, 16, 17
- *Multi-Agent Reinforcement Learning* (Albrecht et al.), cap. 2, 8
- *Multiagent systems: algorithmic, game-theoretic, and logical foundations* (Shoham), cap. 3, 4, 5, 6
- Dafoe, A., Hughes, E., Bachrach, Y., Collins, T., McKee, K. R., Leibo, J. Z., et al. (2020). *“Open Problems in Cooperative AI.”* arXiv preprint arXiv:2012.08630. https://arxiv.org/abs/2012.08630

Paradoja de Braess (1968)



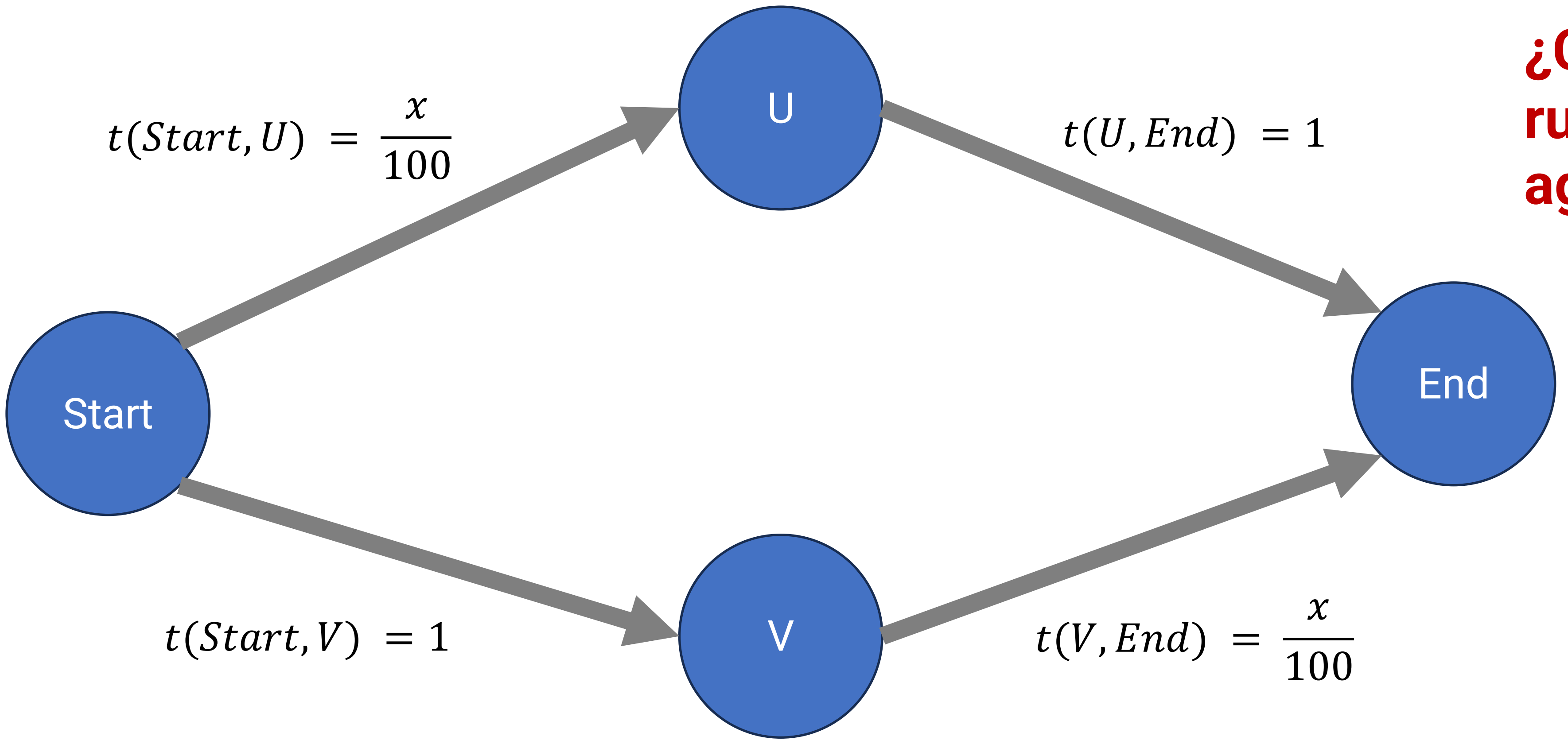
¿Cuál es la mejor ruta para un agente?

Paradoja de Braess (1968)



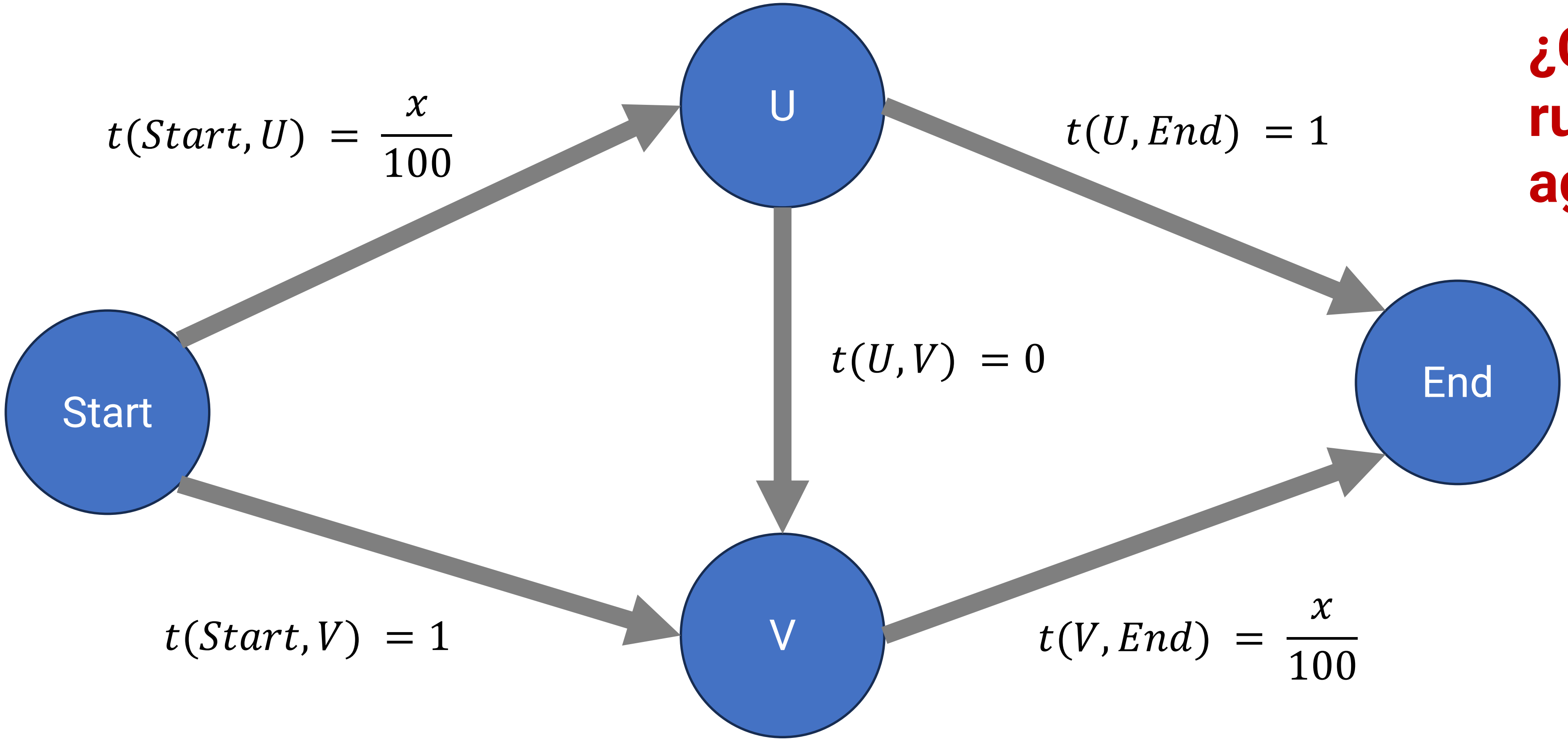
¿Cuál es la mejor ruta para dos agentes?

