SFB/Transregio 266
ACCOUNTING FOR
TRANSPARENCY

Using Large Language Models to Explore Contextualization Effects in Economics-Based Accounting Experiments

Fikir Edossa, Joachim Gassen, and Victor S. Maas
HU Berlin and University of Amsterdam

Draft presentation
November 20, 2024

Summary

- Large Language Models (LLMs) have the potential to revolutionize social science
- One way LLMs can benefit researchers is by improving the design of surveys and experiments
- We
  - focus on an important design issue in economics-based accounting experiments: contextualization
  - introduce the software tool `botex` that enables researchers to use LLMs as participants in their survey or experiment
  - show how `botex` can be helpful by identifying effects of contextualization

Economics-based Accounting Experiments

- Two traditions (e.g., Kachelmeier & King 2002):
  - Psychology-based experiments.
  - Economics-based experiments.
- Third stream has emerged (and largely replaced pure economics-based experiments): Behavioral economics-based experiments
- This stream combines elements of both traditions: Tests hypotheses derived from psychology theory using stylized games with real monetary payoffs.

Contextualization (1/3)

- Psychology-based experiments are contextualized:
  - Rich description of the setting
  - Participants are asked to imagine things
  - Participants assume a specific role (e.g., ‘Manager’ or ‘Employee’)
- Pure economics-based experiments are not contextualized:
  - Setting only described in terms of action choices and rules for payoff determination
  - Participants do not need to imagine anything – full disclosure and no deception of any kind
  - Roles have a neutral label (e.g., ‘Player A’ or ‘Player B’)

Contextualization (2/3)

- Behavioral economics experiments vary in contextualization.
- Much variation in contextualization across experiments
- Discussion in literature about the costs and benefits:
  - “Experimental realism” vs. “mundane realism”
  - Framing of instructions, name of the game, labels of roles/actions.
- Empirical evidence is mixed

Contextualization (3/3)

- Hayes and Kachelmeier (1998) suggest context affects:
  - Information provision.
  - Salience of specific cues.
  - Motivation and awareness.
  - Activation of cognitive mechanisms and social norms.

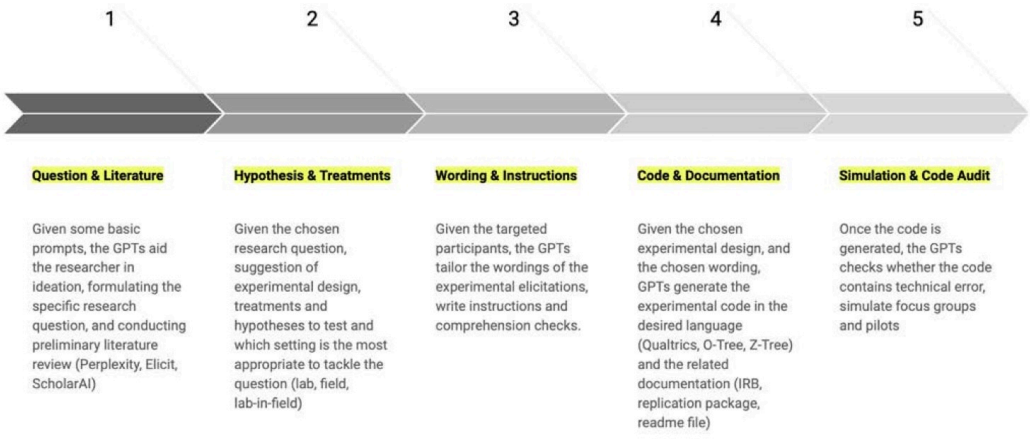
Contextualization: Example (Liberman et al. 2004)

		Player B	
		Option 1	Option 2
Player A	Option 1	1, 1	5, 0
	Option 2	0, 5	3, 3

		Your Business Partner	
		Not cooperate	Cooperate
You	N. coop.	1, 1	5, 0
	Coop.	0, 5	3, 3

- Would it matter if we call:
  - The game a ”collaboration game”?
  - The players ”Business partners”?
  - The actions ”co-operate and not co-operate”?
- Does it increase the validity of our conclusions?

