

Learning with Misspecified Models: The case of overconfidence

Jimena Galindo

October 9, 2023

Overestimation: Belief that type is higher than it truly is

Overestimation: Belief that type is higher than it truly is

Seems to be persistent in various 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)

Ultimately it leads to sub-optimal choices

Models of Learning

Focus on setting with 2 parameters:

- An **Ego-Relevant** parameter
- An **Exogenous** parameter

Some of the assumptions that theory has incorporated to rationalize overconfidence are:

- Dogmatism
- Paradigm shifts
- Motivated beliefs
- Myopic Bayesian

Four Theories of Misspecified Learning

1. **Self-defeating equilibrium** (Heidhues et al. (2018))
 - Bayesian about exogenous parameters
 - Dogmatic about ego-relevant parameters
2. **Bayesian hypothesis testing** (Schwarstein and Sunderam (2021), Ba (2022))
 - Bayesian about exogenous parameters
 - Paradigm shift for ego-relevant parameters
3. **Motivated Beliefs / Self-Attribution Bias** (Brunnermeier and Parker (2005), Bracha and Brown (2012))
 - Optimally biased updating
 - Utility from held beliefs
4. **Myopic Bayesian** (Hestermann and Le Yaouanq, (2021))
 - Bayesian about both
 - Maximizes flow utility only

Which of the proposed theories gives a better explanation of behavior?

Do the theories apply only to misspecifications about ego-relevant parameters?

- Can the same theories explain the prevalence of stereotypes?

An Example (from Heidhues et al. (2018))

A student has unknown **intrinsic ability** θ^* (ego-relevant parameter)

They choose a level of **effort** $e \geq 0$ (choice)

Effort and ability are evaluated by a **grading system** ω (exogenous parameter)

The student wants to maximize:

$$u(e) = (\theta^* + e)\omega - \frac{1}{2}e^2 + \varepsilon$$

Regardless of their own type and of their beliefs about it, they should choose

$$e^*(\omega) = \omega$$

Learning is Possible

This exercise is repeated for $t = 0, 1, \dots$

$$y_t = (\theta^* + e_t)\omega - \frac{1}{2}e_t^2 + \varepsilon_t$$

Note that both parameters are identified in this setting:

- Choosing \hat{e} and $\hat{e} + 1$ over multiple periods allows identification of ω
- Once ω is known, θ can be backed out

Why do people not learn the true values of the parameters?

From the proposed mechanisms:

- Dogmatism and Bayesian Updating do not seem to explain the behavior
- Some evidence supporting Hypothesis Testing
- Most evidence supporting Motivated Beliefs
- Some indications of forward-looking behavior

Road-map

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

Framework

A Unifying Framework

Ego-relevant parameter: $\theta \in \{\theta_H, \theta_M, \theta_L\}$

Exogenous parameter: $\omega \in \{\omega_H, \omega_M, \omega_L\}$ with $p(\omega_k) = 1/3$

Choices: $e \in \{e_H, e_M, e_L\}$

Binary Outcomes: $s_t \in \{\text{success}, \text{failure}\}$ with $p[\text{success}|e, \omega, \theta]$ and p is an order-preserving transformation of $u(x)$

The Data Generating Process

The probability of success is given by:

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

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

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

The Data Generating Process

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

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

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

The Data Generating Process

Diagram illustrating the data generating process, showing three payoff matrices for different types (θ_L , θ_M , θ_H) across three states (ω_H , ω_M , ω_L). The rows represent the sender's strategy (e_H , e_M , e_L).

Payoff matrices for θ_L , θ_M , and θ_H :

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

θ_L

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

θ_M

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

θ_H

A Stable Misspecified Belief

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

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

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

The Stable Beliefs

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

θ_L

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

θ_M

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

θ_H

Mechanisms and Predictions

An Example

- True type is θ_M
- True parameter is $\omega_M \rightarrow$ the student believes it is uniformly distributed

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

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

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

The Dogmatic Modeler

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

A student who dogmatically believes he is θ_H but truly is θ_M

The exogenous parameter is ω_M

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-confirming equilibrium

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

θ_L

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

θ_M

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

θ_H

path

The Switcher (paradigm shifts)

