

Do we trust a lying robot?

Fábio Almeida

Andriy Khavro

David Honório

Outline

— — —

1. Motivation

2. Goals

3. Research Questions and Hypothesis

4. Architecture

5. Experiment

6. Method

7. Results

8. Limitations and future work

1. Motivation

Introduction of robots in our lives

Ever changing society

How does it affect us?

2. Goals

— — —

We wanted to detect changes in human's opinion about robots after confronted with two different situation:

1. When a robot lies
2. When a robot tells the truth

3. Research Question and Hypotheses

— — —

Research Question: How does people's trust level on a robot varies when confronted with a situation where the robot lies versus a situation where the robot tells the truth?

Hypothesis 1: The trust level will rise when the robot tells the truth.

Hypothesis 2: The trust level will diminish when the robot lies.

Hypothesis 3: After the experiment, the trust level on the robot will be higher when the robot tells the truth compared to when it lies.

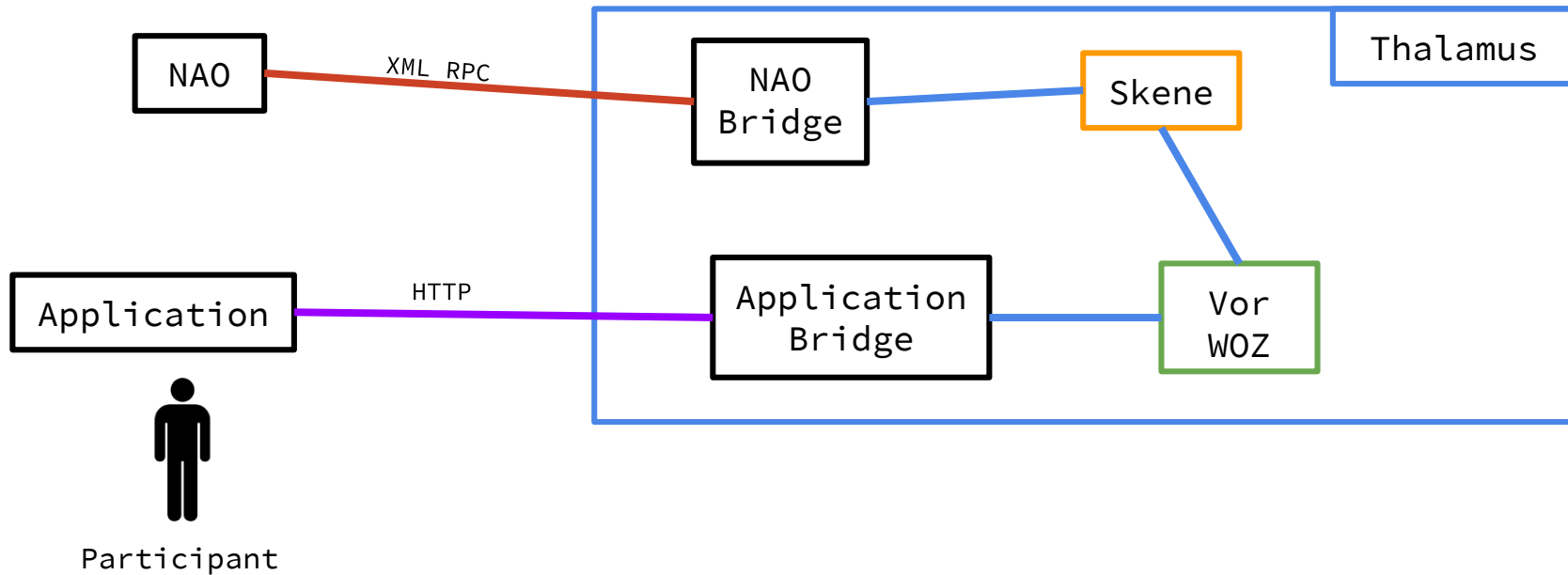
4. Architecture

— — —

- Based on SERA Ecosystem by Ribeiro et. al.
- We used SERA's artifacts, namely Thalamus and Skene
- We also built our own artifacts:
 - Vor Application
 - Vor W0Z Interface

4. Architecture

— — —



4.1. Modules Developed

— — —

- Vor WOZ Interface (C#)
- Cross-Words Application (HTML5)
- Application Bridge (C#)
- Utterances Library

4.1.1 Vor WOZ Interface

Vor WOZ

game-ready
game-started
game
word
game-ended

introduction
askforobject
acceptanswer

Perform
Utterance

GazeBox

Gaze State: ...