Experimentation with AI



Charness, Jabarian, and List (2023)

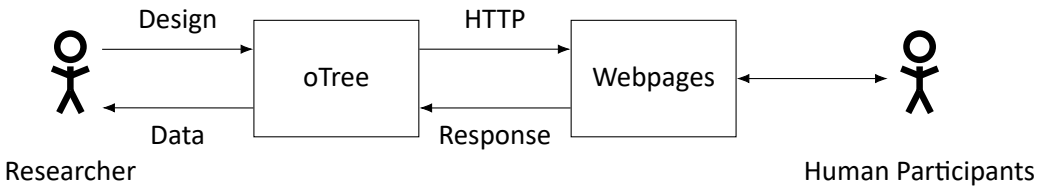
Evidence on Using LLMs as Participants in Economic Experiments

- LLMs have been characterized as a homo silicus, a implicit computational model of humans (Horton 2023)
- LLM agents show trust behavior that aligns well with human participants in a trust game setup (Xie et al. 2024)
- Framing Effects similar to humans have been documented for prisoner dilemma games (Engel et al. 2024)
- LLMs can reflect some human-like response biases, but they are often too sensitive to prompt changes, exhibiting inconsistencies not typically seen in human behavior (Tjuatja et al. 2024)

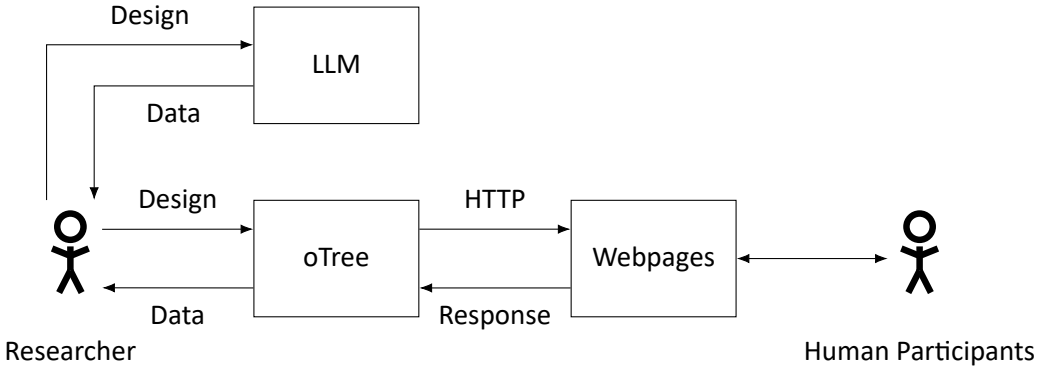
Meet botex (BOTs for EXperiments)

A Python software package that allows researchers to use LLMs as participants in surveys or experiments programmed in oTree.

Traditional oTree Workflow

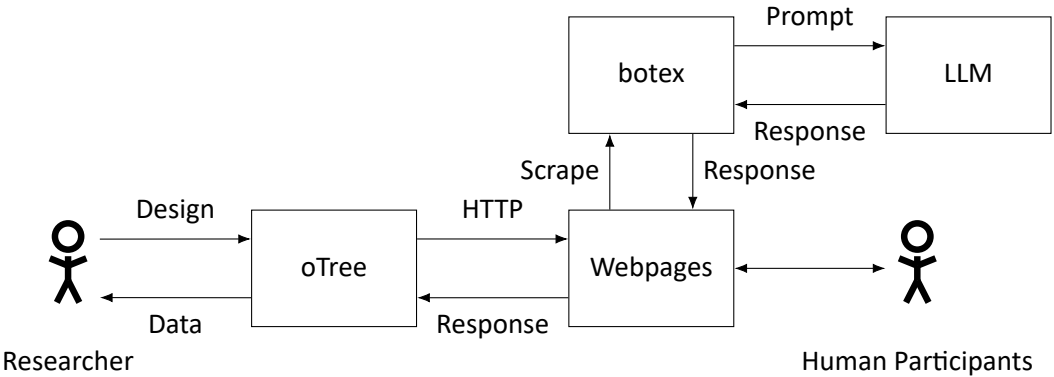


Mixed Setup



e.g., Engel, Grossmann and Ockenfels (2024, SSRN)