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

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. Avoids the self-defeating equilibrium if the likelihood of θ_M becomes larger than that of θ_H

path

Self-Attribution Bias / Optimal Expectations

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

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

Bias is such that

$$c(\theta_H, \omega, \text{good news}) \leq c(\theta_M, \omega, \text{good news}) \leq c(\theta_L, \omega, \text{good news}) \leq 1 \quad \forall \omega$$

And

$$c(\theta, \omega_L, \text{bad news}) \leq c(\theta, \omega_M, \text{bad news}) \leq c(\theta, \omega_H, \text{bad news}) \leq 1 \quad \forall \theta$$

1. Chooses e that maximizes utility according to priors
 - Belief on $\mathbb{E}[\omega]$ deteriorates a lot after bad news \rightarrow big change in effort
 - Belief on $\mathbb{E}[\theta]$ increases a lot after good news \rightarrow small positive (or negative) change in effort

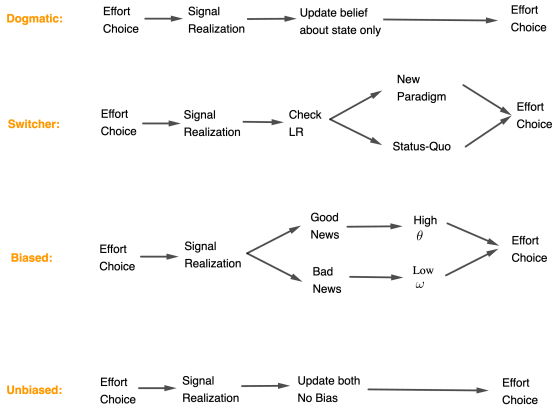
path

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

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

But if they start with a prior that is “tight” around a self-defeating equilibrium they will never learn

All Models



Predictions

	Good News	Bad News	
Dogmatic:	Increase Effort	Decrease Effort	Reacts more than Bayesian
Switcher:	<p>Decrease Effort ← Paradigm Shift → Increase Effort</p> <p>Increase Effort ← Status-Quo → Decrease Effort</p>		Depends on hypothesis test
Biased:	Small Increase in Effort or Decrease Effort	Decrease Effort	<p>Reacts more than Bayesian to bad news</p> <p>Reacts less than Bayesian to good news</p>
Unbiased:	Increase Effort	Decrease Effort	Benchmark

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, Us Geography, Sports and Video games
- How many questions do you think you answered correctly in each quiz?
 - 0 to 5 (θ_L)
 - 6 to 15 (θ_M)
 - 16 or more (θ_H)
- How sure are you about your guess?
 - Random guess $\rightarrow 1/3$
 - Another is equally likely $\rightarrow 1/2$
 - Fairly certain $\rightarrow 3/4$
 - Completely sure $\rightarrow 1$

“Effort” 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

Stereotype condition

Observe the characteristics of a 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

Eliciting 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 success rate matrix for only one type.

- Track the matrices they choose to see in each round

Based on the other participant's Science and Technology Quiz results

Which probability matrix would you like to see?

Low Score

Mid Score

High Score

Your Previous Outcomes

Choice

Successes

Failures

You have no data for this task yet

See History

Next

Based on the other participant's 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	<input type="radio"/>	40	45	65
B	<input type="radio"/>	30	65	69
C	<input type="radio"/>	5	50	80

Your Previous Outcomes

Choice

Successes

Failures

You have no data for this task yet

See History

Next

The Data

The Data

Subject pool:

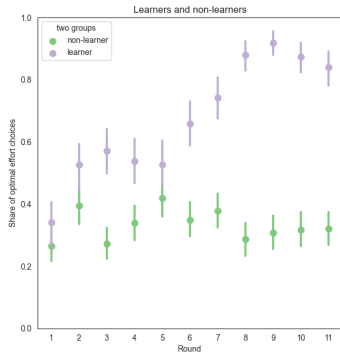
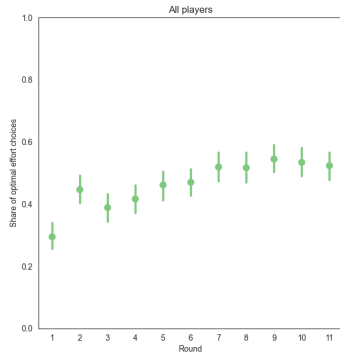
- Run at the CESS lab in person
- 45 subjects in Ego
- 33 subjects in Stereotype

