Learning with Misspecified Models: The Case of Overconfidence

Jimena Galindo

November 2, 2023

Example

A student needs to choose how much to study for an exam.

Their choice depends on two factors:

1. What they think their ability is
2. How generous they expect the grading system to be

Their choice will affect their grade.

Example (continued)

If their grade is surprisingly high, they can attribute it to two things:

1. Their ability is higher than they thought
2. Grading is more generous than they expected

The way in which they incorporate the information will affect their study choice for the next exam.

Example (continued)

If their grade is surprisingly high, they can attribute it to two things:

1. Their ability is higher than they thought
2. Grading is more generous than they expected

The way in which they incorporate the information will affect their study choice for the next exam.

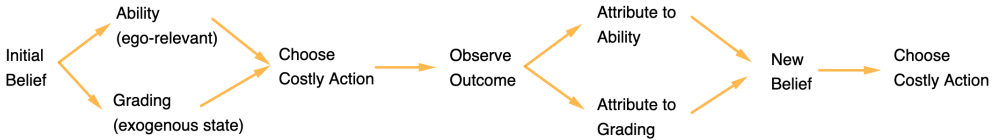
This Project: I experimentally study the mechanisms through which the student might hold incorrect beliefs in the long-run

The Mechanisms

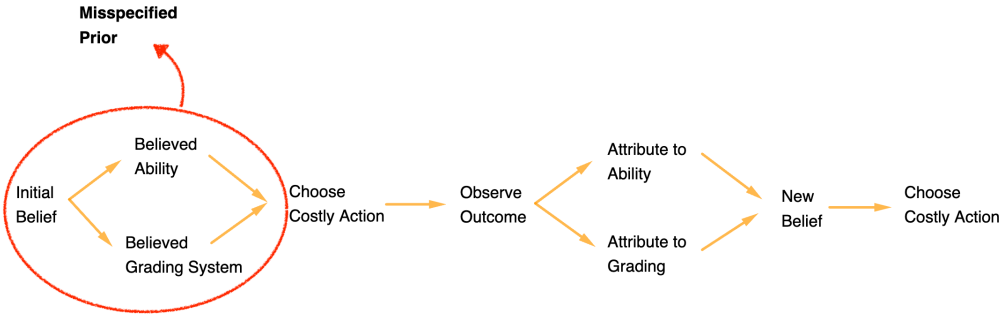
Three mechanisms for incorrect learning:

1. Misspecified initial belief
2. Learning traps
3. Incorrect updating procedure

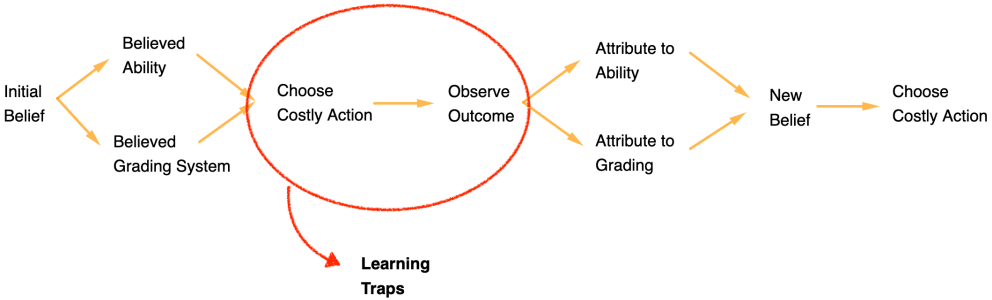
The Learning Problem



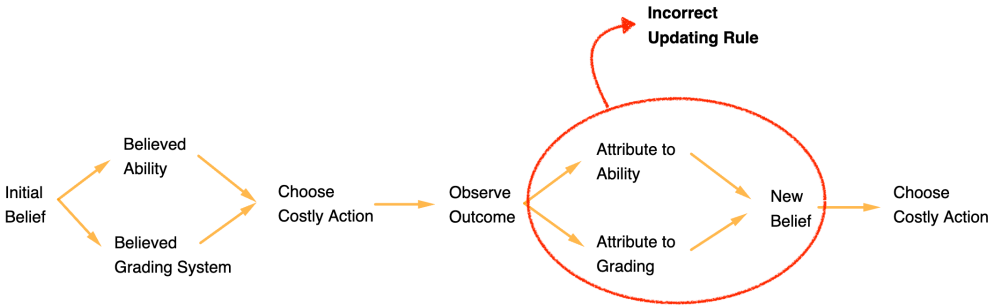
Reason 1: Misspecified Initial Belief



Reason 2: Learning Traps



Reason 3: Incorrect Updating



Incorrect Learning

Overestimation: Belief that the value of a parameter is larger than it truly is.

- *e.g. Believing IQ is 150 when it is actually 100*
- Called overconfidence if the belief is about the ego-relevant parameter

Incorrect Learning

Overestimation: Belief that the value of a parameter is larger than it truly is.

- *e.g. Believing IQ is 150 when it is actually 100*
- Called overconfidence if the belief is about the ego-relevant parameter

It is prevalent in diverse settings:

- Excess entry of entrepreneurs (Camerer and Lovallo, 1999)
- Suboptimal genetic testing and savings (Oster et al. 2013)
- Workers overestimate their productivity (Hoffman and Burks, 2020)

Overestimation leads to costly choices

Four Theories of Misspecified Learning

1. **Myopic Bayesian** (Hestermann and Le Yaouanq (2021))
   - Learning Traps
2. **Motivated Beliefs/Attribution Bias** (Brunnermeier and Parker (2005), Bracha and Brown (2012), Mobius et al. (2014))
   - Biased updating
3. **Paradigm Shifts** (Schwarstein and Sunderam (2021), Ba (2022))
   - Misspecified initial beliefs
   - Belief updating through hypothesis tests
4. **Dogmatic Modelers** (Heidhues et al. (2018))
   - Misspecified initial beliefs
   - Learning traps

All rationalize the prevalence of overconfidence

Research Questions

1. Which of the proposed mechanisms gives a better explanation of behavior in the lab?

2. Can the same mechanisms explain incorrect beliefs when the parameters are not ego-relevant?
    - Can they explain the prevalence of stereotypes?