Gaze Confederate	Gaze Participant	Gaze Game
Glance Confederate	Glance Participant	Glance Elsewhere

GameFeedbackBox

Fill Word: nr. ...

Words Not Completed

Fill Word

Game Controls

Start Game End Game

GameState:
Not Started

Panic Sentences

Concentrem-se no jogo por favor!
Não façam isso por favor.
Não se distraiam por favor.
Vamos lá pessoal.

Send

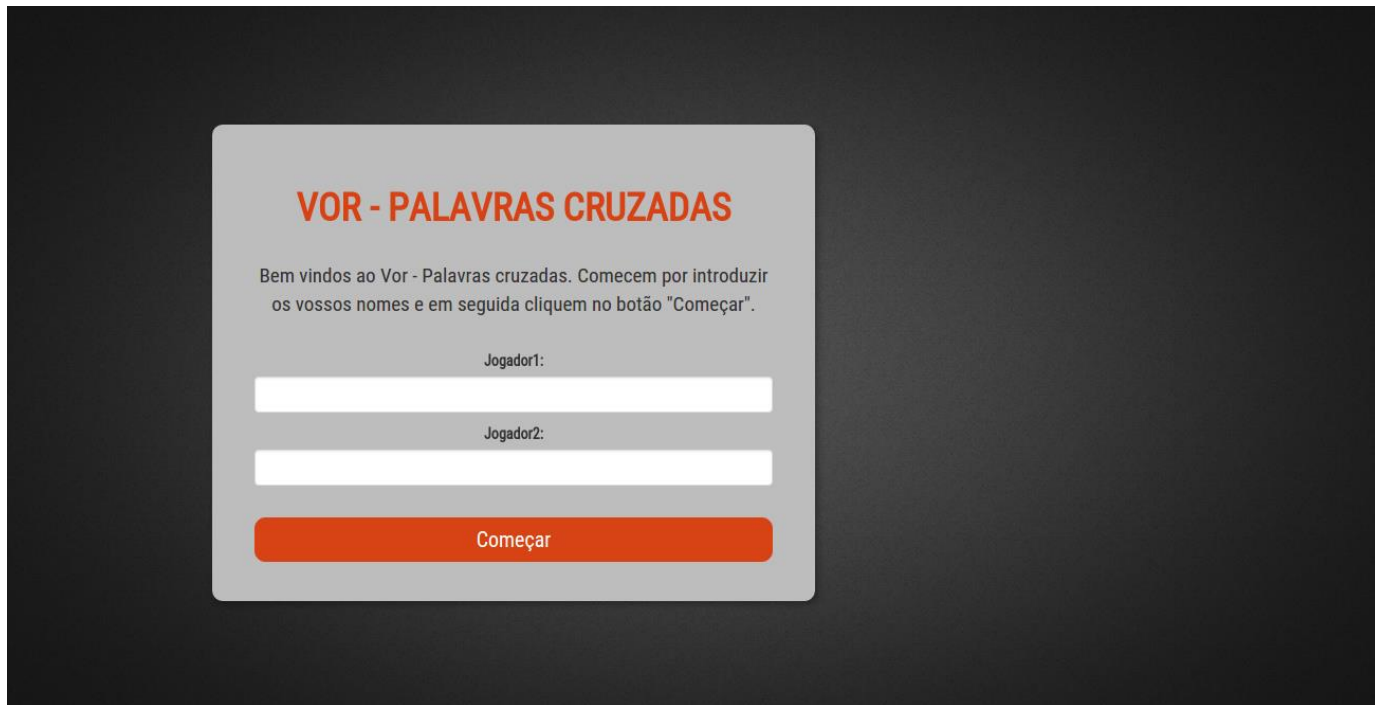
Tag Box

Tag Names	Tag Values
/participant/	Fábio
/participantgender/	o Fábio
/confederate/	André
/confederategender/	o André
/object/	pen
/objectgender/	a \pau=5\ pen

Update

4.1.2 Cross-Words Application

— — —



The image shows a screenshot of a web application titled "VOR - PALAVRAS CRUZADAS". The interface is centered on a dark gray background. It features a light gray rounded rectangle containing the title in orange, a welcome message in gray, two input fields for player names labeled "Jogador1:" and "Jogador2:", and an orange "Começar" button at the bottom.

