Slide 1:

AI-Powered Individualism and Relational AI

Hirokazu Shirado

Assistant Professor
Human-Computer Interaction Institute
School of Computer Science
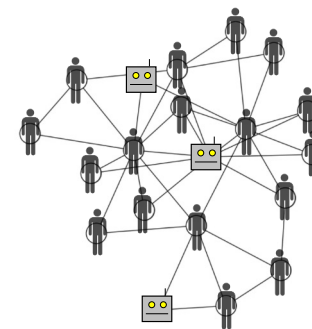Carnegie Mellon University

HCII • Human-Computer Interaction Institute

1

Slide 2:

2

Slide 3:

3

Slide 4:
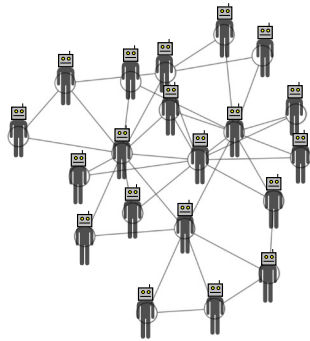
Hybrid Social Networks of Humans and Machines

Social coordination with noisy bots
(*Nature* 2017)

Cooperation with defriending bots
(*iScience* 2020)

Resource sharing with stingy bots
(*Scientific Reports* 2023)

4

Hybrid Actors of Humans and Machines

Driving coordination
between AI-assisted drivers
(*PNAS* 2023)

Voice communication
between AI-assisted speakers
(CHI 2025 Best Paper)

Intergroup cooperation
with LLM message suggestions
(CHI 2025 Best Paper)

5

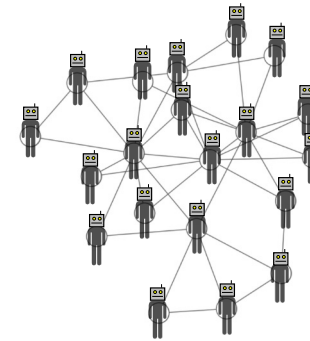## **Hybrid Actors** of Humans and Machines

**Driving coordination
between AI-assisted drivers
(*PNAS* 2023)**

Voice communication
between AI-assisted speakers
(*CHI* 2025 *Best Paper*)

Intergroup cooperation
with LLM message suggestions
(*CHI* 2025 *Best Paper*)

6

"Go Straight" or "Swerve"?

7

Chicken Game

		Your partner	
		Go straight	Swerve
You	Go straight	0	b
	Swerve	$b-c$	$b-c$

$$b > c$$

8

Chicken Game

		Your partner	
		Go straight	Swerve
You	Go straight	0 ↓	b ↑
	Swerve	b-c	b-c

$$b > c$$

Strategic Interdependence

9

Chicken game

		Your partner	
		Go straight	Swerve
You	Go straight	0 ↓	(b)
	Swerve	(b-c)	b-c

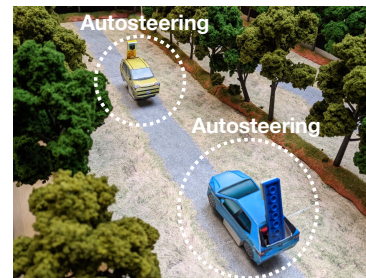**Strategic Interdependence** ⟶ **Alternating Reciprocity**

10

Autonomous Car Technology for Individual Safety

e.g., Nissan's autonomous emergency steering system
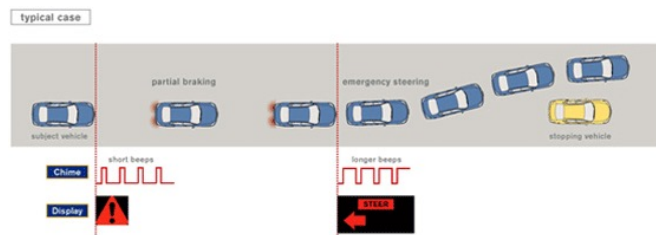
https://www.nissan-global.com/EN/INNOVATION/TECHNOLOGY/ARCHIVE/AUTONOMOUS_EMERGENCY_STEERING_SYSTEM/

11

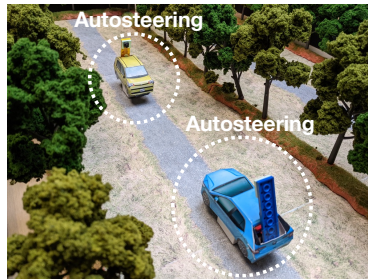## Chicken Game with **Safety Assistance**