Road-map

1. Unifying Framework
2. Mechanisms and Predictions
3. Experimental Design
4. The Data
5. Results

Framework

A Unifying Framework

Type (Ability): $\theta \in \{\theta_H, \theta_M, \theta_L\}$

State (Grading): $\omega \in \{\omega_H, \omega_M, \omega_L\}$ drawn from a discrete-uniform distribution

A Unifying Framework

Type (Ability): $\theta \in \{\theta_H, \theta_M, \theta_L\}$
State (Grading): $\omega \in \{\omega_H, \omega_M, \omega_L\}$ drawn from a discrete-uniform distribution

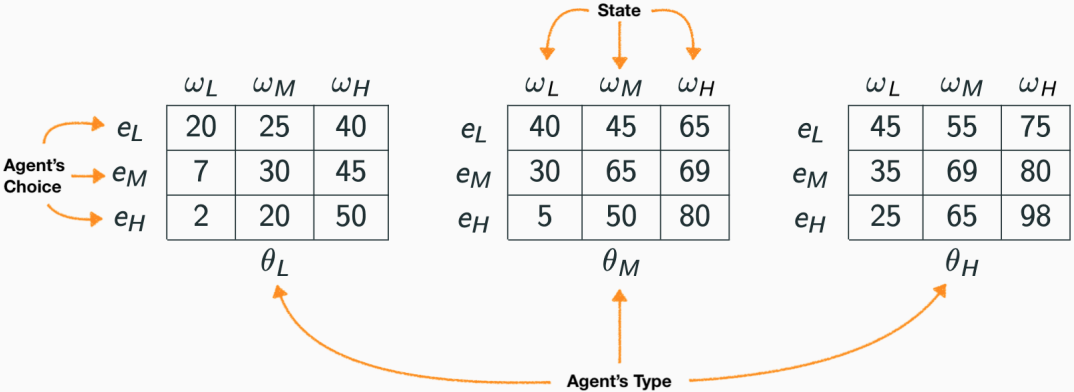
Each period, the agent makes a choice and observes an outcome:

Choice (Study time): $e_t \in \{e_H, e_M, e_L\}$
Outcome: $s_t \in \{\text{success, failure}\}$

Probability of success: $p\left[success_t | e_t, \omega, \theta\right]$
Payoff: $v > 0$ if the outcome is a success, 0 otherwise

Choose e to maximize the probability of success at each period $t = 1, 2, ...$

The probability of success: $p[success|e, \omega, \theta]$



An Example

For an agent of type θ_M and a state ω_M, the probability of success is given by:

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

The Probability of Success: Optimal Choices

Conditional on:

- A type (Matrix)
- A state (Column)

The choice (row) that maximizes the probability of success is the one that matches the state

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

Learning Correctly is Possible

- Suppose they are of type θ_M and the state is ω_M
- But they believe they are θ_H

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

1. Choose 2 distinct actions for T periods each
2. There is a unique column that rationalizes the average number of successes for both choices

Learning Correctly is Possible

- Suppose they are of type θ_M and the state is ω_M
- But they believe they are θ_H

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

1. Choose e_L for 100 periods $\rightarrow$ 45% success rate

Learning Correctly is Possible

- Suppose they are of type θ_M and the state is ω_M
- But they believe they are θ_H

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

1. Choose e_L for 100 periods $\rightarrow$ 45% success rate
2. Choose e_M for 100 periods $\rightarrow$ 65% success rate

Why do incorrect beliefs persist?

Mechanisms and Predictions

Bayesian Benchmark

Based on Hesterman and Le Yaouanq, (2021)

Myopic Bayesian

Start with a diffused prior over (θ, ω) and updates correctly

$$p_{t+1}(\theta, \omega | s_t) = \frac{p_t(s_t|\theta, \omega) p_t(\theta, \omega)}{\sum_{(\theta', \omega')} p_t(s_t|\theta', \omega') p_t(\theta', \omega')}$$

Is myopic: maximizes the period utility and not the future flow of payoffs

$P(success|e, \theta, \omega)$ is increasing in ω and θ

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

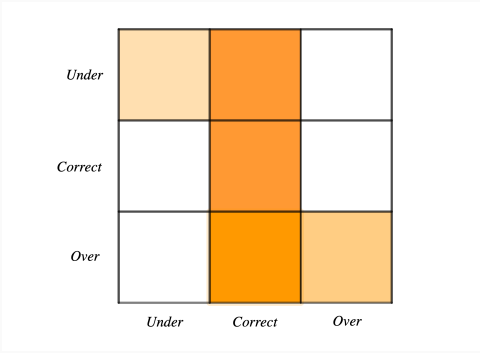
θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

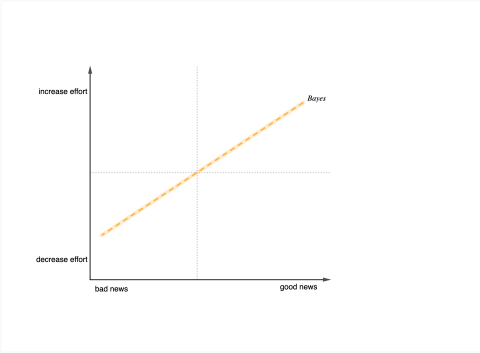
θ_H

- Streaks of successes will be attributed to higher θ and $\omega \rightarrow$ higher e
- Streaks of failures will be attributed to lower θ and $\omega \rightarrow$ lower e

Bayesian: Predictions



Predicted Transition Matrix.