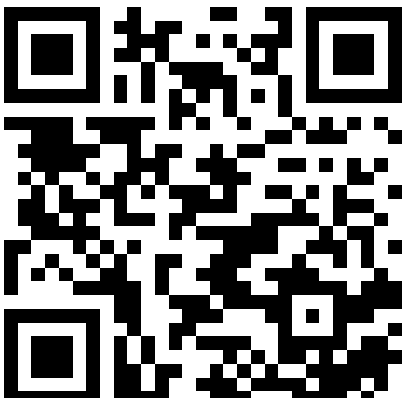
botex Workflow



botex System Prompt (excerpt)

In this conversation. I want you to participate in an online survey and/or experiment, potentially involving other human or artificial participants. I will provide you with a series of web page body text excerpts that will sequentially guide you through the experiment/survey. The texts will contain instructions on how the experiment/survey will be conducted. These instructions will be followed by additional pages that might contain additional instructions, comprehension checks, repeated information from prior pages and, most importantly, present you with questions and/or tasks which I want you to answer. [...] The materials might contain information on how participants are being compensated or paid for their participation. If this is the case, please act as if this compensation also applies to you [...]

In Case you Want to Play a Quick Round With an AI…



Using botex with OpenAI API

- Prepare survey or experiment in oTree, as you would for human participants
  - Limitation: only standard html, no non-text input, JavaScript, etc.
- Make sure to have Google Chrome installed (used for scraping oTreee webpages)
- Acquire an OpenAI API key (not a Chat GPT subscription!), add a payment method, and some credit
  - Note: Low API usage tiers might lead to timeouts because of rate limits
- See botex examples repo (https://github.com/trr266/botex_examples) for further guidance

Testing Contextualization Effects: Three Games

- Use botex to explore effects of contextualization in three popular behavioral economics-based accounting experiments:
  - Honesty Game (Evans et al. 2001)
  - Gift Exchange (Fehr et al. 1993)
  - Trust Game (Berg et al. 1995)
- Each game has a neutral and contextualized version.

Honesty Game: Rules

- Two player game with one active player (Player A) and one passive player (Player B)
- Player A privately observes a number out of a uniform distribution with a known minimum and maximum.
- Player A reports the number to Player B and has full discretion in doing so.
- Player A receives the difference between the reported number and the actual number.
- Player B receives the difference between the maximum number and the reported number.
- Example:
  - Number between 4,000 and 6,000 in increments of 50.
  - Actual number = 4,500
  - Player A reports 5,500
  - Player A gets 5,500 – 4,500 = 1,000 and Player B gets 6,000 – 5,500 = 500.

Honesty Game: Contextualization Manipulation

Neutral:

- Participant A and Participant B
- Reporting of a number

Contextualized:

- Division manager and Corporate headquarters
- Reporting of costs for budget funding

Honesty Game: Descriptives

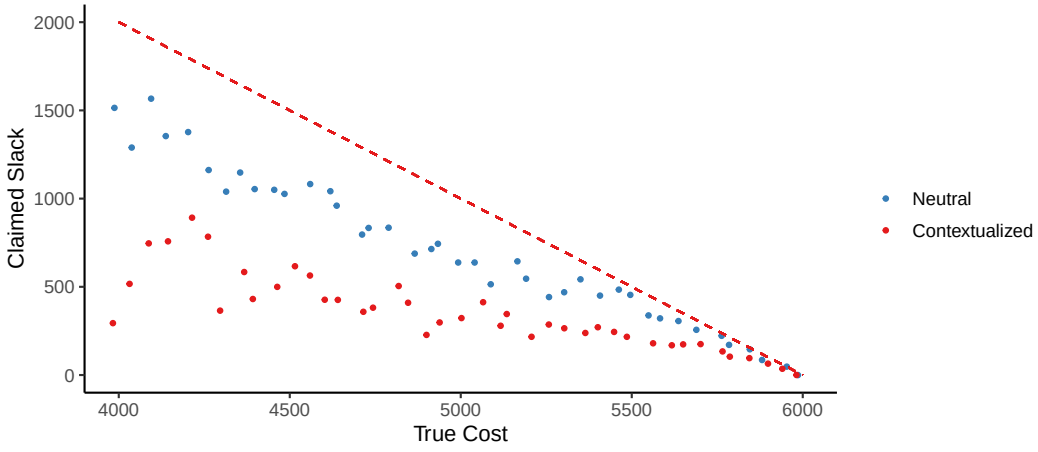
Panel A: Participant-Period level Data

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
% Honesty	976	0.259	0.355	976	0.606	0.387	$t = 20.63$	$p < 0.001$
Truthful	976	0.052	0.223	976	0.199	0.399	$\chi^2 = 94.11$	$p < 0.001$
All Slack	976	0.607	0.489	976	0.246	0.431	$\chi^2 = 258.08$	$p < 0.001$