Paradoja de Braess (1968)



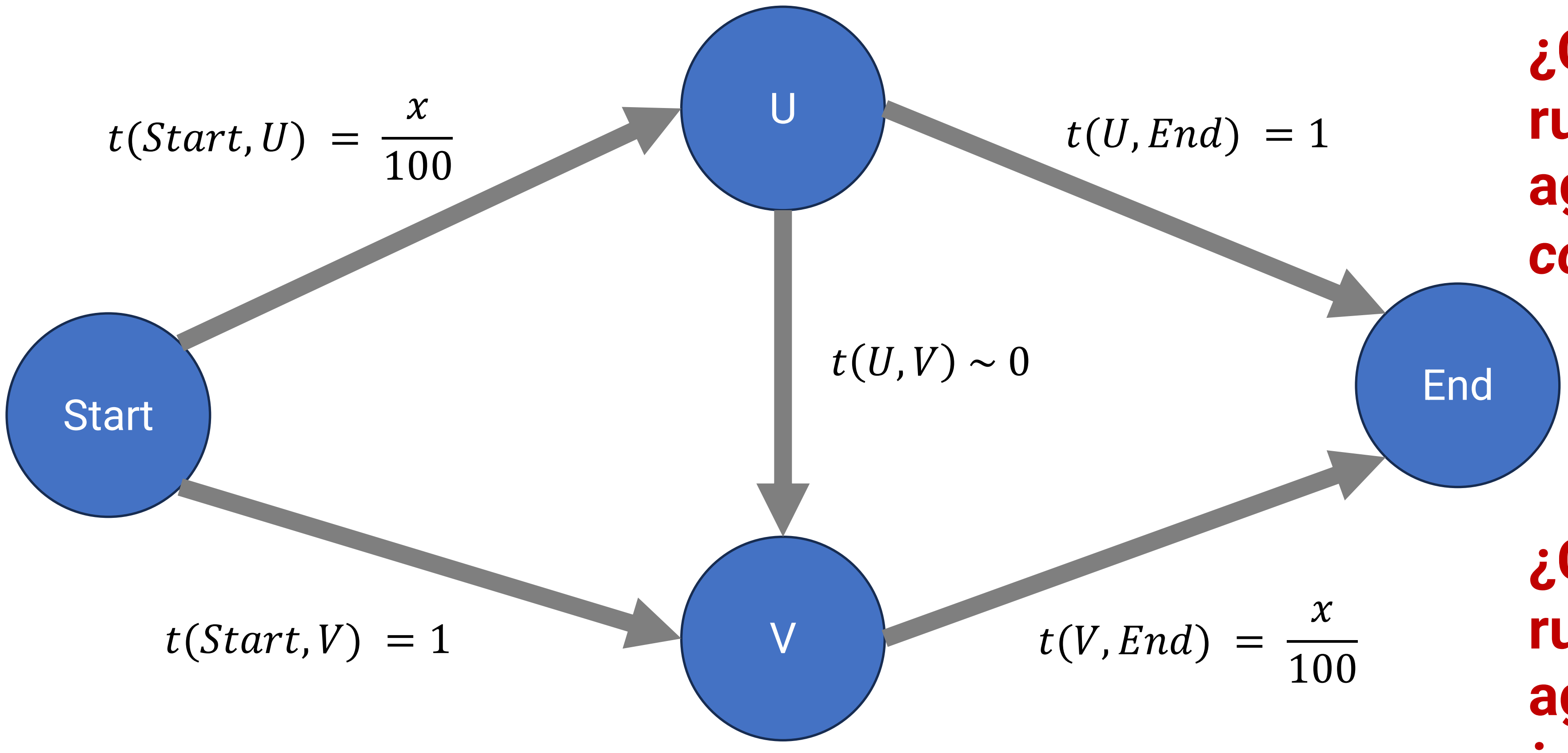
¿Cuál es la mejor ruta para 100 agentes?

Paradoja de Braess (1968)



¿Cuál es la mejor ruta para 100 agentes?

Paradoja de Braess (1968)



¿Cuál es la mejor ruta para 100 agentes, *colectivamente*?

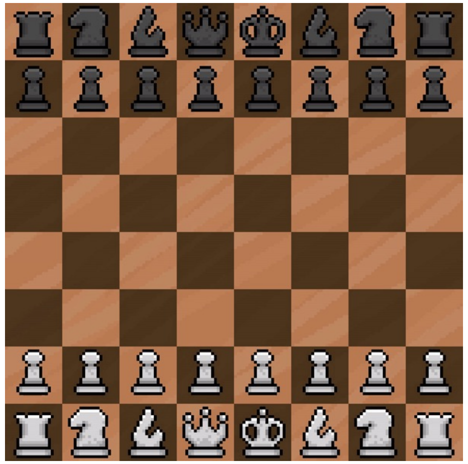
¿Cuál es la mejor ruta para 100 agentes, *individualmente*?

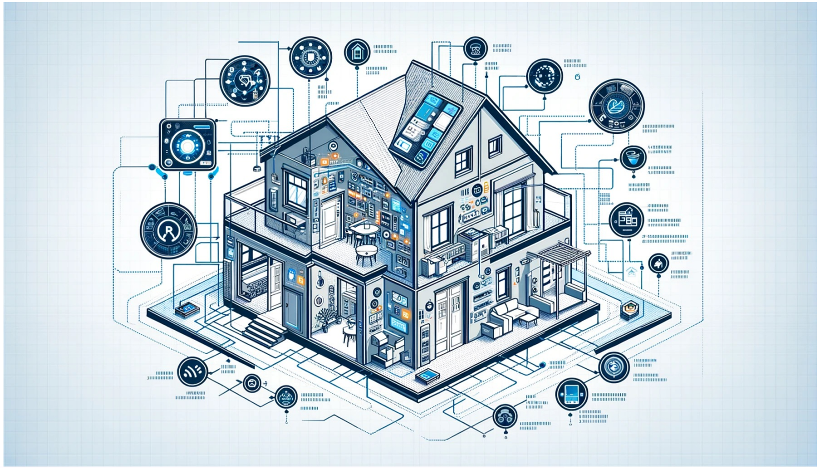
Sistemas Multiagente

Sistemas Multiagente y Teoría de Juegos

Entornos multiagente

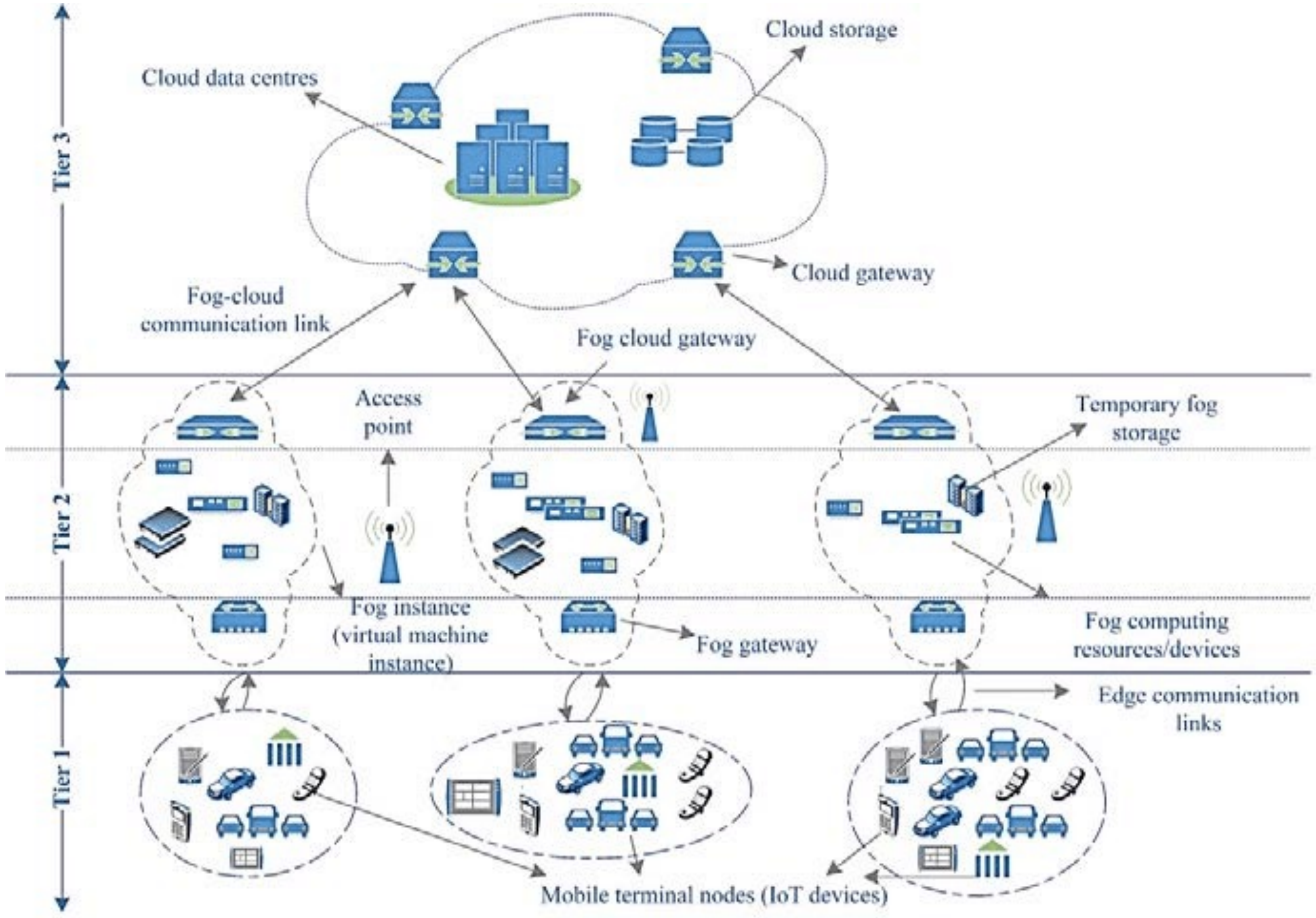
- Hasta ahora, nos hemos centrado en modelos computacionales diseñados para la toma de decisiones en entornos de un agente
- Sin embargo, estos modelos no suelen ser adecuados para entornos multiagente
    - Los agentes relevantes para un agente A son aquellos cuyo comportamiento *racional* puede tener influencia en el rendimiento de A
    - El entorno es **multiagente** si hay como mínimo: **un agente A, y un agente relevante para A**
    - Los entornos multiagente pueden ser cooperativos, competitivos, o una mezcla de ambos