Predicted Reaction to News.

Dogmatic Modeling

Based on Heidhues, Koszegi, and Strack, (2018)

The Dogmatic Modeler (Heidhues et al. (2018))

Agent of true type θ^*

Holds a degenerate belief: type is $\hat{\theta}$ with probability 1

Their belief is potentially misspecified:

- Overconfident if $\hat{\theta} > \theta^*$
- Underconfident if $\hat{\theta} < \theta^*$

Updates $p_t(\omega)$ using Bayes Rule

$$p_{t+1}(\omega|s, \hat{\theta}) = \frac{p_t(s_t|\omega, \hat{\theta})p_t(\omega)}{\sum_{\omega'} p_t(s_t|\omega', \hat{\theta})p_t(\omega')}$$

The Dogmatic Modeler: Mechanism

Agent of type θ_M and state ω_M who dogmatically believes he is θ_H

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

The Dogmatic Modeler: Mechanism

Agent of type θ_M and state ω_M who dogmatically believes he is θ_H

1. Chooses e_H and is disappointed $\rightarrow$ adjust belief about ω downward

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	**98**

θ_H

The Dogmatic Modeler: Mechanism

Agent of type θ_M and state ω_M who dogmatically believes he is θ_H

1. Chooses e_H and is disappointed $\rightarrow$ adjust belief about ω downward

2. Eventually chooses e_M and is disappointed as well $\rightarrow$ adjust belief about ω

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	**69**	80
e_H	25	65	98

θ_H

The Dogmatic Modeler: Mechanism

Agent of type θ_M and state ω_M who dogmatically believes he is θ_H

1. Chooses e_H and is disappointed $\rightarrow$ adjust belief about ω downward

2. Eventually chooses e_M and is disappointed as well $\rightarrow$ adjust belief about ω

3. Eventually chooses e_L and falls into a self-defeating equilibrium

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	**45**	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

The Stable Beliefs: Learning Traps

Dogmatic beliefs can only be sustained when there is a self-confirming equilibrium

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

The Stable Beliefs: Learning Traps

Dogmatic beliefs can only be sustained when there is a self-confirming equilibrium

- **Underconfident** stable beliefs

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

The Stable Beliefs: Learning Traps

Dogmatic beliefs can only be sustained when there is a self-confirming equilibrium

- **Overconfident** stable beliefs

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

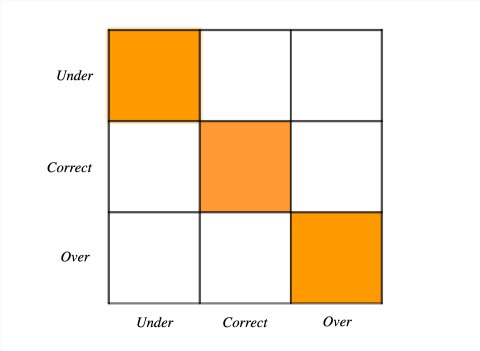
	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

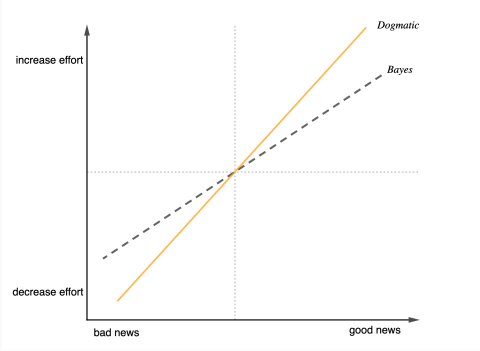
	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

Dogmatic: Predictions



Predicted Transition Matrix.



Predicted Reaction to News.

Paradigm Shifts

Based on Ba, (2022)

The Switcher: (Ba (2022))

Same initial belief as the Dogmatic, but is willing to consider an alternative paradigm θ'

Keeps track of the likelihoods of the two possible paradigms:

- $p_t(s_t|\cdot)$ for $\hat{\theta}$ and θ'

They switch to whichever paradigm is more likely to have generated the signals

$$\frac{p_t(s_t|\theta')}{p_t(s_t|\hat{\theta})} > \alpha \geq 1$$

The Switcher: Mechanism

- Chooses e_H and is disappointed $\rightarrow$ adjust belief about ω downward

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

The Switcher: Mechanism

- Eventually chooses e_M and is disappointed as well $\rightarrow$ adjust belief about ω

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

θ_H

The Switcher: Mechanism

- Avoids the trap if the likelihood ratio of θ_M to θ_H is goes above α

	ω_L	ω_M	ω_H
e_L	20	25	40
e_M	7	30	45
e_H	2	20	50

θ_L

	ω_L	ω_M	ω_H
e_L	40	45	65
e_M	30	65	69
e_H	5	50	80

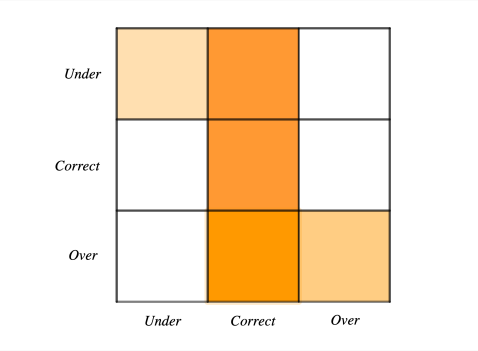
θ_M

	ω_L	ω_M	ω_H
e_L	45	55	75
e_M	35	69	80
e_H	25	65	98

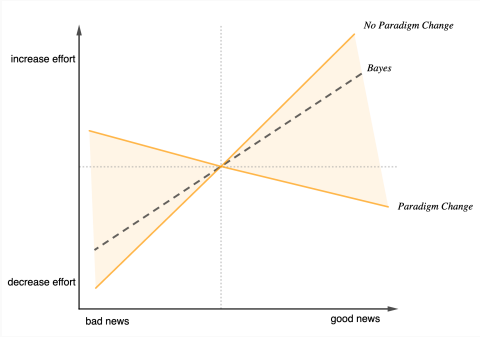
θ_H

A change in paradigm will sometimes induce a change in effort in the opposite direction of the signal

Switcher: Predictions



Predicted Transition Matrix.