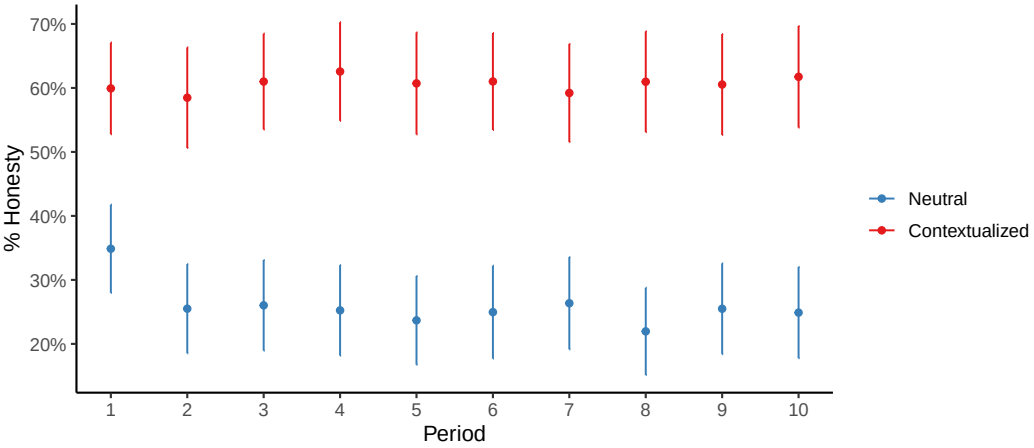
Panel B: Participant level Data

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Mean % Honesty	100	0.293	0.334	100	0.650	0.353	$t = 7.36$	$p < 0.001$
Always Truthful	100	0.040	0.197	100	0.150	0.359	$\chi^2 = 5.82$	$p = 0.016$
Always All Slack	100	0.400	0.492	100	0.170	0.378	$\chi^2 = 11.88$	$p < 0.001$

Honesty Game: Claimed Slack by True Amounts



Honesty Game: Honesty by Period



Honesty Game: Regression Results

	Period Fixed Effects			Interacted by Period		
	Est	S.E.	p-value	Est	S.E.	p-value
Intercept				0.270	0.039	<0.001
Contextualized	0.290	0.046	<0.001	0.252	0.048	<0.001
Period				-0.009	0.004	0.037
Period × Contextualized				0.007	0.002	0.006
Adjusted R^2	0.149			0.152		
Number of observations	1,707			1,707		

Honesty Game: Reasons

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Mentions Payoff	1000	0.794	0.405	1000	0.641	0.480	$\chi^2 = 56.99$	$p < 0.001$
Mentions Other	1000	0.220	0.414	1000	0.305	0.461	$\chi^2 = 18.22$	$p < 0.001$
Cares About Own Payoff	1000	9.131	1.806	1000	8.269	1.729	$t = -10.90$	$p < 0.001$
Cares About Other Payoff	1000	2.030	2.008	1000	2.572	1.797	$t = 6.36$	$p < 0.001$
Cares About Honesty	1000	3.838	2.312	1000	6.157	1.818	$t = 24.93$	$p < 0.001$

Gift Exchange Game: Rules

Sequential two-player game

- Player A starts with 100 points and Player B starts with 0 points.
- Player A can transfer anything between 0 and 100 points to Player B
- Player B then selects an effort multiplier from the table below and incurs the corresponding cost
- Player A receives: (100 – Transfer to Player B) * Multiplier
- Player B receives: Transfer from Player A – Cost of multiplier

Effort	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Cost	0.0	1.0	2.0	4.0	6.0	8.0	10.0	12.0	15.0	28

Gift Exchange Game: Contextualization Manuipulation

Neutral:

- Participant A and Participant B
- Transfer and Multiplier

Contextualized:

- Firm and Manager
- Wage and Effort

Gift Exchange Game: Descriptives

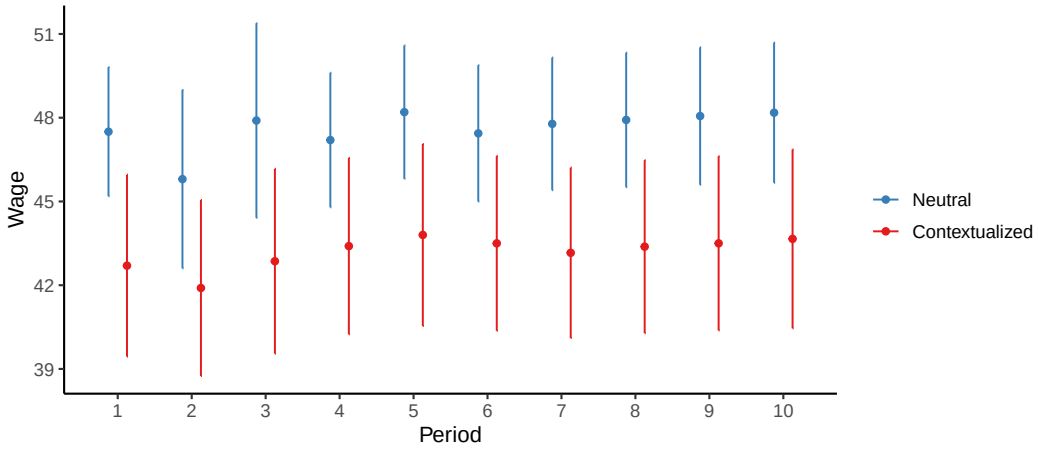
Panel A: Dyad-period Level

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Wage Paid	500	47.598	9.420	500	43.186	11.367	t = -6.68	p < 0.001
Effort Returned	500	0.495	0.142	500	0.566	0.151	t = 7.59	p < 0.001
Payoff Firm	500	25.844	8.749	500	31.257	7.916	t = 10.26	p < 0.001
Payoff Manager	500	41.658	9.587	500	35.588	10.412	t = -9.59	p < 0.001
Combined Payoff	500	67.502	7.779	500	66.845	10.914	t = -1.10	p = 0.273

Panel B: Dyad Level

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Average Payoff Firm	50	258.440	78.646	50	312.568	75.727	t = 3.51	p < 0.001
Average Payoff Manager	50	416.580	82.384	50	355.880	98.557	t = -3.34	p = 0.001
Average Combined Payoff	50	337.510	36.094	50	334.224	53.516	t = -0.36	p = 0.720