		Your partner	
		Go straight	Swerve
You	Go straight	*b-c**	b
	Swerve	b-c	b-c

Strategic Interdependence

12

3

Slide 13

Chicken Game with **Safety Assistance**

Autosteering
Autosteering

		Your partner	
		Go straight	Swerve
You	Go straight	$b-c*$	b
	Swerve	$b-c$	$b-c$

$c > c*$

Strategic Interdependence

13

Slide 15

Miniature field

Robot   Robot

Two players interact with the first-person view over 10 rounds.
Each player earns more money when they reach the pink-taped goal earlier.
When they drive off-road, their driving speed decreases by 75%.

15

Slide 16

Chicken Game with **Safety Assistance**

Autosteering
Autosteering
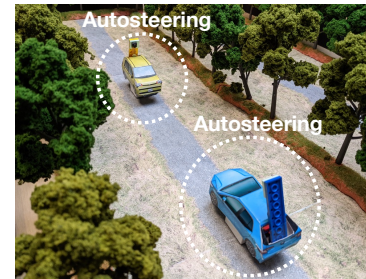
		Your partner	
		Go straight	Swerve
You	Go straight	$b-c*$	b
	Swerve	$b-c$	$b-c$

$c > c*$

Alternating Reciprocity
???

		$d1$	
		$d2$	
Manual	Warning		
Auto braking	Warning **Stop**		
Auto steering	Warning	**Steer**	

Game round 2 (Your current bonus: $0.34)
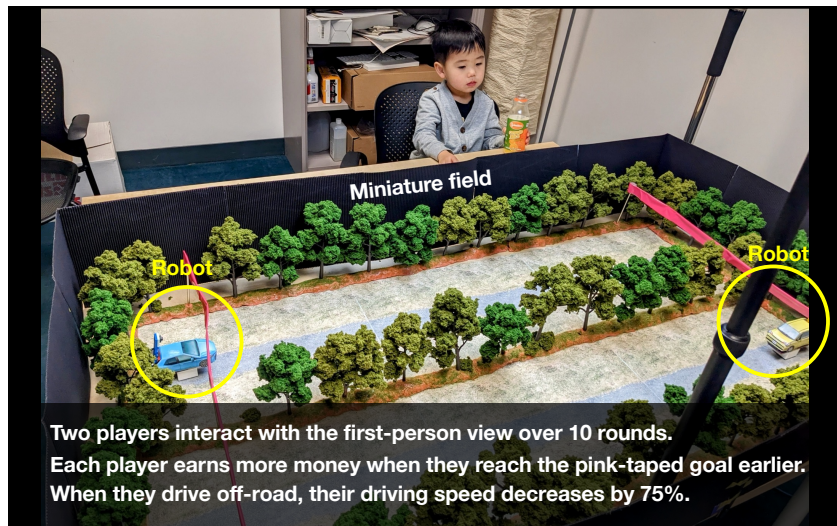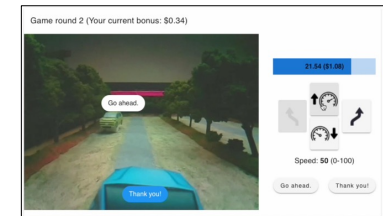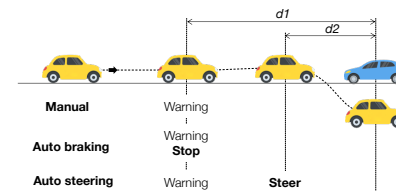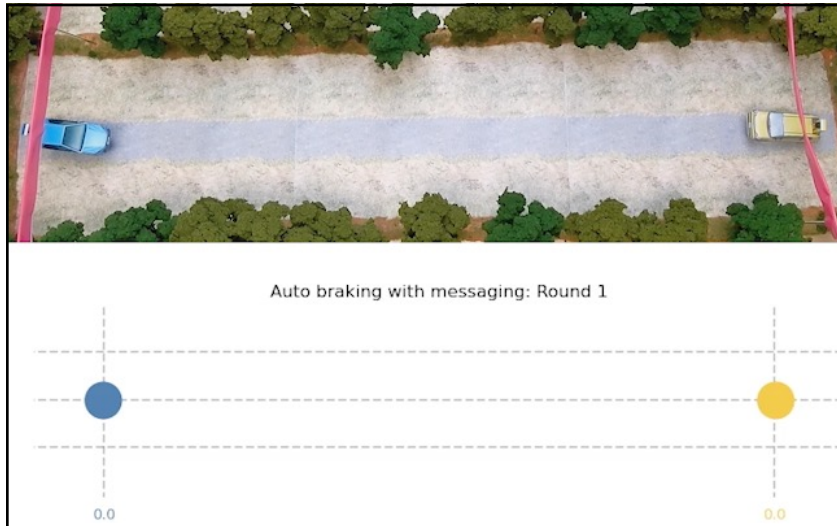
Active safety assistance   X   Messaging capability

