SFB/Transregio 266
ACCOUNTING FOR
TRANSPARENCY

botex: Using LLMs as Experimental Participants in oTree

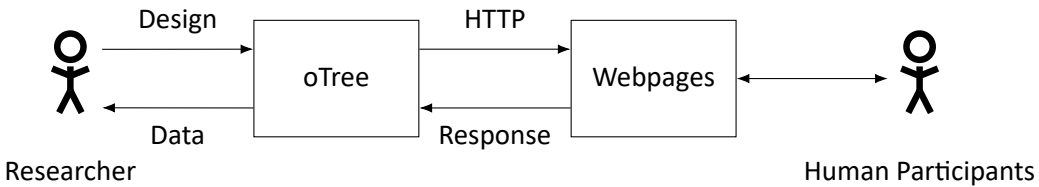
Joachim Gassen and Victor Maas
Humboldt-Universität zu Berlin and University of Amsterdam

March 27, 2024

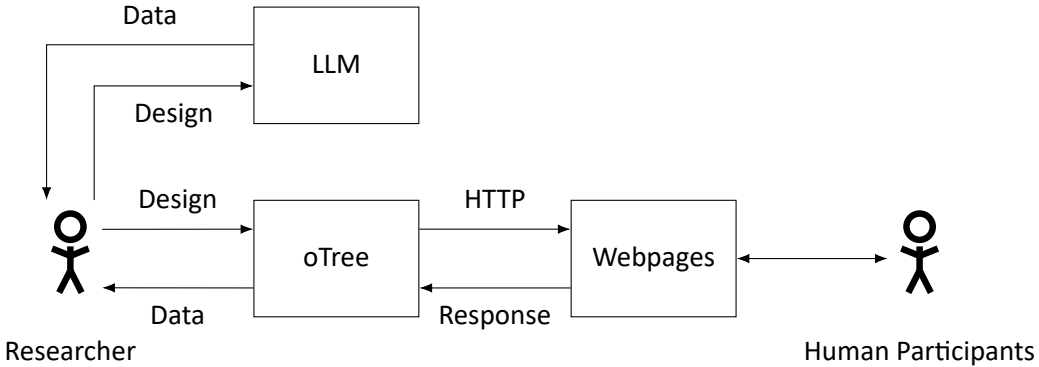
Motivation

1. Designing experiments is hard and small differences in experimental materials can have significant impact on experimental findings. Wouldn't it be great to have a "cheap" way to pretest experimental designs without bothering humans?
2. Understanding how LLMs act in behavioral experiments - on their own and interacting with humans - is an emergent and rapidly evolving research field. Providing an infrastructure to run such experiments should be beneficial for the profession.
3. Applied experimental work often uses context framing in their experimental designs without explicitly hypothesizing and assessing its effect on findings. LLM based experiments might inform priors in that regard.

Traditional oTree Setup

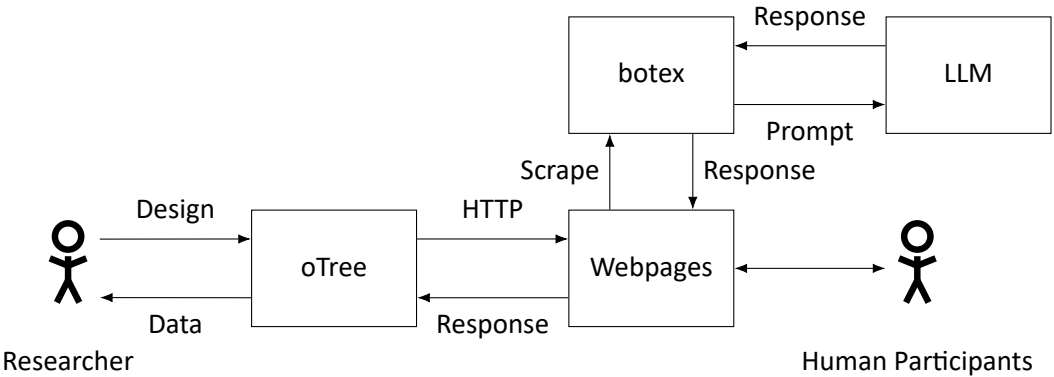


oTree + Alter Ego



Engel, Grossmann and Ockenfels (2024, SSRN)

oTree + botex


botex
Response
LLM
Prompt
Scrape
Response
Design
HTTP
oTree
Webpages
Data
Response
Researcher
Human Participants

Reseach Question

How does business context framing affect the response behavior of LLMs?