Orders: onion
Score: 0
Time Left: 28


TUXEDO

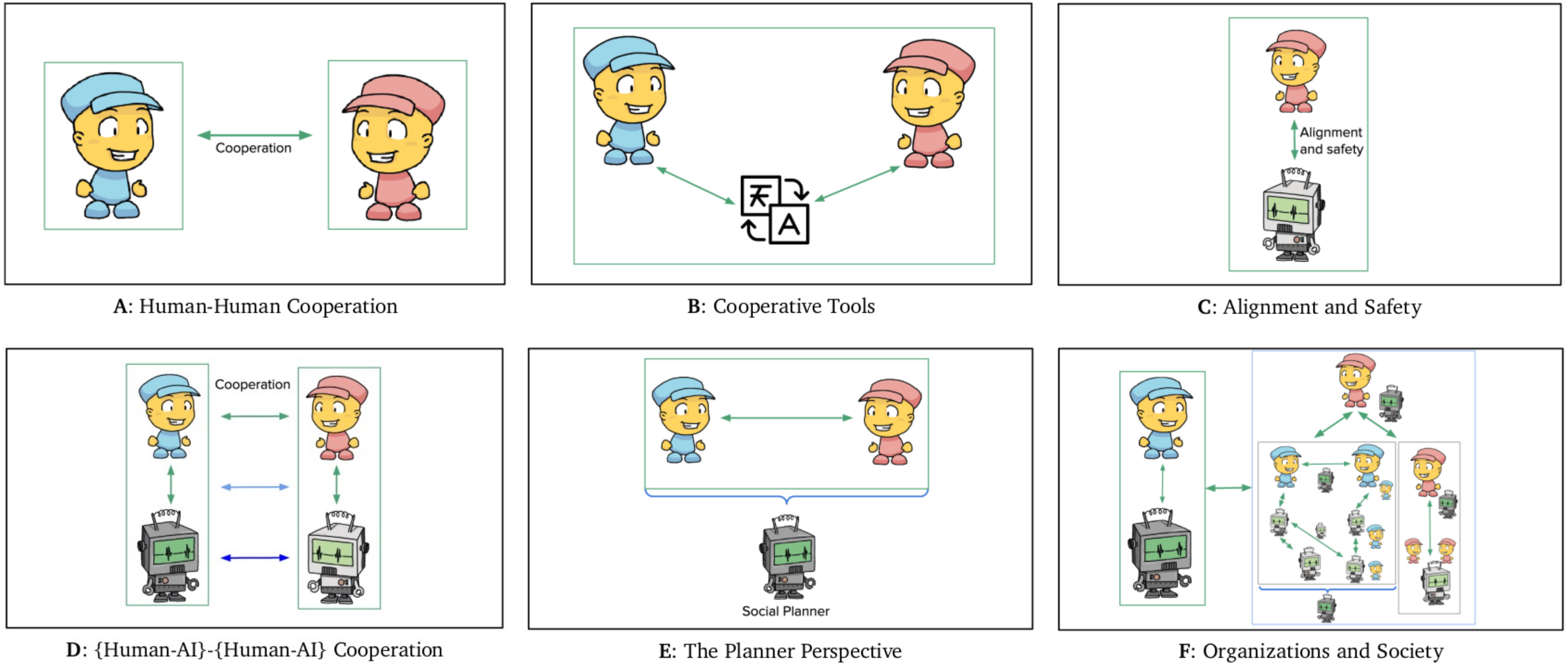


Taneja, M., & Davy, A. (2016). Resource aware placement of data analytics platform in fog computing. Procedia Computer Science, 97, 153-156.
Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE internet of things journal, 3(5), 637-646.

Campo multidisciplinar

- Los modelos computacionales se inspiran y retroalimentan a áreas de conocimiento que utilizan abstracciones de cooperación
  - Biología
  - Ciencias sociales
  - Genética
  - Política
- Los sistemas multiagente se usan para la simulación de componentes con ciertos grados de autonomía
  - Gestión de recursos: redes eléctricas, redes de abastecimiento, transporte
  - Microeconomía: mercados, subastas, comercio
  - Coordinación en situaciones de emergencia o catástrofe
- Hay un interés creciente en seguir avanzando en el área para anticiparse a nuevos problemas presentes y futuros
  - Interacción humano – máquina
  - Robots *opacos*
  - Conducción autónoma
  - Seguridad, equidad, interpretabilidad, fiabilidad

Modos de cooperación humano-máquina



Dafoe, A., Hughes, E., Bachrach, Y., Collins, T., McKee, K. R., Leibo, J. Z., et al. (2020). Open Problems in Cooperative AI. arXiv preprint arXiv:2012.08630.

Preguntas abiertas

- ¿Es posible que los agentes tengan mecanismos computacionales que permitan generar y ”entender” objetivos colectivos y tomar decisiones en base a ellos?
  - ¿Puede un agente aprender estos mecanismos?
- ¿Qué mecanismos pueden existir que faciliten la comunicación entre agentes inteligentes diseñados de forma independiente?
- ¿Puede emerger la cooperación entre sociedades de agentes egoístas?
  - ¿Qué forma debería tener un artefacto que represente este vínculo entre agentes?
- ¿Pueden los agentes establecer y hacer cumplir compromisos estables que restrinjan la consecución de sus propios objetivos?
- ¿Es posible garantizar que un agente, interactuando con otros agentes incluyendo humanos, se pueda comportar de manera acorde a unos estándares de comportamiento colectivo?

Modelos computacionales: MDPs

- MDPs
- **Partially Observable Markov Decision Processes (POMDPs)**
  - $\langle \boldsymbol{S}, \boldsymbol{A}, \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{B}} \rangle$ tal que
    - Ω es el conjunto (finito) de observaciones
    - $\mathcal{O}(o|a,s)$ es la probabilidad de observar o cuando el agente toma la acción a y sucede una transición a s
    - $\mathcal{B}(s_t) = Pr(s_t = s | s_0, a_1, o_1, a_2, o_2, \ldots, a_{t-1}, o_{t-1})$
- **Decentralised POMDPs (Dec-POMDPs)**
  - $\langle \boldsymbol{S}, \{\boldsymbol{A}\}_{\boldsymbol{i}}, \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mu}, \boldsymbol{\gamma}, \{\boldsymbol{\Omega}_{\mathbf{i}}\}, \boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{B}} \rangle$ tal que
    - A_i es el conjunto de acciones para el agente i
    - Ω_i es el conjunto de observaciones para el agente i

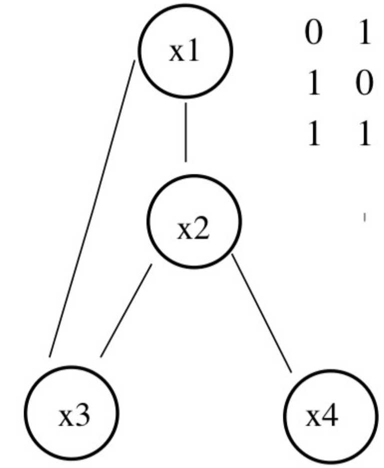
$\}$ **P**

$\}$ **PSPACE-completos**

$\}$ **NEXP-completos**

Modelos computacionales: DCOPs

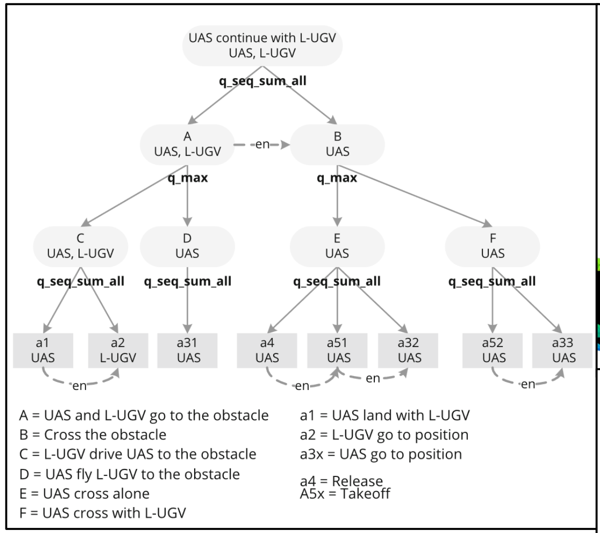
- **Decentralized Constraint Optimization Problems**, $\langle A, V, \mathcal{D}, f, \alpha, F \rangle$:
  - A es el conjunto de agentes
  - V es el conjunto de variables
  - $\mathcal{D}$ es el conjunto de dominios para las variables V
  - f es una función de coste para cada asignación de valor a una variable
  - α es una función de asignación de variables a agentes, generalmente $A = V$
  - F es una función de agregación de todos los costes, generalmente $\sum f$
- **Permiten definir el comportamiento** de los agentes en base al estado de todos
- Existen **algoritmos totalmente descentralizados** para resolver DCOPs
- Adecuados para dominios altamente distribuidos: IoT, smart grids, redes Wireless o P2P, control de tráfico…
- Otros formalismos: redes de Petri, event calculus, situation calculus, π-calculus