Predicted Reaction to News.

Attribution Bias/Motivated Beliefs/Optimal Expectations

Based on Benjamin (2019)

Attribution Bias

Start with a diffused prior over (θ, ω) but updates with a bias

$$p_{t+1}(\theta, \omega | s_t) = \frac{p_t(s_t|\theta, \omega)^{c(\theta,\omega,s_t)} p_t(\theta, \omega)}{\sum_{(\theta',\omega')} p_t(s_t|\theta', \omega')^{c(\theta',\omega',s_t)} p_t(\theta', \omega')}$$

And bias is such that:

- Successes are attributed to high θ
- Failures are attributed to low ω

Self-Attribution: Mechanism

Chooses e that maximizes utility according to current belief

- Belief on ω deteriorates a lot after bad news $\rightarrow$ overreaction in effort
- Belief on θ increases a lot after good news $\rightarrow$ underreaction in effort (or in opposite direction)

Dogmatic: Predictions

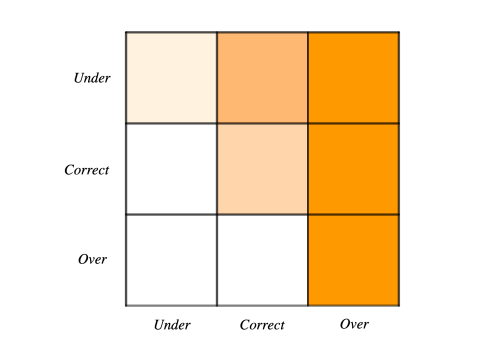


Figure 1: Predicted Transition Matrix.

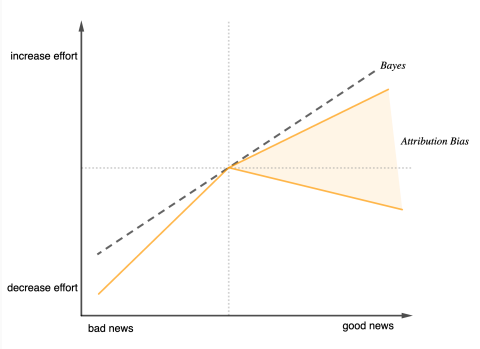


Figure 2: Predicted Reaction to News.

Predictions

- **Dogmatic**:
  - Overreact to signals relative to the Bayesian
  - If there is a trap, they fall into it
- **Switcher:**
  - If status-quo: overreacts
  - If paradigm-shift: underreacts or opposite
  - Able to escape traps (most of the time)
- **Attribution Bias**:
  - Overreacts to bad news
  - Underreacts to good news (or in opposite direction)
  - Become overconfident even when initially correct

Experimental Design

The Experiment

Two parts:

1. Setting the types
2. Updating

Two treatments:

1. Ego
2. Stereotype

Set the Types

- Quiz: Answer as many questions as you can in 2 minutes
  - Math, Verbal, Pop-Culture, Science, U.S. Geography, Sports and Video games

- For each topic, how many questions do you think you answered correctly?
  - 0 to 5 (θ_L)
  - 6 to 15 (θ_M)
  - 16 or more (θ_H)

Science and Technology Quiz

Time left to complete this page: **1:19**

Which cell organelle is also called powerhouse of the cell?

- Ribosome
- Endoplasmic reticulum
- Cytoplasm
- Mitochondria

Next

Choice and Update

Choice and feedback (One topic at a time)

- A success rate is drawn at random (A, B or C)
- Choose a gamble: A, B or C (effort)
- Receive a sample of 10 signal realizations

x 11 per topic

Eliciting Beliefs?

I do not directly elicit beliefs:

- Track their belief about ω with their choices
- Eliciting beliefs for θ can incentivize learning in a way that is not consistent with the theory

Allow them to see the probability matrix for only one type

- Track the matrix they choose to see in each round

Screen: Belief Reminder

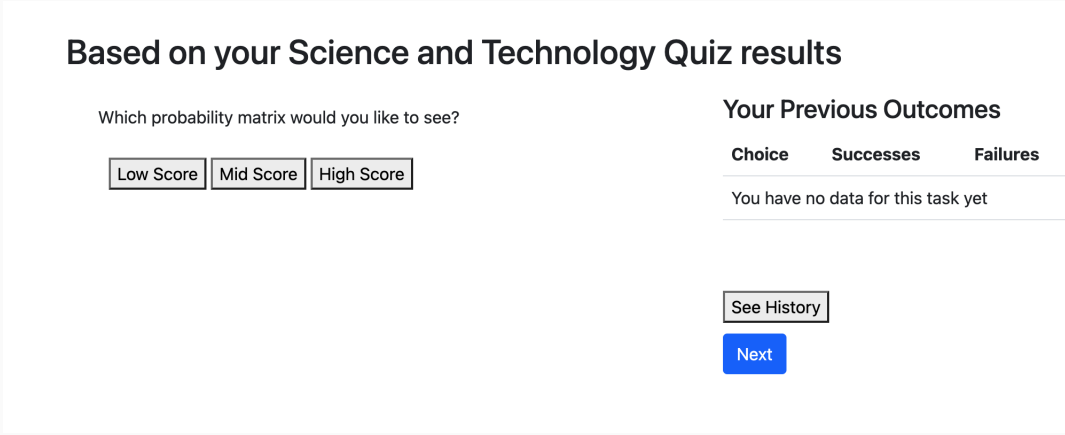
Science and Technology

The next 11 rounds will be based on your Science and Technology quiz results.