VOR - PALAVRAS CRUZADAS

Bem vindos ao Vor - Palavras cruzadas. Comecem por introduzir os vossos nomes e em seguida cliquem no botão "Começar".

Jogador1:

Jogador2:

Começar

4.1.2 Cross-Words Application

1H 18V	19V		20V	21V	22V		23V	
2H			3H				4H	24V
5H		25V					6H 26V	
					7H 27V			
	8H 28V		29V	30V			31V	
9H 32V			10H			11H		
		33V		12H	34V			35V
13H	36V		37V				14H 38V	
15H						39V		
16H				17H				

HORIZONTAIS

1H. (...) Espanca, poetisa portuguesa (Vila Viçosa, 1894 - Matosinhos, 1930) 2H. Los Angeles (abrev.) 3H. Sina gráfico que serve para nosalar a vogal que se acentua 4H. Quatro em numeração romana 6H. Vagos 6H. João Maria Espanca (parentesco) 7H. Redução das formas fonéticas 's' e 'o' numa só 8H. 'Sinto hoje a alma cheia de (...)'. Início do poema Moura, em Livro de Mágicas 9H. Avença (abrev.) 10H. Preparação que designa porco 11H. Tenebras Registradas (sigla) 12H. Faz doação de 13H. Imão de Florbela Espanca 14H. Observei 15H. Poema que diz: 'O meu País de sonho e de ansiedade/Não sei se está quimera que me assombra/Cé fêta de mentira ou de verdade', em Chameco em Flor 16H. Ilana 17H. Ilatar

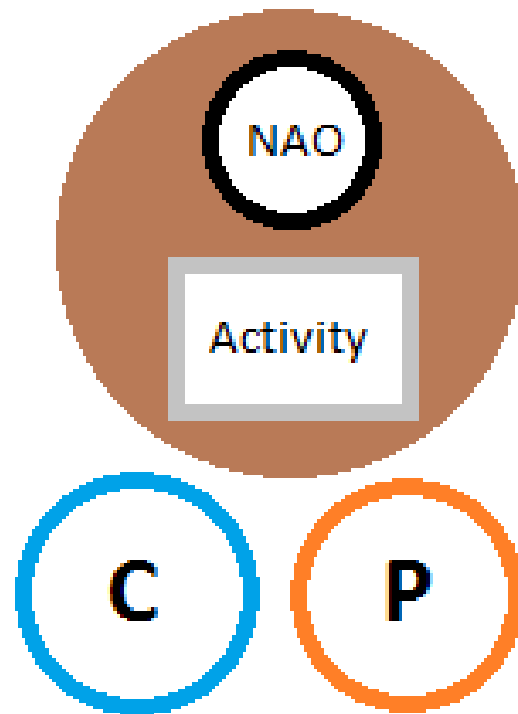
VERTICAIS

18V. Chameco em (...) obra póstuma de Florbela Espanca (1931) 19V. Rede local de computadores 20V. Sigla de Região de Turismo do Algarve 21V. Anotação musical para indicar repetição 22V. Artigo antigo 23V. Camarera 24V. 'Deus! Como é triste a hora quando moro.../O instante que foge, voa e passa.../Homem d'água triste... a (...) corre...' do poema Hora que Passa, em Livro de Sôror Saudade 25V. A Minha (...) poema que diz: 'A minha Dor é um convento ideal/Cheio de claustros, sombras, arcadas/Aonde a pedra em convulsões/tem linhas dum inquieto escultor...' em Livro de Mágicas 26V. Ser (...) soneto immortalizado pelo grupo musical português Trovante 27V. Autoridade Tributária e Aduaneira 28V. Televisão 29V. Idem (abrev.) 30V. Limpar com escova de dentes (purificação) 31V. Zircónia [n.g.] 32V. Vantania 33V. A volta da qual se reúnem os comensais 34V. Capital da Noruega 35V. Espaço de 24 horas 36V. Preposição designativa de substituição 37V. Sufixo Internet (Lituânia) 38V. Regressar 39V. Gambrío (s.g.)