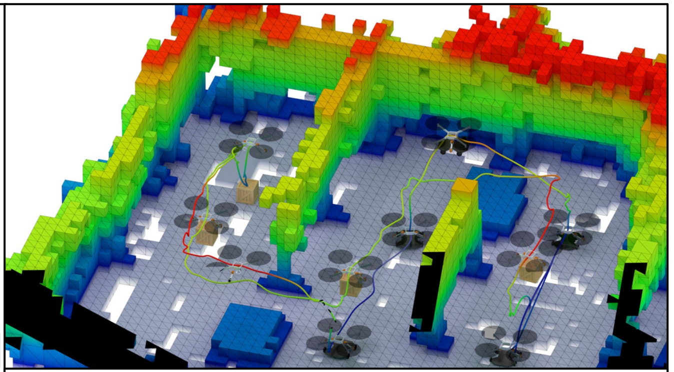


xi	xj	f(xi,xj)
0	0	1
0	1	2
1	0	2
1	1	0

Modelos computacionales: planificación

(Generalized) Partial Global Planning: construcción de planes colectivos a partir de la conjunción de los objetivos individuales, cumplimiento a partir de compromisos colectivos



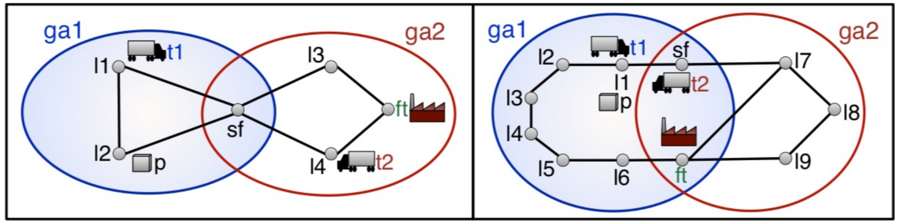


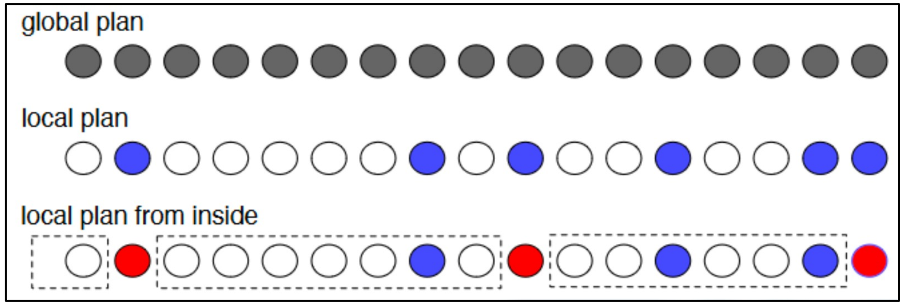
k	Schedules and commitments
1	UAS: $((a1, 0, 3), (a52, 3, 6), (a33, 6, 8))$
	L-UGV: $((a2, 0, 10))$
	Comm(UAS): $(a1, 3, a2)$
	Comm(L-UGV): $(a2, 10, B)$
2	UAS: $((a1, 0, 3), (slack, 3, 10), (a52, 10, 13), (a33, 13, 15))$
	L-UGV: $((a2, 3, 13))$
	Comm(UAS): $\emptyset$
	Comm(L-UGV): $(a2, 13, B)$
3	UAS: $((a1, 0, 3), (slack, 3, 13), (a52, 13, 16), (a33, 16, 18))$
	L-UGV: $((a2, 3, 13))$
	Comm(UAS): $\emptyset$
	Comm(L-UGV): $\emptyset$

Arbanas, B., Ivanovic, A., Car M., Orsag M., Petrovic, T. and Bogdan S., "Decentralized planning and control for UAV–UGV cooperative teams", Autonomous Robots (2018) 42:1601–1618

Versiones descentralizadas de algoritmos clásicos de planificación

- **MA-PDDL/MA-STRIPS**
- **MA-A***



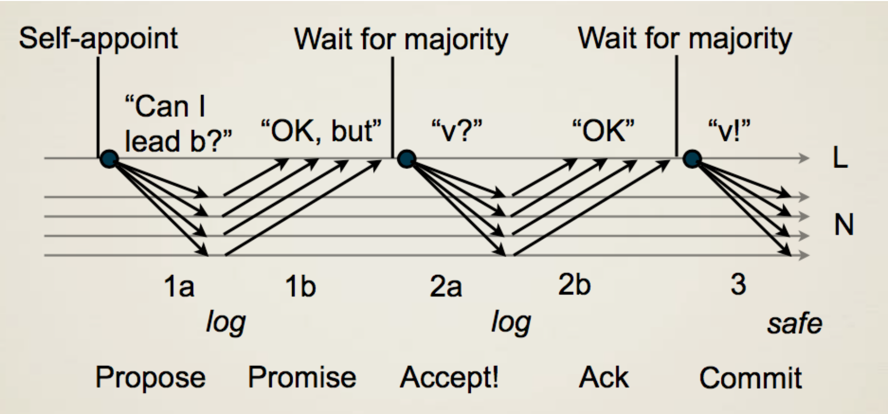


Torreño, Alejandro, et al. "Cooperative multi-agent planning: a survey." ACM Computing Surveys (CSUR) 50.6 (2018): 84.

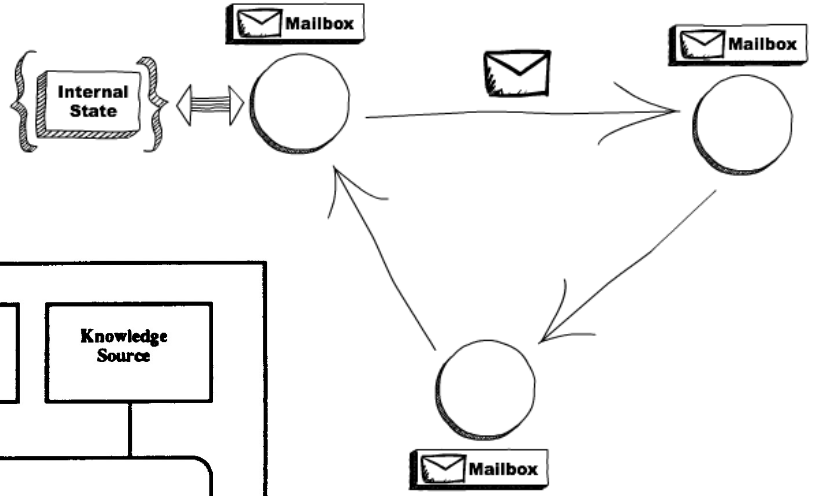
Modelos computacionales: sistemas distribuidos

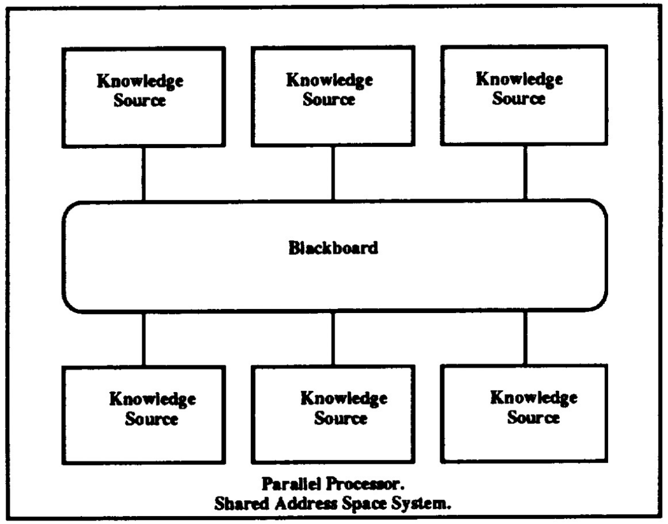
Algoritmos de consenso (e.g. **Raft, Paxos**)

- Máquinas de estados replicadas asíncronamente en todos los agentes que almacenan un **valor**
  - **El valor puede ser quién es el líder o cuál es la asignación actual de tareas**
- Garantías de convergencia y tolerancia a fallos
- Fundamentales en bases de datos distribuidas y en sistemas tipo Blockchain



Modelos de concurrencia: memoria compartida (**blackboard**) vs paso de mensajes (**actor model**)