You guessed that your score was **High-Score .**

Next

Screen:Choose a Matrix



Screen:Choose a Gamble

Based on your Science and Technology Quiz results

Which probability matrix would you like to see?

Low Score | Mid Score | High Score

Choose a gamble :	Rate A	Rate B	Rate C
A ○	45	55	75
B ○	35	69	80
C ○	25	65	98

Your Previous Outcomes

Choice	Successes	Failures
You have no data for this task yet		

See History

Next

Screen:Feedback

Based on your Science and Technology Quiz results

The outcome of your gamble is: **4** successes and **6** failures

Success	Failure	Failure	Success	Success
Failure	Failure	Failure	Success	Failure

Next

Stereotype Condition

Observe the characteristics of another participant

- Gender
- US National or not

Answer the same questions about self and other

Belief updating and effort choice:

- The DGP depends on the θ the other participant

x 11 per topic

The Data

Subject pool:

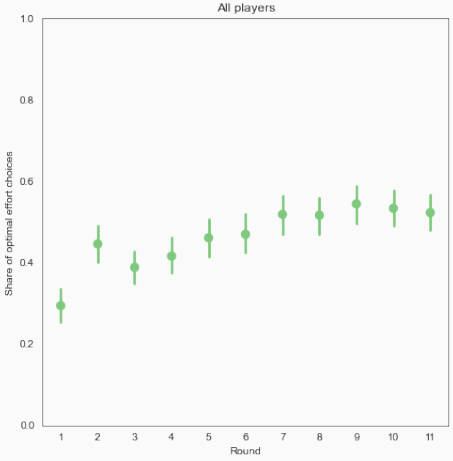
- Run at the CESS lab in person
- 45 subjects in Ego
- 41 subjects in Stereotype

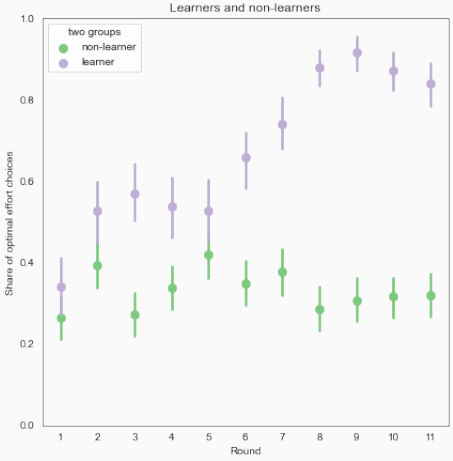
Sessions:

- 9 sessions
- About 45 minutes long
- Average payment: $23
  - $10 show-up fee
  - $0.20 per correct answer
  - $0.20 per success
  - Paid one topic at random

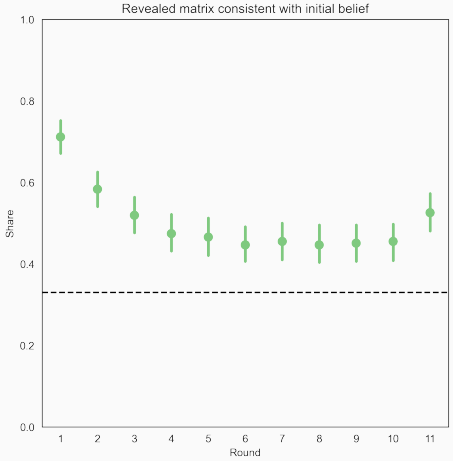
Results

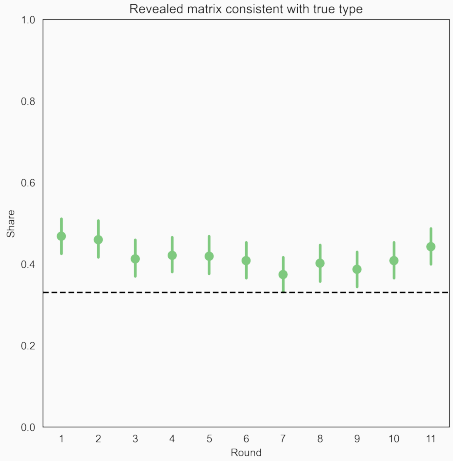
Are they learning ω?





Are they learning Θ?



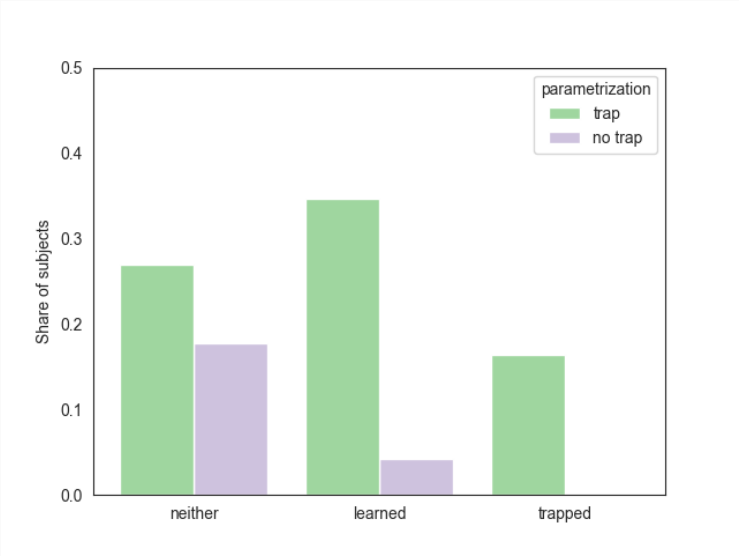


Reasons for lack of learning

- Learning traps
- Attribution Bias
- Considering the wrong alternative paradigms

Learning Traps

Are people falling into traps?