16

Auto braking with messaging: Round 1

0.0   0.0

17

Manual	Alternating Reciprocity
Autosteering assistance	Both Chicken

Results: What facilitates/suppresses **reciprocity**?

Manual   Auto braking   Auto s...

Session

Messaging

1 2 3 4 5 6 7 8 9 10
Round

■ Unilateral turn by Blue car

■ Bilateral turn (Both chicken)

■ Crash
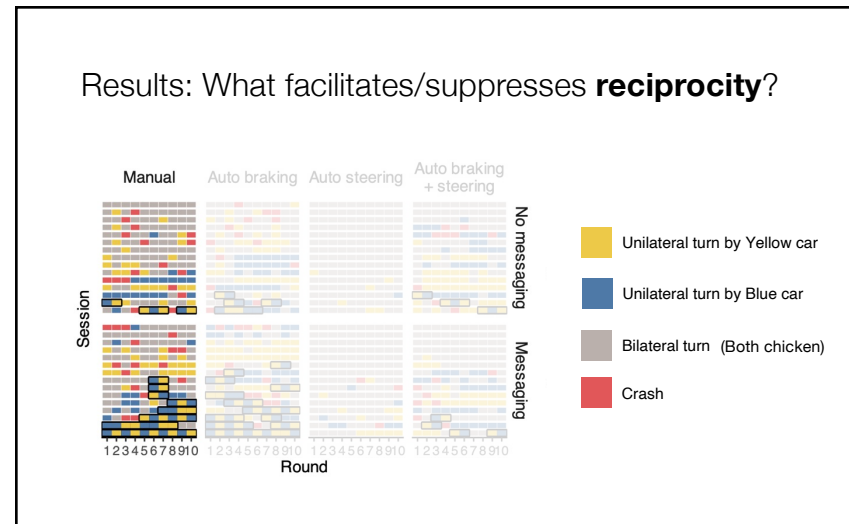
19

Results: What facilitates/suppresses **reciprocity**?

Manual   Auto braking   Auto s...

Session

Messaging

1 2 3 4 5 6 7 8 9 10
Round

■ Unilateral turn by Blue car

■ Bilateral turn (Both chicken)

■ Crash

20

Slide 21

Results: What facilitates/suppresses **reciprocity**?

Manual Auto braking Auto s...

Session

Round
1 2 3 4 5 6 7 8 9 10 1 2 3 4 5 6 7 8 9 10 1 2 3 4 5 6 7 8 9 10 1 2 3 4 5 6 7 8 9 10

saging

Messaging

■ Unilateral turn by Blue car

■ Bilateral turn (Both chicken)

■ Crash

21

Slide 22

Results: What facilitates/suppresses **reciprocity**?

Manual Auto braking Auto s...

Session

Round

saging

Messaging

■ Unilateral turn by Blue car

■ Bilateral turn (Both chicken)

■ Crash

Slide 23

Results: What facilitates/suppresses **reciprocity**?

Manual Auto braking Auto s...

Session

Round
1 2 3 4 5 6 7 8 9 10 1 2 3 4 5 6 7 8 9 10 1 2 3 4 5 6 7 8 9 10 1 2 3 4 5 6 7 8 9 10

saging

Messaging

■ Unilateral turn by Blue car

■ Bilateral turn (Both chicken)

■ Crash

23

Slide 24

Chicken Game with **Safety Assistance**

Auto braking + steering

Autosteering

Strategic interdependence

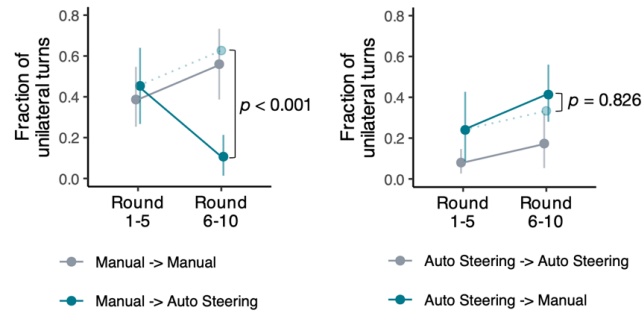|	Your partner	
	Go straight	Swerve
You Go straight	$b-c^*$ ↓	b ↑
Swerve	$b-c$	$b-c$

$c > c^*$

$c < c^*$

24

Egocentric Driving Persists After Safety Assistance Is Lifted.