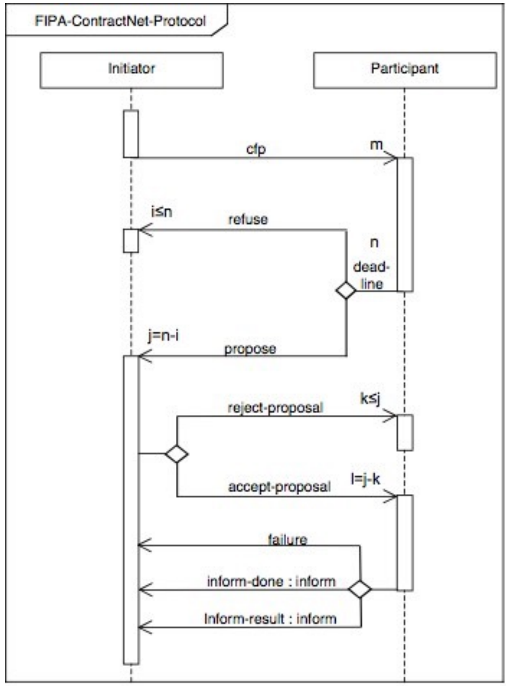


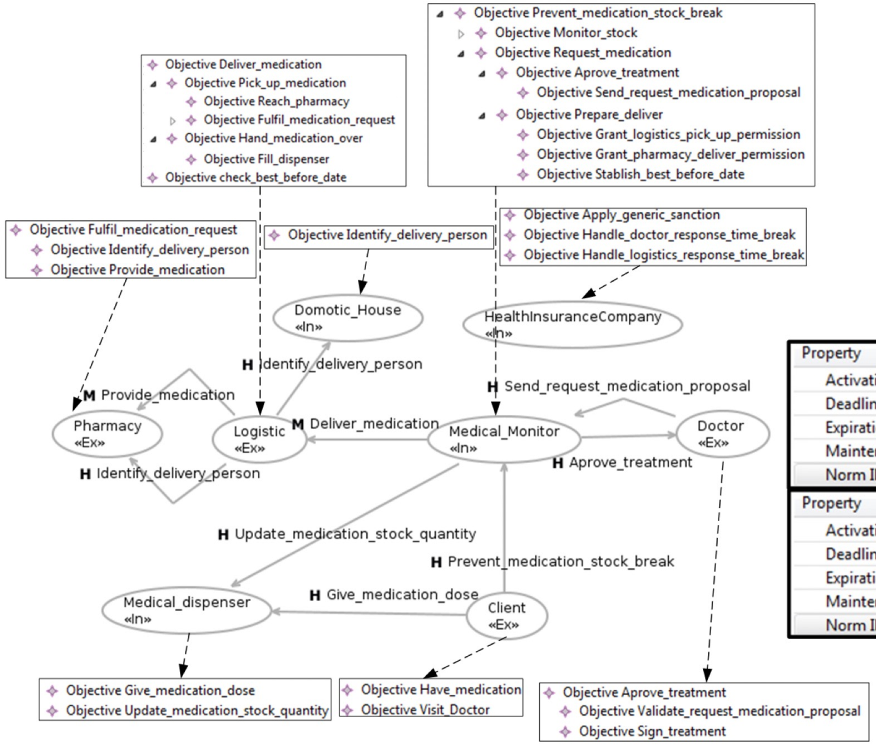


Modelos computacionales: sistemas sociotécnicos

Modelos simbólicos basados en el lenguaje (*speech acts*) y la comunicación

Protocolos de interacción





Normas, instituciones y organizaciones

Property	Value
Activation Condition	isActive(Provide_medication)
Deadline	
Expiration Condition	isFullfilled(check_best_before_date)
Maintenance Condition	¬ ~isFullfilled(Provide_medication)
Norm ID	o_check_best_before_date_before_Provide_medication

Property	Value
Activation Condition	isViolated(O_check_best_before_date_before_fill_dispenser)
Deadline	
Expiration Condition	isFullfilled(Apply_generic_sanction)
Maintenance Condition	
Norm ID	o_Apply_Generic_Sanction_fill_dispenser

Gómez-Sebastià, I., Garcia-Gasulla, D., Barrué, C., Vázquez-Salceda, J., & Cortés, U. (2010, August). A Flexible Agent-Oriented Solution to Model Organisational and Normative Requirements in Assistive Technologies. In CCIA (pp. 79-88).

Modelos computacionales: teoría de juegos

Teoría de Juegos

Sistemas Multiagente y Teoría de Juegos

Definiciones

- “The branch of mathematics concerned with the **analysis of strategies** for dealing with competitive situations **where the outcome of a participant's choice of action depends critically on the actions of other participants**. Game theory has been applied to contexts in war, business, and biology. Compare with decision theory.” – *Oxford Dictionary of English*
- “The study of mathematical models of **conflict and cooperation** between **intelligent rational decision-makers**” – *An Introduction to Game Theory* (Roger B. Myerson, 1984)

Dilema del prisionero

agente columna → ↓ agente fila	**C(ooperate)**	**D(efect)**
C(ooperate)	-1, -1	-20, 0
D(efect)	0, -20	-5, -5

Representación de **juego en forma normal**:

- Agente fila, agente columna (más agentes: más matrices)
- Matriz de recompensas $u_{fila}, u_{columna}$ (*payoffs* / utilidades)
- Suponemos una única decisión, simultánea e independiente

¿Cuál es la mejor estrategia para un agente racional?

- Dos sospechosos de un delito
- Se les separa e interroga, sin posibilidad de comunicación
- A cada uno se le ofrece la posibilidad de confesar y traicionar al otro (**D**efect), pero puede permanecer en silencio (**C**ooperate)
- Si los dos confiesan (**D**, **D**), pasarán 5 años en la cárcel
- Si uno confiesa (**D**) y el otro no (**C**), el primero sale libre y el otro pasará 20 años en la cárcel
- Si los dos se callan (**C**, **C**), pasarán ambos un año en la cárcel

Dilema del prisionero

- Vamos a suponer que somos el agente fila (en todo caso la matriz es simétrica)
- Una posibilidad conservadora es la **estrategia maximin**: **maximizar la mínima recompensa que podemos recibir**
  - La mínima recompensa que puedo recibir es -20 si coopero (**C**), -5 si traiciono (**D**): elijo traicionar (**D**)
- Dado un conjunto de agentes $\{i\} \cup -i$, y un conjunto de estrategias para cada agente S_i:

$$\underline{u_i} = \max_{s_i} \min_{s_{-i}} u_i(s_i, s_{-i})$$

agente columna → ↓ agente fila	**C(ooperate)**	**D(efect)**
C(ooperate)	-1, -1	-20, 0
D(efect)	0, -20	**-5**, -5

Dilema del prisionero

- Otra posibilidad, más competitiva y arriesgada, es la **estrategia minimax**: **minimizar la máxima recompensa que el agente columna puede recibir**
  - La máxima recompensa que el agente columna puede recibir es 0 si coopero (**C**), -5 si traiciono (**D**): elijo traicionar (**D**)
- Dado un conjunto de agentes $\{i\} \cup -i$, y un conjunto de estrategias para cada agente S_i:

$$\overline{u_i} = \min_{s_{-i}} \max_{s_i} u_i(s_i, s_{-i})$$

agente columna → ↓ agente fila	C(ooperate)	D(efect)
C(ooperate)	-1, -1	-20, 0
D(efect)	0, -20	-5, **-5**

Las estrategias maximin y minimax buscan seguir un modelo conductual concreto que modela una cierta aversión al riesgo y/o competitividad, pero **¿cómo podemos estar seguros de si son óptimas?**

Dilema del prisionero

- ¿Hay alguna estrategia que supere a las demás, **incondicionalmente**?
- Si el agente columna elige cooperar (**C**):
  - Si coopero (**C**), recibo -1, si traiciono (**D**), recibo 0
  - Por lo tanto, prefiero (**D**)
- Si el agente columna elige traicionar (D):
  - Si coopero (**C**), recibo -20, si traiciono (**D**) recibo -5
  - Por lo tanto, prefiero (**D**)

agente columna → ↓ agente fila	**C(ooperate)**	**D(efect)**
C(ooperate)	-1, -1	-20, 0
D(efect)	0, -20	-5, -5

- La estrategia **D domina estrictamente**
- Entonces, ¿cuál es el dilema?
  - Lo veremos más adelante, pero **¿es ésta la mejor opción para el sistema multiagente?**

Estrategia dominante

- Dado un conjunto de agentes $\{i\} \cup -i$, un conjunto de estrategias para cada agente S_i y dos estrategias $s_i, s_i' \in S_i$:
  - s_i **domina débilmente** a $s_i{}'$ si
    $\forall s_{-i} \in S_{-i}, u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$
  - s_i **domina estrictamente** a $s_i{}'$ si
    $\forall s_{-i} \in S_j, u_i(s_i, s_{-i}) > u_i(s_i', s_{-i})$
- Una **estrategia dominante** se suele denotar como s_i^* y se considera **óptima para *i*** y cumple:
  $\forall s' \in S_i, \forall s_{-i} \in S_{-i}, u_i(s_i^*, s_{-i}) \geq u_i(s_i', s_{-i})$

agente columna → ↓ agente fila	C(ooperate)	D(efect)
C(ooperate)	-1, -1	-20, 0
D(efect)	0, -20	-5, -5

$u_{fila}(D, C) = 0$

$u_{columna}(D, D) = -5$

En teoría de juegos, se asume que un agente racional **siempre** preferirá la estrategia óptima

¿Estrategia dominante?

- La existencia de una estrategia dominante por parte de un agente garantiza la existencia de una **solución** al juego para ese agente
- ¿Existe siempre una estrategia dominante?
- Tenemos dos agentes que habían quedado para cenar y se han quedado sin batería en el móvil
  - No recuerdan si habían quedado en la pizzería o en la hamburguesería
  - Vayan donde vayan, como no pueden comunicarse, se quedarán a cenar ahí, aunque sea a solas
  - Un agente (fila) preferiría la pizzería y el otro (columna) la hamburguesería
  - En cualquier caso, preferían no cenar a solas
  - **¿Dónde deberían ir?**

¿Estrategia dominante?