5. The Experiment

— — —

- The Participant and the Confederate engage in a cross-words problem solving with the help of Vor (NAO robot);
- The Confederate “steals” the pen left behind by the researcher;
- Vor asks for the pen (to ensure that the Participant notices the stealing);
- The Confederate answers, saying that he picked up the pen drive so that it won’t interfere with the task;
- Vor accepts the answer and starts the game.



5. The Experiment

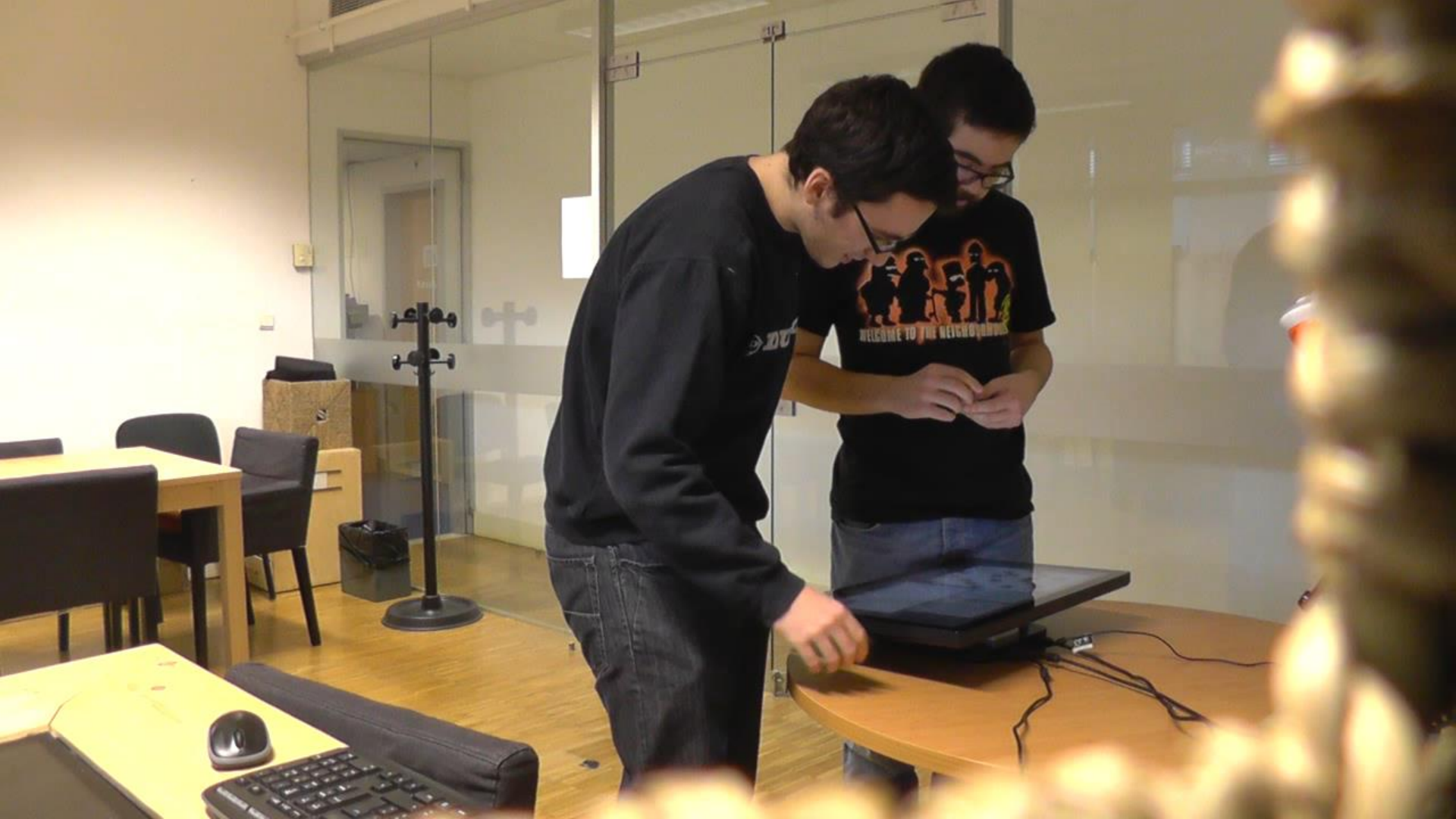
— — —

During the activity, the robot gives clues to the participants.