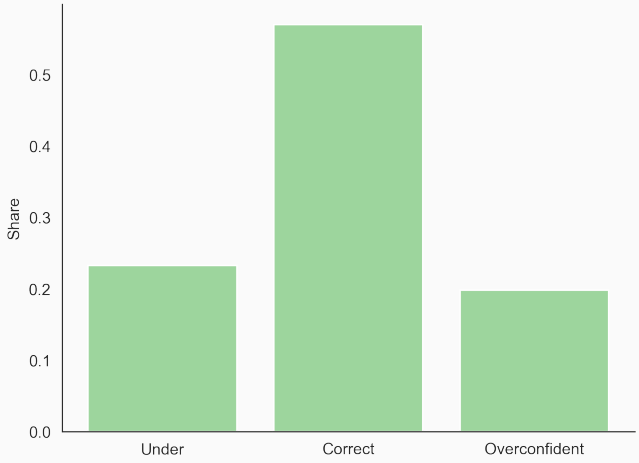
Learners, Trapped and Others

So far we have seen that:

- 44% of the subjects learn the true state
- 16% of the subjects fall into traps
- 40% of the subjects don't learn correctly and don't fall into traps
    - From these 60% were facing parameters for which there were traps

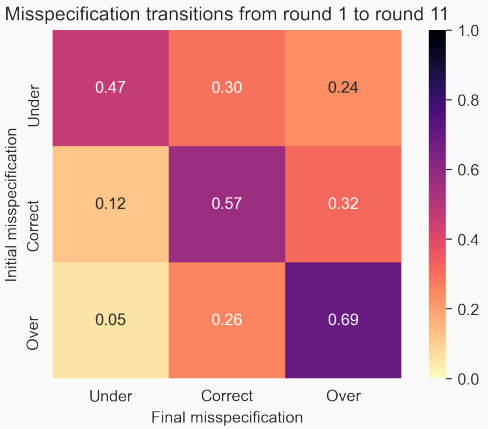
Attribution Bias

Initial Specifications

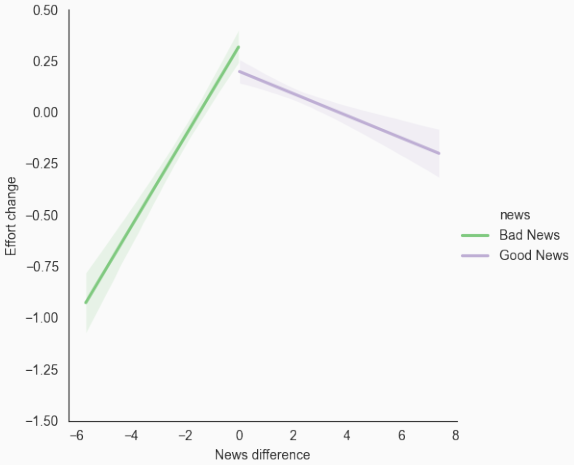


certainties seterotypes

Transition Matrix



Good News v. Bad News



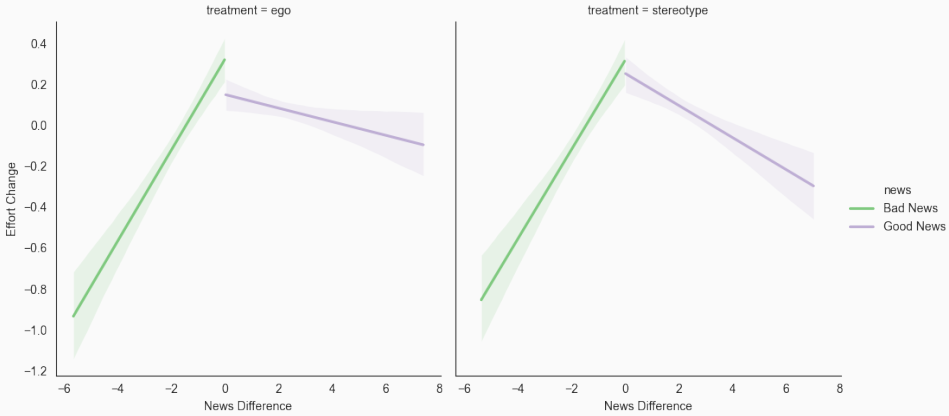
other

Regression Results

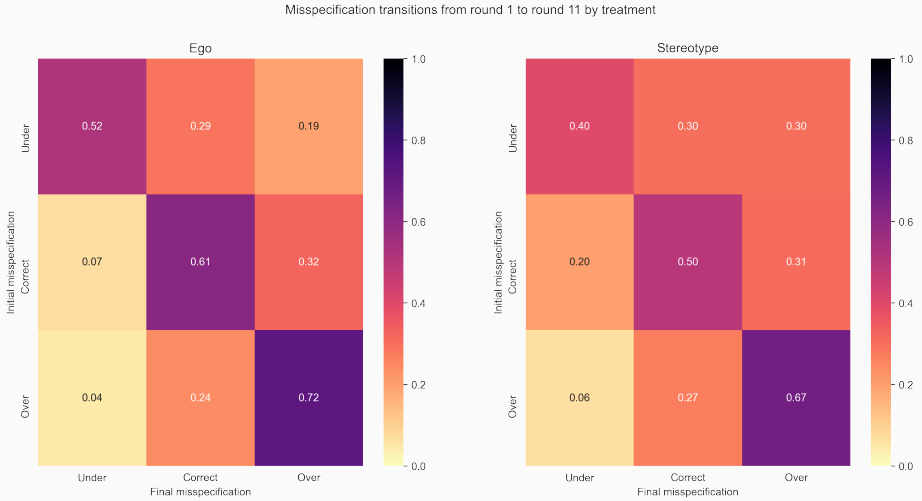
	Dependent variable:				
	Change in effort				
	All	Ego-relevant	Stereotype	Bayesian Simulation	Dogmatic Simulation
	(1)	(2)	(3)	(4)	(5)
Good news	−0.12**	−0.16***	−0.05	0.08	−0.08
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
News difference	0.22***	0.22***	0.21***	0.06***	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
News difference * Good news	−0.27***	−0.25***	−0.29***	−0.04	−0.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	0.31***	0.31***	0.30***	−0.08*	0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Observations	4,680	2,700	1,980	4,680	4,680
R^2	0.04	0.04	0.04	0.05	0.06
Adjusted R^2	0.04	0.04	0.04	0.05	0.06
Note:					*p<0.1; **p<0.05; ***p<0.01