	P	**H**
P	2, 1	0, 0
H	0, 0	1, 2

Un agente (fila) preferiría la pizzería (+1) y el otro (columna) la hamburguesería (+1)

En cualquier caso, preferían no cenar a solas (+1/-1)

- Si el agente columna elige P:
  - Si el agente fila elige P recibe 2, y si elige H recibe 0, por lo tanto prefiere P
- Si el agente columna elige H:
  - Si el agente fila elige P recibe 0, y si elige H recibe 1, por lo tanto prefiere H
- **NO hay una estrategia dominante**

Equilibrio de Nash

	P	H
P	2, 1	0, 0
H	0, 0	1, 2

Un agente (fila) preferiría la pizzería (+1) y el otro (columna) la hamburguesería (+1)

En cualquier caso, preferían no cenar a solas (+1/-1)

- NO hay una estrategia dominante
- **Tampoco hay estrategia maximin (es indiferente)**
- **Estrategia minimax: P (agente fila), H (agente columna)**
  - Pero… ¿queremos seguir una estrategia tan competitiva en este juego?

Equilibrio de Nash

	P	**H**
P	2, 1	0, 0
H	0, 0	1, 2

Un agente (fila) preferiría la pizzería (+1) y el otro (columna) la hamburguesería (+1)

En cualquier caso, preferían no cenar a solas (+1/-1)

- NO hay una estrategia dominante
- Sin embargo, hay dos casos que *contentarían* a ambos agentes: (P, P) y (H, H)
  - En ambos casos, ninguno de los agentes querría haber escogido una estrategia diferente
  - Esto nos da una pista sobre cómo formular soluciones para juegos sin estrategias dominantes

Equilibrio de Nash

	P	H
P	**2, 1**	0, 0
H	0, 0	**1, 2**

Un agente (fila) preferiría la pizzería (+1) y el otro (columna) la hamburguesería (+1)

En cualquier caso, preferían no cenar a solas (+1/-1)

- Un equilibrio de Nash es una situación en la que, si los demás agentes mantienen una estrategia, cualquier agente presente en el sistema preferirá mantener la suya propia
  - *"If you keep doing what you're doing, I'll keep doing what I'm doing"* (Tim Roughgarden)
- (P, P) y (H, H) son **equilibrios de Nash**
- **Los equilibrios de Nash se consideran soluciones del sistema**

Equilibrio de Nash

- Dado un conjunto de agentes $\{i\} \cup -i$, un perfil de estrategias $s^* = (s_i^*, s_{-i}^*)$ es un equilibrio de Nash si

$$\forall s_i \in S_i, u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*)$$

- Como hemos visto, pueden existir múltiples equilibrios de Nash
- Si el agente no tiene una preferencia entre los múltiples equilibrios, se dice que estos equilibrios son *débiles*
- Un equilibrio de Nash es *fuerte* cuando

$$\forall s_i \in S_i, s_i \neq s_i^* \rightarrow u_i(s_i^*, s_{-i}^*) > u_i(s_i, s_{-i}^*)$$

¿Equilibrio de Nash?

- ¿Existe siempre un equilibrio de Nash?
- Vamos a jugar a otro juego, en este caso de suma cero: *piedra, papel, tijera*
    - Un juego de suma cero es aquel en el que la suma de recompensas es constante independientemente del perfil de estrategias

	piedra	**papel**	**tijera**
piedra	0, 0	-1, 1	1, -1
papel	1, -1	0, 0	-1, 1
tijera	-1, 1	1, -1	0, 0

¿Equilibrio de Nash?

	piedra	papel	tijera
piedra	0, 0	-1, 1	1, -1
papel	1, -1	0, 0	-1, 1
tijera	-1, 1	1, -1	0, 0

- *“If you keep doing what you’re doing, I’ll keep doing what I’m doing”*
  - Si el agente fila supiera que el agente columna va a escoger piedra, entonces preferiría escoger papel
  - Pero si el agente columna supiera que el agente fila va a escoger papel, entonces preferiría escoger tijera
  - Y así, *ad infinitum*
- No es posible estabilizar este juego si hemos de escoger estrictamente entre piedra, papel o tijera
- Aparentemente, no hay equilibrio de Nash
  - **¿Qué estrategia escogeríais?**

Equilibrios mixtos

- Si la estrategia consiste en escoger una, y solo una, de las opciones (**estrategia pura**), **no hay garantía de equilibrio de Nash**
- Sin embargo, **sí que hay equilibrio garantizado si permitimos estrategias estocásticas** (**estrategias mixtas**), donde asignamos una distribución de probabilidad sobre las posibles estrategias puras (*Non-Cooperative games*, John Forbes Nash, 1950). Por ejemplo, en *piedra, papel, tijera*:

$$\left(p_{piedra}, p_{papel}, p_{tijera}\right) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

es una estrategia mixta que es estrategia dominante y equilibrio de Nash

- Una estrategia pura se puede ver, de hecho, como un caso especial de estrategia mixta:

$$\exists s_i \in S_i, p_{s_i} = 1 \wedge \forall s_i' \in S_i, \left[s_i \neq s_i' \rightarrow \; p_{s_i'} = 0\right]$$

Recompensa esperada

- Sea:
  - Un agente i, escogido de un total de n agentes
  - S_i el conjunto de estrategias puras del agente i
  - $\sigma_i(s_i)$ la estrategia mixta del agente i, formulada como la probabilidad de que i elija la estrategia pura s_i
  - $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ el perfil de estrategias mixtas para los n agentes
- La recompensa (utilidad) esperada para el agente i es:

$$E[u_i(\sigma)] = \sum_{s_i \in S_i} \sum_{s_j \in S_j} \dots \sum_{s_n \in S_n} \left(\prod_{j=1}^{n} \sigma_j(s_j) \right) \cdot u_i(s_1, s_2, \dots, s_n)$$

Recompensa esperada

$$\boldsymbol{E[u_i(\sigma)]} = \sum_{s_i \in S_i} \sum_{s_j \in S_j} \dots \sum_{s_n \in S_n} \left(\prod_{j=1}^{n} \boldsymbol{\sigma_j(s_j)} \right) \cdot \boldsymbol{u_i(s_1, s_2, \dots, s_n)} = \sum_{s_i \in S_i} \sum_{s_j \in S_j} \left(\prod_{j=1}^{2} \sigma_j(s_j) \right) \cdot u_i(s_1, s_2) =$$

$$= \sum_{s_j \in S_j} \left(\prod_{j=1}^{2} \sigma_j(s_j) \right) \cdot u_i(piedra, s_2) + \sum_{s_j \in S_j} \left(\prod_{j=1}^{2} \sigma_j(s_j) \right) \cdot u_i(papel, s_2) + \sum_{s_j \in S_j} \left(\prod_{j=1}^{2} \sigma_j(s_j) \right) \cdot u_i(tijera, s_2) =$$

$$= \left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(piedra, piedra) + \left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(piedra, papel) + \left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(piedra, tijera) +$$

$$\left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(papel, piedra) + \left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(papel, papel) + \left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(papel, tijera) +$$

$$\left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(tijera, piedra) + \left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(tijera, papel) + \left(\frac{1}{3} \cdot \frac{1}{3} \right) \cdot u_i(tijera, tijera) =$$

$$= \frac{1}{9} \cdot 0 + \frac{1}{9} \cdot -1 + \frac{1}{9} \cdot 1 + \frac{1}{9} \cdot 1 + \frac{1}{9} \cdot 0 + \frac{1}{9} \cdot -1 + \frac{1}{9} \cdot -1 + \frac{1}{9} \cdot 1 + \frac{1}{9} \cdot 0 = \mathbf{0}$$

Ineficiencia del equilibrio

- Volviendo al dilema…

agente columna → ↓ agente fila	**C(ooperate)**	**D(efect)**
C(ooperate)	-1, -1	-20, 0
D(efect)	0, -20	-5, -5

- El equilibrio de Nash es $\sigma = (D, D)$
  - Equivalente a: $\sigma = \left(\sigma_{fila}, \sigma_{columna}\right)$ con $\sigma_{fila}(D) = \sigma_{columna}(D) = 1$
  - $u_{fila}(D, D) = u_{columna}(D, D) = -5$
- **¿Es ésta la mejor solución para el sistema multiagente?**

Ineficiencia del equilibrio

- Obviamente, depende de lo que definamos *mejor* desde el punto de vista del sistema
  - Necesitamos una definición de *racionalidad* que se extienda al ámbito colectivo
  - Veremos múltiples maneras de formalizar esta racionalidad durante el resto del curso
- Una fórmula de racionalidad estándar (llamada *utilitarianista*) en teoría de juegos es la maximización de la función de bienestar W (*welfare function*) que es, simplemente, la suma de las recompensas para todos los agentes:

$$W(\sigma) = \sum_i u_i(\sigma)$$

- Una fórmula alternativa es la llamada *igualitaria*: $W'(\sigma) = \min(u_i(\sigma))$

Ineficiencia del equilibrio

Para normalizar sumamos 20: *"¿de cuántos años de cárcel me libro?"*

- El precio de la anarquía se define como el ratio máximo entre los valores *normalizados* de:
  - El bienestar óptimo
  - El peor bienestar de los equilibrios