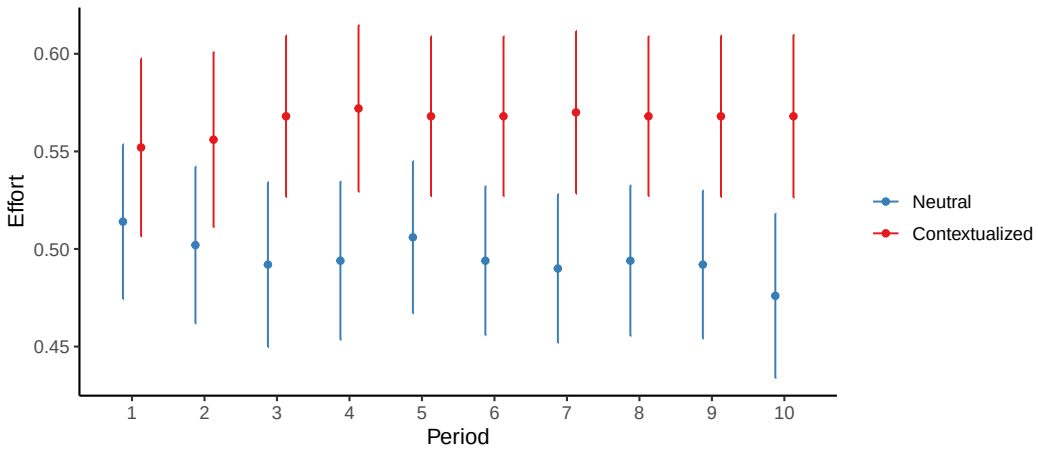
Gift Exchange: Wage by Period



Gift Exchange Game: Wage Regression Results

	Period Fixed Effects			Interacted by Period		
	Est	S.E.	p-value	Est	S.E.	p-value
Intercept				46.831	1.129	<0.001
Contextualized	-4.412	1.887	0.044	-4.359	1.609	0.024
Period				0.140	0.066	0.063
Period × Contextualized				-0.010	0.099	0.924
Adjusted R^2	0.036			0.041		
Number of observations	1,000			1,000		

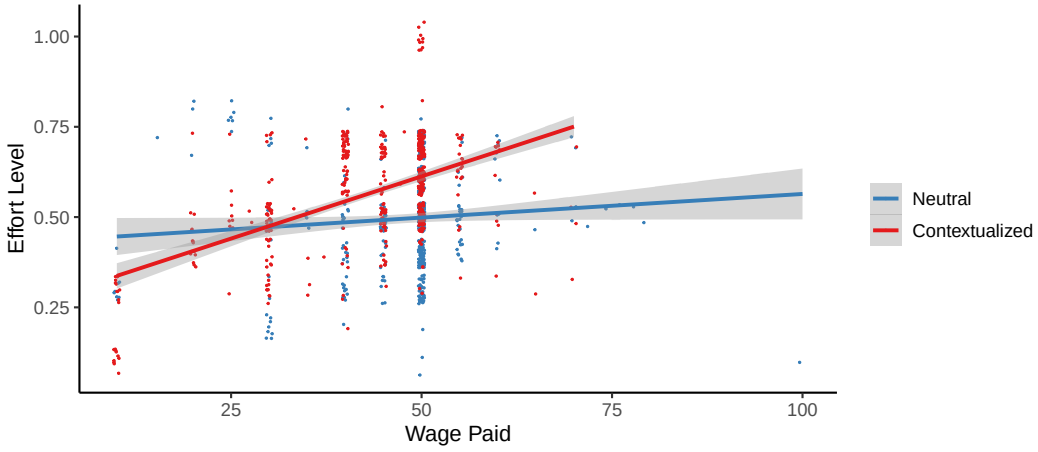
Gift Exchange: Effort by Period



Gift Exchange Game: Effort Regression Results

	Period Fixed Effects			Interacted by Period		
	Est	S.E.	p-value	Est	S.E.	p-value
Intercept				0.510	0.016	<0.001
Contextualized	0.070	0.028	0.033	0.049	0.024	0.075
Period				-0.003	0.001	0.018
Period × Contextualized				0.004	0.002	0.029
Adjusted R^2	0.046			0.053		
Number of observations	1,000			1,000		

Gift Exchange: Wage Effort Effect



Gift Exchange Game: Wage Effort Effect Regression Results

	(1)		
	Est	S.E.	p-value
Contextualized	-0.164	0.136	0.257
Wage	0.001	0.002	0.582
Wage × Contextualized	0.006	0.003	0.077
Adjusted R^2	0.184		
Number of observations	1,000		

Gift Exchange Game: Reasons for Wage Decision

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Mentions Payoff	500	0.694	0.461	500	0.902	0.298	$\chi^2 = 65.81$	$p < 0.001$
Mentions Other	500	0.916	0.278	500	0.972	0.165	$\chi^2 = 13.79$	$p < 0.001$
Cares About Own Payoff	500	6.866	1.322	500	7.112	1.122	$t = 3.17$	$p = 0.002$
Cares About Other Payoff	500	5.728	1.289	500	5.592	1.375	$t = -1.61$	$p = 0.107$
Cares About Fairness	500	7.156	1.481	500	6.584	1.483	$t = -6.10$	$p < 0.001$
Cares About Reciprocity	500	6.089	1.657	500	6.046	1.590	$t = -0.42$	$p = 0.673$