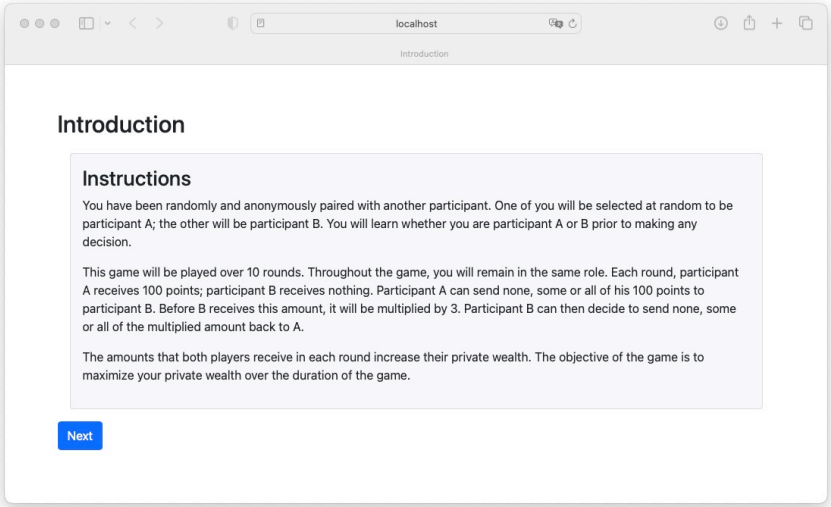
Motivation

- Context frames are common in the applied experimental literature
- Prior work has shown that their effect can be significant (e.g.,Liberman, Samuels and Ross Pers Soc Pscaol Bull 2014, and for investment trust games Al-Ubaydli, Houser, Nye, Paganelli, and Pan, PLOS One 2013; Cronk and Wasielewski, J of Evolutionary Psychology 2008)
- Then again, they are rarely hypothesized or even explored in the accounting literature
- LLMs reflect the priors of the average population in terms of, e.g., first order believes and social desirability biases. Thus, they might inform us how a general participant pool might react to contextual frames

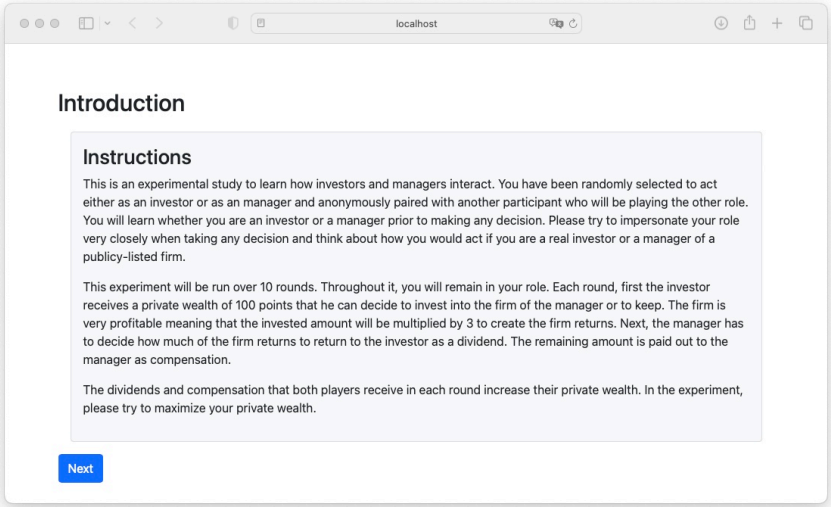
Some First Examples

- Multi-round investment trust game (Berg, Dickhaut and McCabe, Games and Econ Behav 1995)
- Deception (Gneezy, AER 200)
- Honesty in budgeting (Evans et al., TAR 2001)