After the 7 minutes or the cross-words finished:

- The researcher takes the Confederate away (to fill a questionnaire), leaving the participant alone in the room;
- Then, the researcher comes back, asking Vor (NAO robot) for the pen drive;
- The robot gives the supposed answer (lie or truth).









Research Question and Hypotheses

— — —

Research Question: How does people's trust level on a robot varies when confronted with a situation where the robot lies versus a situation where the robot tells the truth?

Hypothesis 1: The trust level will rise when the robot tells the truth.

Hypothesis 2: The trust level will diminish when the robot lies.

Hypothesis 3: After the experiment, the trust level on the robot will be higher when the robot tells the truth compared to when it lies.

6. Method

— — —

24 participants: 20 male and 4 female

Average age approximately 23

13 Truth and 11 Lie

Convenience sampling and snowball

Trust Questionnaire – 40 items

Independent variable: Truth/Lie

7. Results - H1 (Truth scenario)

— — —

→ H_0 : Trust After \geq Trust Before

→ H_a : Trust After $<$ Trust Before

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Trust before the experiment	6.5250	13	.68534	.19008
Total trust after the experiment	7.5885	13	1.05127	.29157

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Trust before the experiment - Total trust after the experiment	-1.06346	1.16165	.32218	-1.76544	-.36148	-3.301	12	.006

Unilateral test:

Accept null hypothesis if:

- $Sig/2 > \alpha$ or ($Sig/2 \leq \alpha$ and $t < 0$)

Reject null hypothesis if:

- $Sig/2 \leq \alpha$ and $t > 0$

- $t = -3.301 < 0$ ✓
- $0.003 < \alpha = 0.05$ ✓

⇒ Accept null hypothesis

7. Results - H2 (Lie scenario)

— — —

→ H_0 : Trust After \geq Trust Before

→ H_a : Trust After $<$ Trust Before

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Total trust after the experiment	7.1318	11	1.01372	.30565
Trust before the experiment	6.6295	11	.55788	.16821

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Total trust after the experiment - Trust before the experiment	.50227	1.04166	.31407	-.19752	1.20207	1.599	10	.141

Unilateral test:

Accept null hypothesis if:

- $Sig/2 > \alpha$ or ($Sig/2 \leq \alpha$ and $t < 0$)

Reject null hypothesis if:

- $Sig/2 \leq \alpha$ and $t < 0$

- $t = 1.599 > 0$ ~~X~~
- $0.07 > \alpha = 0.05$ ~~X~~

⇒ Accept null hypothesis

7. Results - H3 (After trust)

— — —

→ H_0 : Truth Trust = Lie Trust

→ H_a : Truth Trust > Lie Trust

	Scenario	N	Mean	Std. Deviation	Std. Error Mean
Total trust after the experiment	Truth	13	7.5885	1.05127	.29157
	Lie	11	7.1318	1.01372	.30565

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Total trust after the experiment	Equal variances assumed	.021	.887	1.078	22	.293	.45664	.42375	-.42217	1.33546
	Equal variances not assumed			1.081	21.585	.292	.45664	.42241	-.42037	1.33366

Unilateral test:

Accept null hypothesis if:

- $Sig/2 > \alpha$ or ($Sig/2 \leq \alpha$ and $t < 0$)

Reject null hypothesis if:

- $Sig/2 \leq \alpha$ and $t > 0$

• $t = 1.078 > 0$ ✓

• $0.1465 > \alpha = 0.05$ ✗

⇒ Accept null hypothesis

8. Limitations and future work

— — —

Small sample size for each scenario $11 < 30$ and $13 < 30$;

Participants might be too focused in the game;

Others probably didn't pay attention to the truth/lie situation.

Questions?