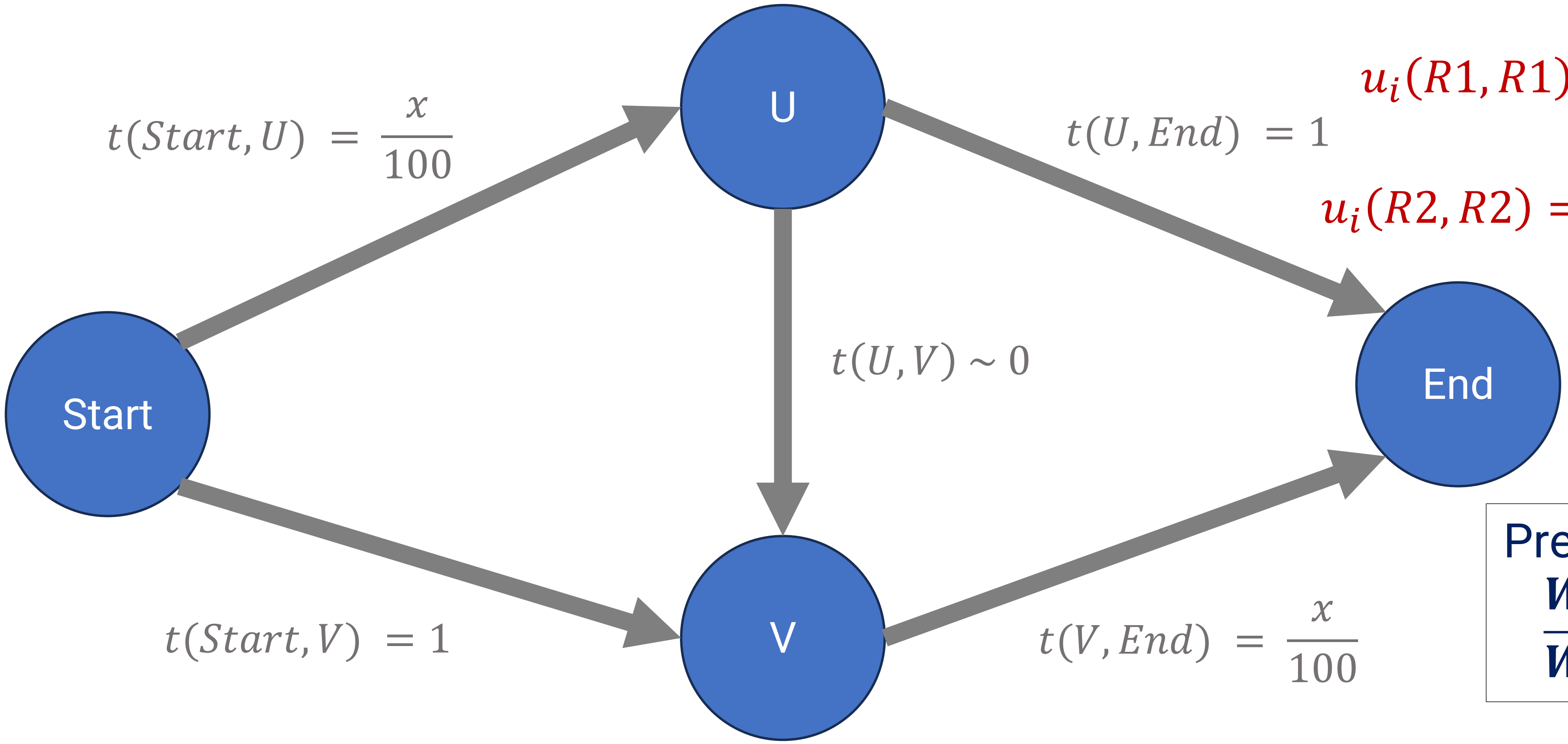
$$\frac{W(C,C)}{W(D,D)} = \frac{38}{30} = 1.27$$

- En este juego, el equilibrio de Nash es ineficiente con respecto del bienestar
- Este valor (> 1) representa la ineficiencia existente entre el comportamiento racional individual (egoísta) y un comportamiento hipotético centralizado

agente columna → ↓ agente fila	C(ooperate)	D(efect)
C(ooperate)	19, 19	0, 20
D(efect)	20, 0	15, 15

$$u_{fila}(C,C) = u_{columna}(C,C) = 19$$
$$u_{fila}(C,D) = u_{columna}(D,C) = 0$$
$$u_{fila}(D,C) = u_{columna}(C,D) = 20$$
$$u_{fila}(D,D) = u_{columna}(D,D) = 15$$
$$W(C,C) = 38$$
$$W(C,D) = W(D,C) = 20$$
$$W(D,D) = 30$$

Precio de la anarquía: paradoja de Braess



Supongamos estas dos estrategias mixtas:

$R1 = 100\% \ [Start \to U \to V \to End]$

$R2 = 50\% \ [Start \to U \to End], 50\% \ [Start \to V \to End]$

$$u_i(R1, R1) = \frac{1}{t(R1)} = \frac{1}{1+1} = 0.5$$

$$u_i(R2, R2) = \frac{1}{t(R2)} = \frac{1}{1 + \frac{50}{100}} = 0.67$$

Precio de la anarquía:

$$\frac{\boldsymbol{W(R2, R2)}}{\boldsymbol{W(R1, R1)}} = \frac{\mathbf{67}}{\mathbf{50}} = \mathbf{1.33}$$

Optimalidad de Pareto

- Método alternativo de mejorar la eficiencia del sistema: buscar el óptimo de Pareto
- Un perfil σ ofrece una **mejora de Pareto** (o tiene una **dominancia de Pareto**) respecto de otro perfil σ' cuando, en el cambio de σ a σ', como mínimo un agente mejora su recompensa sin que haya ningún agente que empeore la suya:

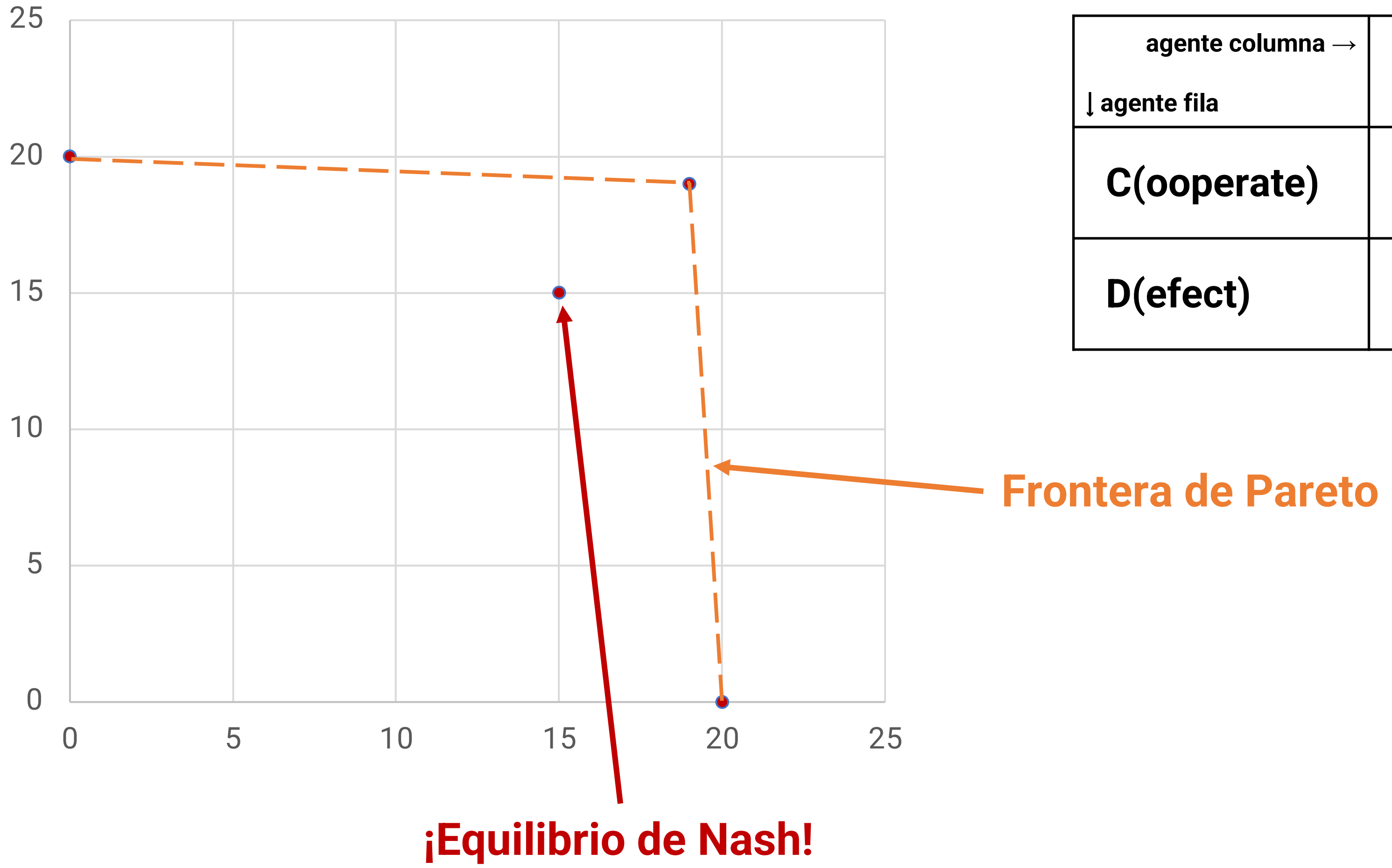
$$\forall i, u_i(\sigma) \geq u_i(\sigma') \land \exists j, u_j(\sigma) > u_j(\sigma')$$