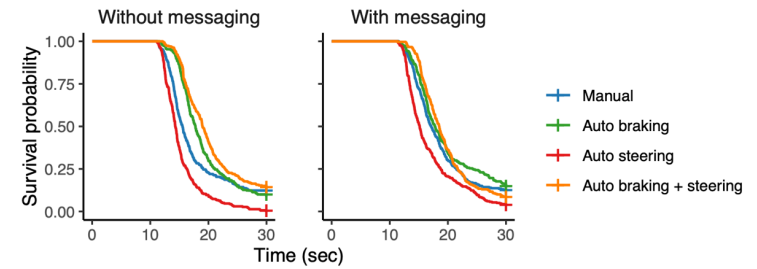What happens when assistance is activated or deactivated halfway?
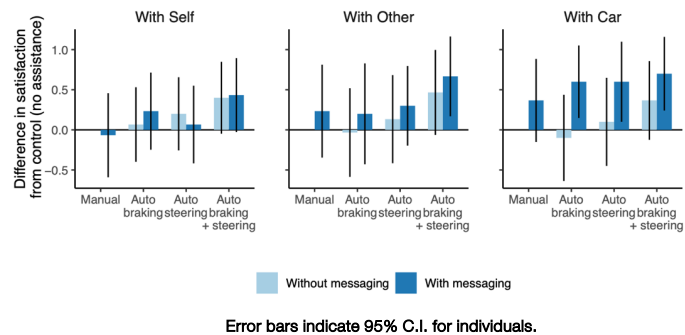

Results: Autosteering Helps Speed Up

Survival Probability:
Percentage of drives to reach the goal within the indicated time.

Results: No Difference in **Subjective Satisfaction**

Error bars indicate 95% C.I. for individuals.

AI-Powered Individualism

Humans: no willingness to coordinate

Hybrid actors: successful coordination

25

26

27

28

Slide 29

AI-Powered Individualism
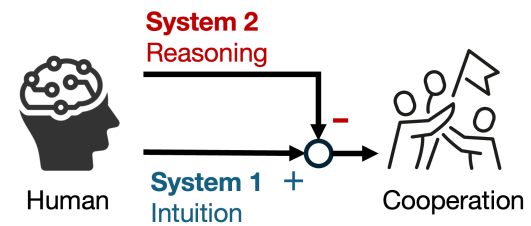

Humans: no willingness to coordinate

Hybrid actors: successful coordination

29

Slide 30

Dual-Process Hypothesis for Cooperation

System 2
Reasoning

−

System 1 +
Intuition

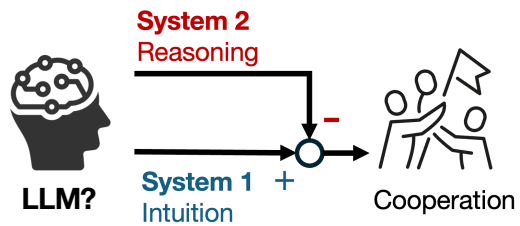Human   Cooperation

Rand, D. G., Greene, J. D. & Nowak, M. A. Spontaneous
giving and calculated greed. *Nature* **489**, 427–430 (2012).

30

Slide 31

Dual-Process Hypothesis for Cooperation

System 2
Reasoning

−

System 1 +
Intuition

LLM?   Cooperation

Li, Y. & Shirado, H. Spontaneous giving and calculated greed in language models.
arXiv [cs.CL] (2025) https://arxiv.org/abs/2502.17720.