Stereotypes

Asymmetric Updating in the Stereotype Condition



Do misspecifications persist more often in the Ego condition?



Differences across treatments

Small differences across treatments

- Less stickiness in initial beliefs in Stereotype
- Attribution bias in Ego condition
- Possible self-censoring in Stereotype

Concluding Remarks

Summary

Three mechanisms through which an agent might hold incorrect long-run beliefs:

- Incorrect initial beliefs
- Learning traps
- Attribution bias

Results: Attribution bias is the best explanation for aggregate behavior

- Asymmetric treatment of good and bad news
- Tendency to become overconfident

Ego-relevance v. Stereotypes: Similar pattern for different reasons

- Over-correction of initial biases about others

Other explanations

Other Results:

I estimate the structural parameters of the models.

- α is identified from paradigm changes
- $c(\theta, \omega, \textit{news})$ is estimated using SMM

I sort subjects into the best-fitting model:

- Attribution bias is the best fit for most subjects
- Some better explained by paradigm shifts
- Very few dogmatic and Bayesian

The end

Thank you!

Other Explanations

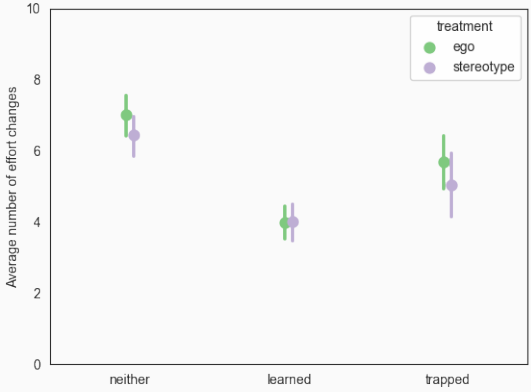
Intentional Exploration

Hestermann and Le Yaouanq (2021) propose a model with endogenous exploration

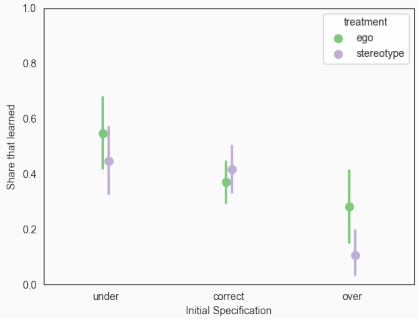
- Overconfident agents are more likely to explore
- Underconfident agents are always pleasantly surprised and do not explore as much