- Un perfil σ es **Pareto-óptimo** o **Pareto-eficiente** si no hay ningún perfil σ' accesible desde s que mejore la recompensa de como mínimo un agente sin empeorar la de otro

$$\forall \sigma', \forall i, u_i(\sigma) \geq u_i(\sigma')$$

- Llamamos **conjunto o frontera de Pareto** al conjunto de perfiles Pareto-óptimos

Optimalidad de Pareto



agente columna → ↓ agente fila	C(ooperate)	D(efect)
C(ooperate)	19, 19	0, 20
D(efect)	20, 0	15, 15

Conceptos de solución

- No hay una única solución a un juego en forma normal, sino diversos **conceptos de solución** que serán más o menos adecuados dependiendo de los objetivos del sistema multiagente:
  - Estrategia maximin / estrategia minimax
  - Dominancia estratégica
  - Equilibrio de Nash
  - Precio de la anarquía
  - Optimalidad de Pareto
  - …

Propiedades importantes

- En un juego de suma cero, la suma de recompensas para todos los agentes no varía sea cual sea el perfil de estrategias: las pérdidas de un agente son las ganancias de los otros

$$\forall \sigma, \sigma', \sum_i u_i(\sigma) = \sum_i u_i(\sigma')$$

- Teorema Minimax (Von Neumann, 1928): en juegos de suma cero de dos jugadores, existe (y se puede encontrar en tiempo polinómico) un valor v tal que
  - Existe una estrategia s_i^* tal que el jugador i puede garantizar una recompensa mínima de v y
  - Existe una estrategia s_j^* tal que el jugador j puede garantizar una recompensa máxima de v
  - **Corolario: equilibrio de Nash = estrategia minimax = estrategia maximin**
- En juegos en forma normal, en general, **encontrar un equilibrio de Nash es PPAD-completo** (Papadimitriou, 1994), incluso para dos agentes (Chen, 2006)
  - $\text{PPAD} \subseteq TFNP \subseteq FNP \subseteq NP$

Otras formas de representación

- En forma normal (o forma estratégica), el foco está en el análisis de las estrategias y equilibrios a partir de una fotografía estática del entorno y las funciones de utilidad de los agentes
- Forma repetida
  - Repeticiones de la misma decisión en forma normal
- Forma secuencial
  - Las decisiones no son únicas y se toman en secuencia
  - Las estrategias en este caso se desarrollan a lo largo del tiempo
- Forma extensiva
  - Se suelen representar en forma de árbol, donde los nodos representan puntos de decisión y las ramas representan acciones posibles
- Veremos estas formas más adelante (Sesión ~10)

Coordinación multiagente

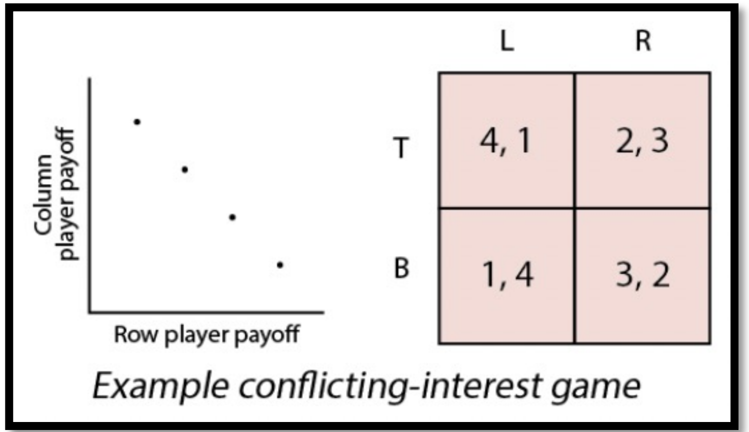
Sistemas Multiagente y Teoría de Juegos

Toma de decisiones multiagente

- En las próximas semanas, vamos a estudiar la influencia de diferentes modelos computacionales en el proceso de toma de decisiones de agentes situados y autónomos
- Asumiremos que cada agente es *racional*, según la definición usada hasta ahora
- Sin embargo, nos preocuparemos también sobre cómo combinar esa racionalidad con los objetivos colectivos (organizacionales, institucionales…)
- Llamamos **coordinación** al **proceso que siguen los agentes para gestionar las dependencias entre sus actividades**
- La aplicabilidad de cada modelo dependerá en buena parte del tipo de cooperación/competición esperable en cada situación
  - La situación depende del entorno y de las preferencias de los agentes
  - Podemos usar conceptos de teoría de juegos para caracterizar las situaciones

Competición: agentes con intereses conflictivos

- En este tipo de situación, los agentes están diseñados para cumplir objetivos (o funciones de utilidad) inherentemente conflictivos, y/o simplemente son puramente egoístas
- El incremento en las recompensas de un agente va asociado a un decremento en la recompensa de los demás (e.g. juegos de suma cero)
- Una solución a estas situaciones es la aplicación de algoritmos para el cálculo de estrategias competitivas (e.g. **minimax**)
- Otra posibilidad es modificando los equilibrios de Nash
    - Se puede intervenir o diseñar el entorno para que los agentes conflictivos cumplan con los objetivos colectivos como efecto colateral de su comportamiento racional: **diseño de mecanismos**

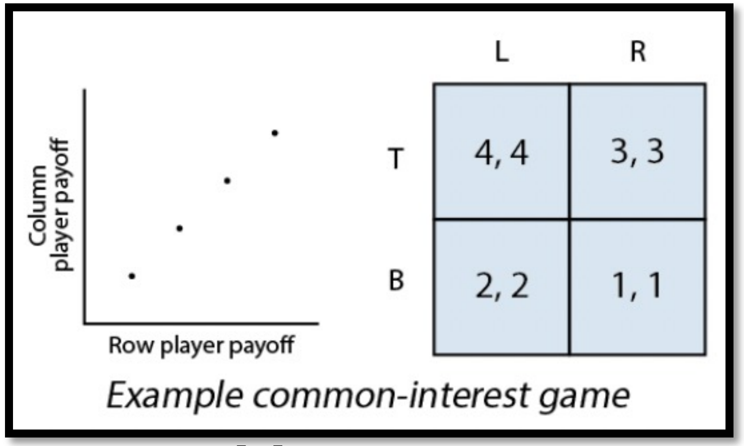


[8] (Dafoe et al)

No suele haber equilibrios de Nash. Si los hay, nunca serán Pareto-óptimos

Cooperación: agentes con intereses comunes

- Cualquier incremento en la recompensa de un agente irá generalmente asociado a un incremento en la recompensa de los demás agentes
- Desde un punto de vista colectivo, el objetivo es mantener el rendimiento y la estabilidad colectivos sin afectar el comportamiento autónomo de los agentes
- Por lo tanto, los métodos a aplicar en estos casos son aquellos que permiten una comunicación efectiva, por ejemplo:
  - **Protocolos de interacción**
  - **Algoritmos de consenso**
  - **Elección social**

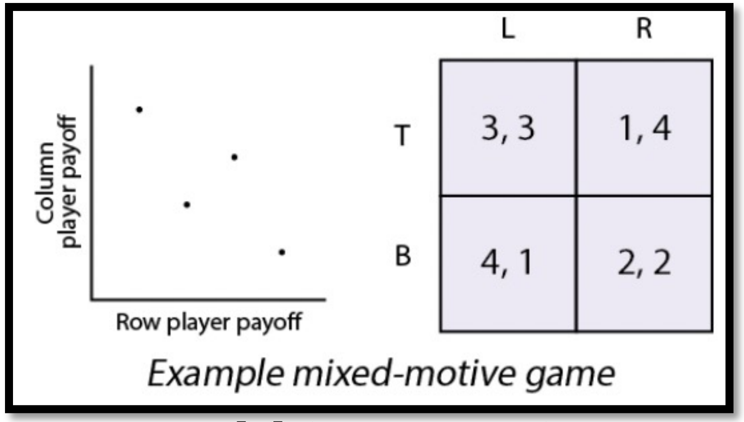


[8] (Dafoe et al)

Hay equilibrios de Nash que coinciden con los óptimos de Pareto

Agentes con intereses mixtos

- Los incentivos de los agentes pueden estar contextualmente alineados o en conflicto
  - Durante la ejecución del sistema multiagente, podemos variar entre la cooperación y la competición
- Es el tipo de situación más interesante porque también es la más habitual en contextos reales y en entornos simulados complejos
- En estos casos buscaremos combinar los métodos que se suelen usar en cooperación y en competición
- Una posibilidad es encontrar particiones de los agentes que sean capaces de cooperar: **formación de coaliciones**



[8] (Dafoe et al)

Hay equilibrios de Nash, pero no suelen coincidir con los óptimos de Pareto

¿Por qué es necesaria la coordinación?

- Hay diversos motivos que pueden hacer necesaria la coordinación en un sistema multiagente:
    - La posible asimetría de las funciones de utilidad (preferencias) de los agentes
    - La imposibilidad de que los agentes puedan cumplir sus objetivos por sí mismos, por recursos, por tiempo o por capacidades
    - El mero hecho de que haya coordinación permite mejorar el rendimiento o la estabilidad del sistema multiagente
    - La presencia de recursos limitados
    - Para evitar el fallo del sistema, e.g. paradoja de Braess, tragedia del bien común (*tragedy of the commons*, Hardin 1968 & Ostrom 1990)…

https://www.americanscientist.org/article/garrett-james-hardin-dallas-1915-santa-barbara-2003