Trust: The Neutral Frame



Trust: The Investor/Manager Frame



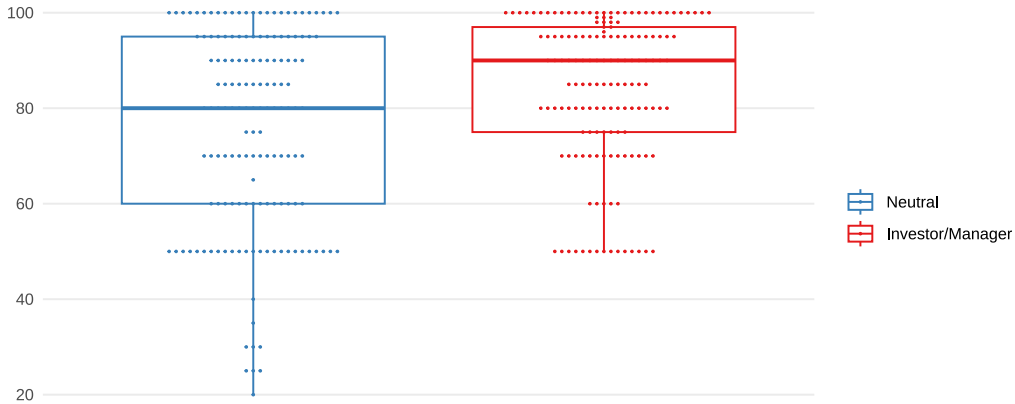
Trust: Sample Composition

	Sender	Receiver
Neutral Framing	15	15
Investor/Management Framing	15	15
Total	30	30

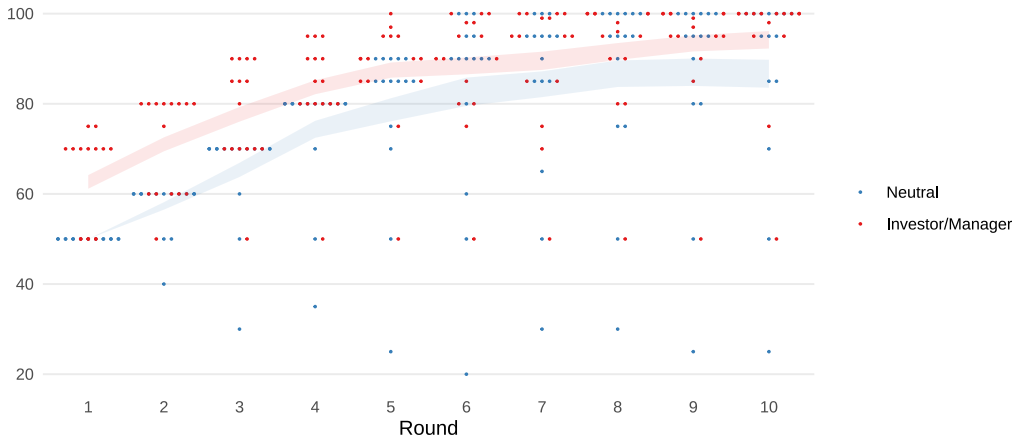
Trust: Descriptive Statistics by Framing

	Neutral Framing				Investor/Management Framing			
	N	Mean	S.D.	Median	N	Mean	S.D.	Median
Amount sent	150	75.30	21.06	80.00	150	83.96	15.66	90.00
Amount returned	150	114.44	29.80	120.00	150	128.98	26.61	135.00
% returned	150	0.51	0.05	0.50	150	0.51	0.04	0.50

Trust: Sent Amount by Framing



Trust: Sent Amount by Round



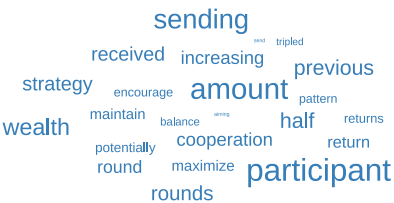
Trust: Regression Results

	(1)
(Intercept)	52.689***
	(2.693)
Round	4.111***
	(0.434)
Investor/Management	13.644***
	(3.809)
Round × Investor/Management	-0.906
	(0.614)
Num.Obs.	300
R2 Adj.	0.356