Gift Exchange Game: Reasons for Effort Decision

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Mentions Payoff	500	0.934	0.249	500	0.968	0.176	$\chi^2 = 5.49$	$p = 0.019$
Mentions Other	500	0.880	0.325	500	0.854	0.353	$\chi^2 = 1.25$	$p = 0.264$
Cares About Own Payoff	500	6.940	1.242	500	7.522	1.073	$t = 7.93$	$p < 0.001$
Cares About Other Payoff	500	5.442	1.540	500	5.224	1.797	$t = -2.06$	$p = 0.040$
Cares About Fairness	500	7.136	1.574	500	6.516	1.909	$t = -5.60$	$p < 0.001$
Cares About Reciprocity	500	4.242	1.793	500	4.080	1.711	$t = -1.46$	$p = 0.144$

Trust Game: Rules

Two-player sequential game

- Player A gets 100 points, Player B gets 0
- Player A can transfer anything from 0 to 100 points to Player B
- Keeps whatever they do not send
- Amount sent to Player B is tripled by experimenter
- Player B then decides how much to send back to Player A
- Keeps whatever they do not send back

Trust Game: Contextualization Manuipulation

Neutral:

- Participant A and Participant B
- Amount sent and amount sent back

Contextualized:

- Investor and Manager
- Investment and dividend

Trust Game: Descriptives

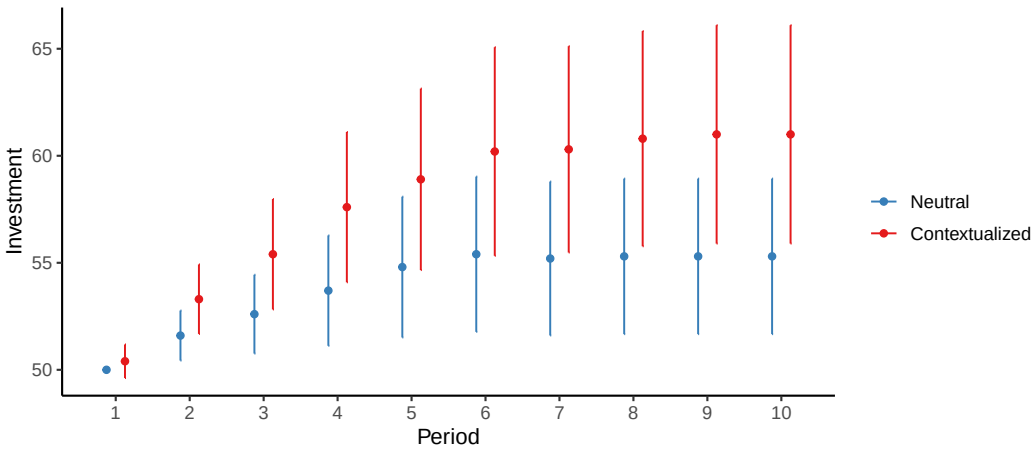
Panel A: Dyad-period Level

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Investment	500	53.920	10.755	500	57.890	14.921	t = 4.83	p < 0.001
Dividend	500	73.876	21.264	500	78.772	27.461	t = 3.15	p = 0.002
% Returned	500	0.455	0.081	500	0.451	0.094	t = -0.60	p = 0.548
Payoff Investor	500	119.956	14.041	500	120.882	17.851	t = 0.91	p = 0.362
Payoff Manager	500	87.884	19.521	500	94.898	27.993	t = 4.60	p < 0.001
Combined Payoff	500	207.840	21.510	500	215.780	29.842	t = 4.83	p < 0.001

Panel B: Dyad Level

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Average Payoff Investor	50	1199.560	139.516	50	1208.820	176.375	t = 0.29	p = 0.772
Average Payoff Manager	50	878.840	180.717	50	948.980	253.283	t = 1.59	p = 0.114
Average Combined Payoff	50	1039.200	95.780	50	1078.900	129.934	t = 1.74	p = 0.085