31

Slide 32

Spontaneous Giving and Calculated Greed in LLM

Dictator Game · Prisoner's Dilemma Game · Public Goods Game

Cooperation / Proportion allocated / Cooperation rate

GPT-4o · o1 · $P < 0.001$

Cooperation   Defection

Li, Y. & Shirado, H. Spontaneous giving and calculated greed in language models.
arXiv [cs.CL] (2025) https://arxiv.org/abs/2502.17720.
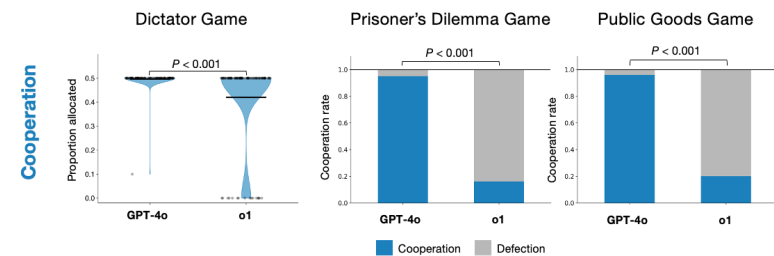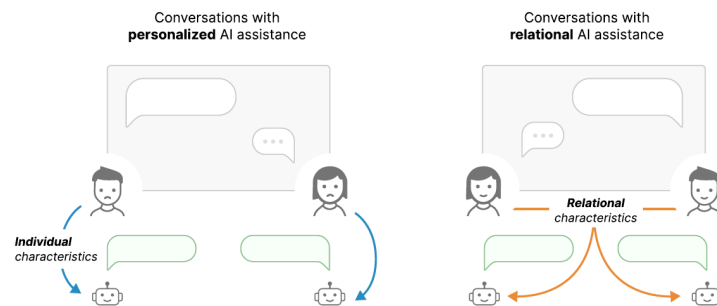
32

How can we address **AI-powered individualism**?

33

Relational AI:

Individuals (nodes) → Relations (edges)

34

Relational AI on Intergroup Communication

Conversations with
personalized AI assistance

Conversations with
relational AI assistance

*Individual
characteristics*

*Relational
characteristics*

Claggett, Kraut, and Shirado. "Relational AI: Facilitating Intergroup Cooperation with Socially Aware Conversational Support." *CHI* 2025 (Best Paper Award).

35

Experiment Procedure

(1) Pre-task survey

How do you feel about this statement? (3 / 7)
I support funding the military

| Strongly disagree | Disagree | Slightly disagree | Neutral | Slightly agree | Agree | Strongly agree |

(2) Pairing (in-group | out-group)

Your partner is from a
different group.

You have been grouped with millions of Americans who all agree with "I support funding the military"

Your partner has an opposing opinion.

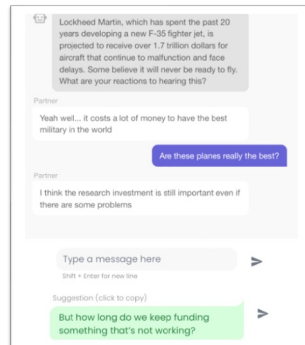Next, you will chat with your partner about an assigned topic.

Continue to Chat

An example of an out-group pairing

Claggett, Kraut, and Shirado. "Relational AI: Facilitating Intergroup Cooperation with Socially Aware Conversational Support." *CHI* 2025 (Best Paper Award).

36

Slide 37

Experiment Procedure

(3) Chatting with AI suggestion

37

Slide 38

Treatments: LLM message suggestions

(3) Chatting with AI suggestion
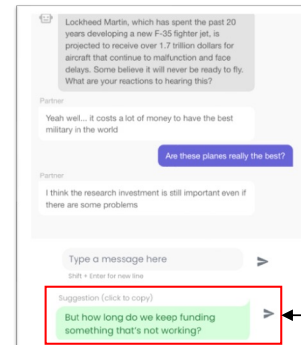
- No AI Assistance

- Personalized AI Assistance
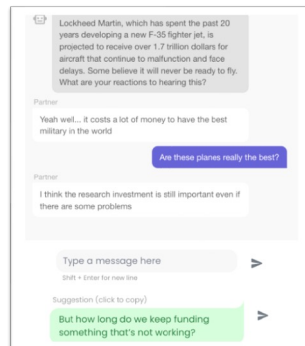Prompting GPT-4
with the user's survey answer

- Relational AI Assistance
Prompting GPT-4
with the pair context,
in-group or out-group,
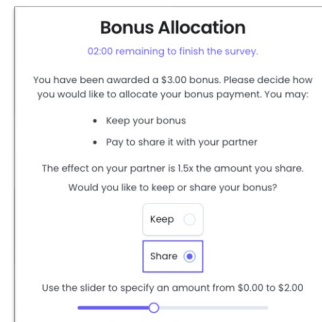based on the user's and partner's
survey answers