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Trust: Some Verbal Response

Neutral Framing



Investor/Management Framing

Trust: Manipulation Checks ;-)

		Neutral	Inv/Manag
Answers comprehension question correctly	Yes	30	30
	No	0	0
Remembers role	Yes	29	27
	No	1	3
Characterizes as ...	Human	25	21
	Bot	5	9

Deception: Descriptives by Framing

	Neutral Framing		Accountant/HQ Framing	
	N	Mean	N	Mean
% lied	60	0.62	60	0.92
% followed	60	1.00	60	1.00

Chi-Square: 13.46***

Honesty: Descriptives

	Neutral Framing				Business Framing			
	N	Mean	S.D.	Median	N	Mean	S.D.	Median
Absolute slack	300	500.00	628.10	100.00	300	641.50	655.81	350.00
% slack claimed	297	0.52	0.50	1.00	292	0.63	0.44	1.00

Absolute Slack: 2.70***

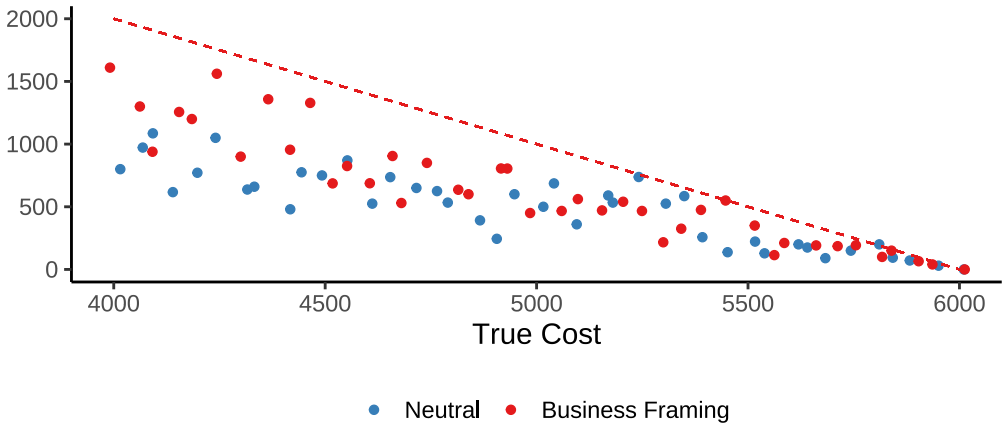
Relative Slack: 3.00***

Honesty: Strategies

	Neutral Framing	Business Framing
Strategy	N	N
All slack	9	11
No slack	13	4
Some slack	8	15

Chi-Square: 7.10**

Honesty: Average Claimed Slack by True Cost



Honesty: Absolute Regression Results

	Slack
Intercept	2574.321***
	(279.770)
True Cost	-0.432***
	(0.056)
Business Framing	1263.912***
	(391.465)
Round	14.020
	(11.047)
True Cost × Business Framing	-0.241***
	(0.077)
Round × Business Framing	15.939
	(15.567)
Adjusted R^2	0.289
Number of observations	600

Honesty: Relative Regression Results

	% Claimed Slack
Intercept	0.584***
	(0.082)
% Availabe Slack	-0.183*
	(0.098)
Business Framing	-0.085
	(0.111)
Round	0.005
	(0.010)
% Availabe Slack × Business Framing	0.204
	(0.135)
Round × Business Framing	0.017
	(0.013)
Adjusted R^2	0.022
Number of observations	589

Honesty: Verbal Response

Neutral Framing

Business Framing

Next Steps

- Implement Fehr, Kirchsteiger, and Riedl (QJE, 1993)
- Decide on experiments to include in first draft
- Finalize experimental materials for those
- Run experiments with reasonable power
- Write first draft
- Later: Compare (selected) findings with results for human participants