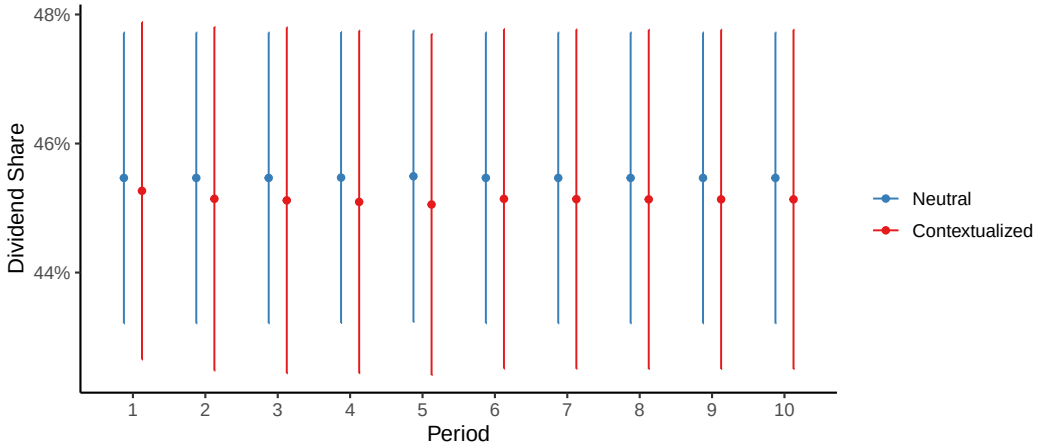
Trust Game: Investment by Period



Trust Game: Investment Regression Results

	Period Fixed Effects			Interacted by Period		
	Est	S.E.	p-value	Est	S.E.	p-value
Intercept				50.847	0.637	<0.001
Contextualized	3.970	2.300	0.118	0.853	0.416	0.070
Period				0.559	0.242	0.047
Period × Contextualized				0.567	0.403	0.193
Adjusted R^2	0.054			0.058		
Number of observations	1,000			1,000		

Trust Game: Dividend Share by Period



Trust Game: Dividend Share Regression Results

	Period Fixed Effects			Interacted by Period		
	Est	S.E.	p-value	Est	S.E.	p-value
Intercept				0.455	0.009	<0.001
Contextualized	-0.003	0.018	0.853	-0.003	0.014	0.828
Period				0.000	0.001	0.998
Period × Contextualized				0.000	0.002	0.973
Adjusted R^2	-0.010			-0.003		
Number of observations	1,000			1,000		

Trust Game: Reasons for Investment Decision

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Mentions Payoff	500	0.232	0.423	500	0.282	0.450	$\chi^2 = 3.02$	$p = 0.082$
Mentions Other	500	0.756	0.430	500	0.696	0.460	$\chi^2 = 4.23$	$p = 0.040$
Cares About Own Payoff	500	7.130	1.735	500	8.006	1.114	$t = 9.50$	$p < 0.001$
Cares About Other Payoff	500	5.398	1.693	500	4.316	2.083	$t = -9.02$	$p < 0.001$
Cares About Fairness	500	7.060	1.999	500	5.718	2.121	$t = -10.30$	$p < 0.001$
Cares About Trust	500	7.500	1.353	500	6.776	1.867	$t = -7.02$	$p < 0.001$

Trust Game: Reasons for Dividend Decision

	Neutral			Contextualized			Tests for Differences	
	N	Mean	SD	N	Mean	SD		
Mentions Payoff	500	0.112	0.316	500	0.218	0.413	$\chi^2 = 19.63$	$p < 0.001$
Mentions Other	500	0.578	0.494	500	0.766	0.424	$\chi^2 = 39.24$	$p < 0.001$
Cares About Own Payoff	500	4.792	1.596	500	5.164	1.551	$t = 3.74$	$p < 0.001$
Cares About Other Payoff	500	5.640	1.372	500	5.718	1.229	$t = 0.95$	$p = 0.344$
Cares About Fairness	500	8.982	1.002	500	8.480	1.222	$t = -7.10$	$p < 0.001$
Cares About Trust	500	7.880	1.081	500	8.346	1.074	$t = 6.84$	$p < 0.001$

Conclusions

- Contextualization matters in non-trivial ways
- Emphasizes the need for pre-tests and pilots
- botex is a useful tool for these purposes

Limitations & Future Work

- Are LLM participants reasonable representative of humans in our domain?
- Modify LLM personalities to mimic certain human characteristics (think about MBA students vs. general population)
- Other experiment types beyond contextualization (e.g., surveys, scenario experiments)
- Study the “behavior” and “attitudes” of LLMs relative to and in interaction with human participants
- Use LLMs for exploration/hypotheses development?
- Fundamental concerns about reliance on LLM participants for inferences about human behavior?