Underconfidence would be more persistent

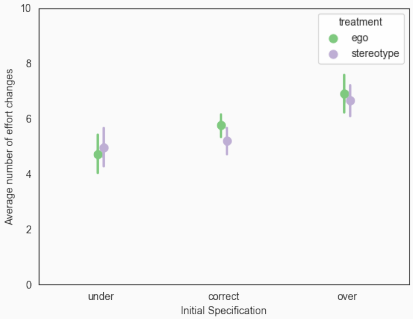
Choice Changes



By Initial Specification



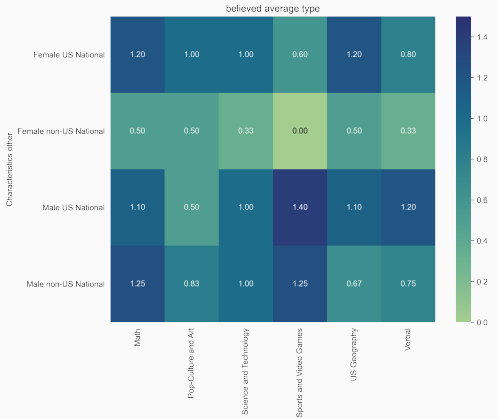
Learning

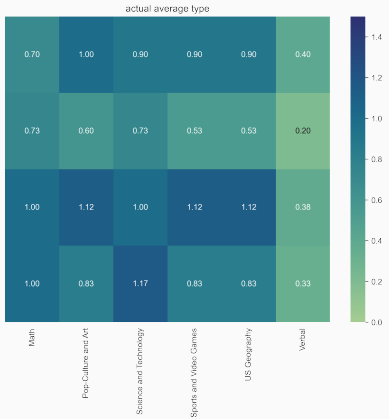


Choice Changes

Back

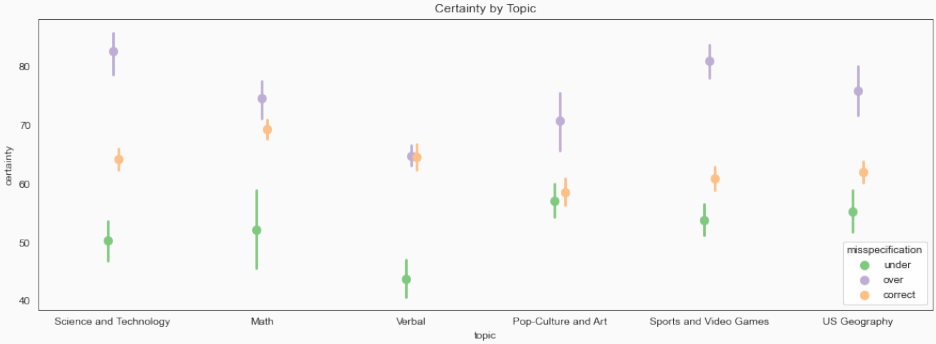
Misspecifications





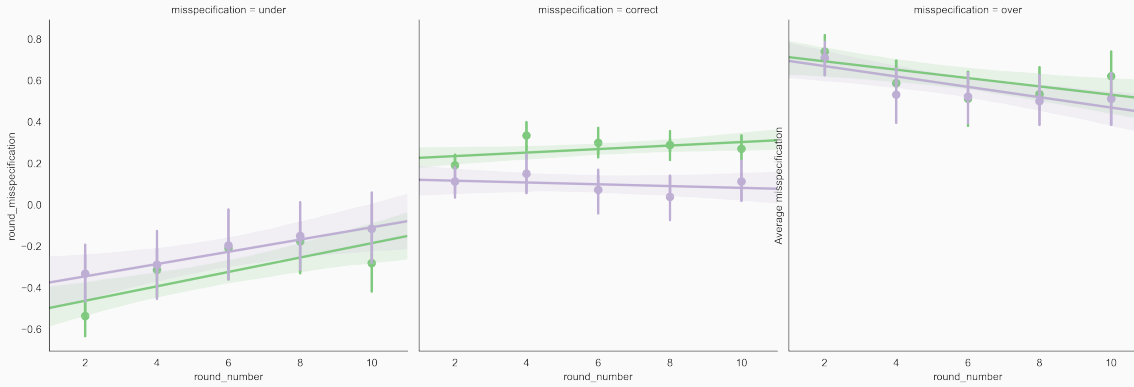
Back

Certainties



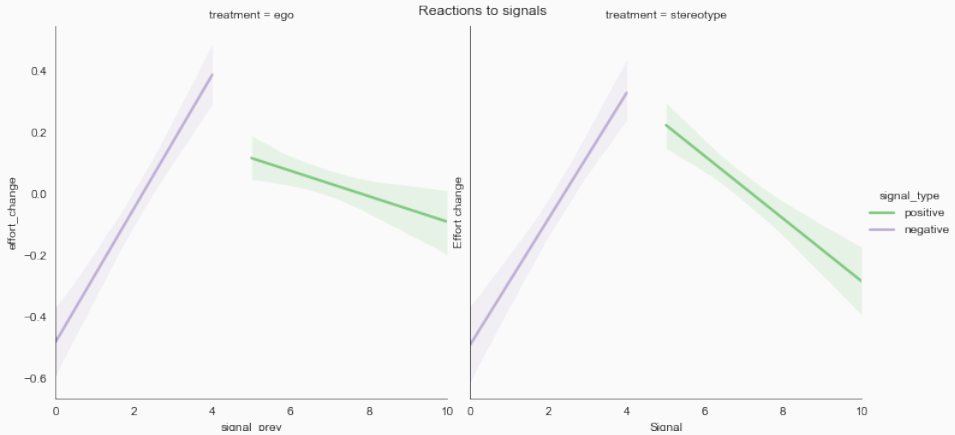
Back

Misspecification changes by treatment



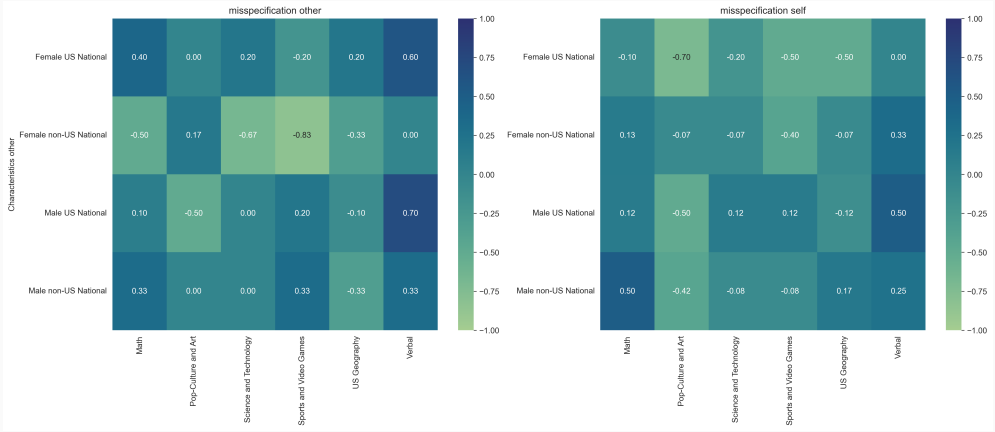
Back

Positive Signals v. Negative Signals



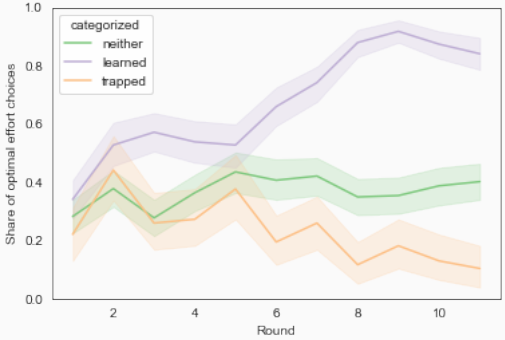
Back

The Stereotypes



types Back

Subject categorization



Back