38

Slide 39

Evaluation: Prisoner's Dilemma Game
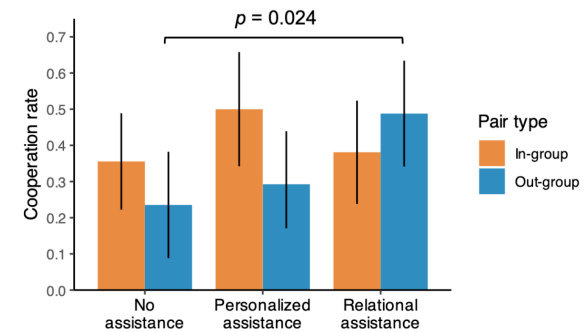
(3) Chatting with AI suggestion   **(4) Post-chat cooperation game**

Bonus Allocation

02:00 remaining to finish the survey.

You have been awarded a $3.00 bonus. Please decide how you would like to allocate your bonus payment. You may:

- Keep your bonus
- Pay to share it with your partner

The effect on your partner is 1.5x the amount you share. Would you like to keep or share your bonus?

Keep ○

Share ◉

Use the slider to specify an amount from $0.00 to $2.00
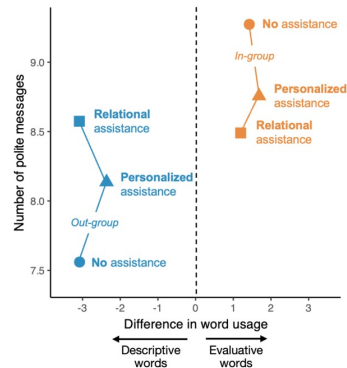
39

Slide 40

Results:
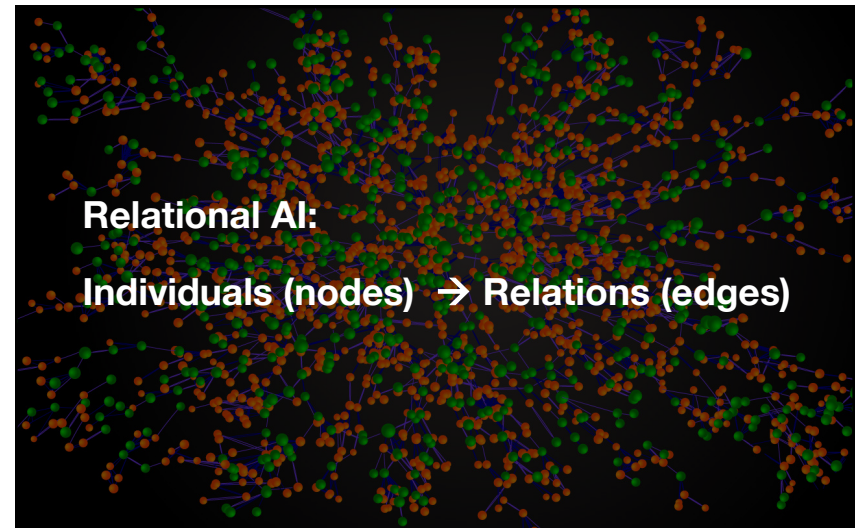Whether both choose cooperation after the pair chat

Claggett, Kraut, and Shirado. "Relational AI: Facilitating Intergroup Cooperation with Socially Aware Conversational Support." *CHI* 2025 (Best Paper Award).

40

Slide 41

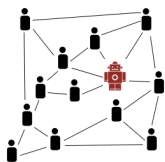Relational AI assistance doesn't change **what you say**, but **how you say it**.

Slide 42

Relational AI:

Individuals (nodes) → Relations (edges)

Slide 43

Contact:
shirado@cmu.edu
https://www.shirado.net/

References:
- Shirado, Kasahara, & Christakis. "Emergence and Collapse of Reciprocity in Semi-Automatic Driving Coordination Experiments with Humans." *PNAS*.(2023).
- Li & Shirado. "Spontaneous Giving and Calculated Greed in Language Models." *arXiv* (2025).
- Claggett, Kraut & Shirado. "Relational AI: Facilitating Intergroup Cooperation with Socially Aware Conversational Support." *CHI* (2025).

Thanks to:
Robert Wood Johnson Foundation
National Science Foundation
NOMIS Foundation