The Sessions:

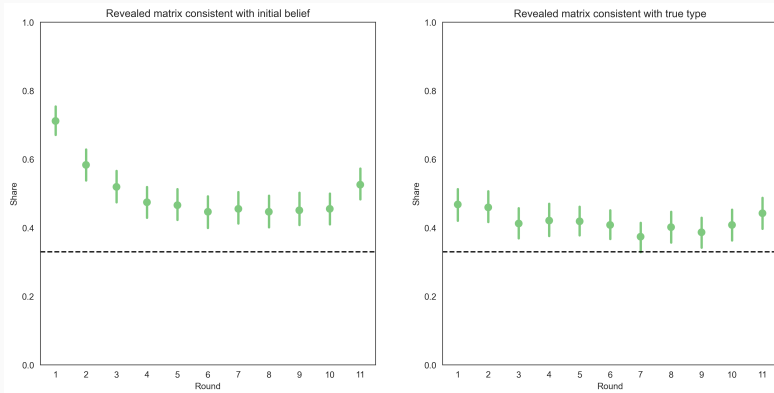
- 8 sessions
- 45 minutes on average
- Average payment: \$23
 - \$10 show-up fee
 - \$0.20 per correct answer
 - \$0.20 per success
 - Paid one topic at random

Learning

Are they learning ω ?



Are they learning Θ

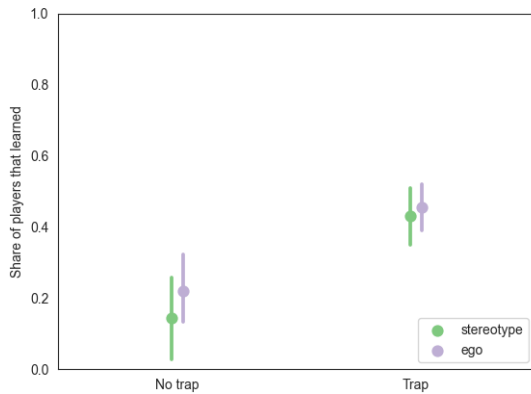


Reasons for lack of learning

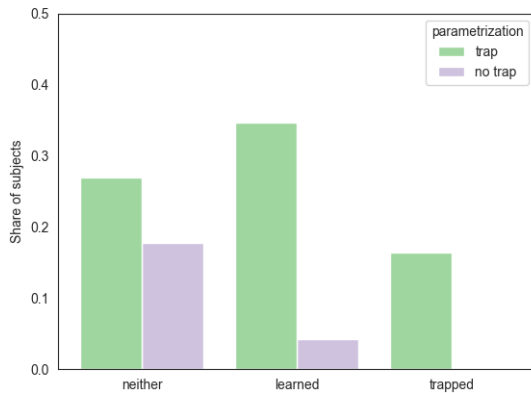
- Learning traps (self-defeating equilibria)
- Misattributions
- Others
 - Considering the wrong paradigms
 - Learning is too costly

Learning Traps

Learning when there are traps

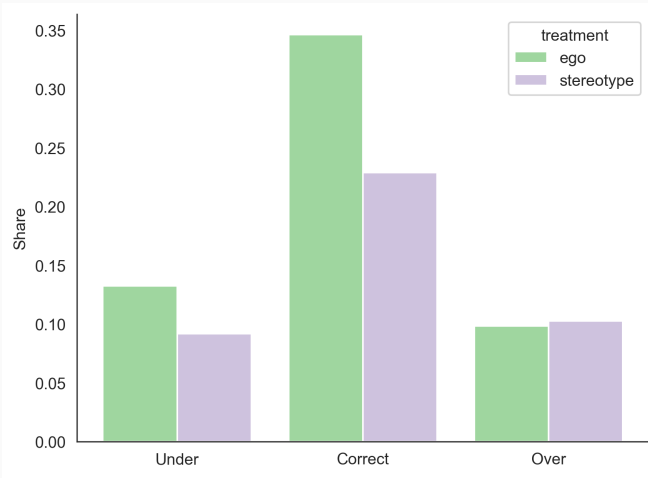


Are people falling into traps?

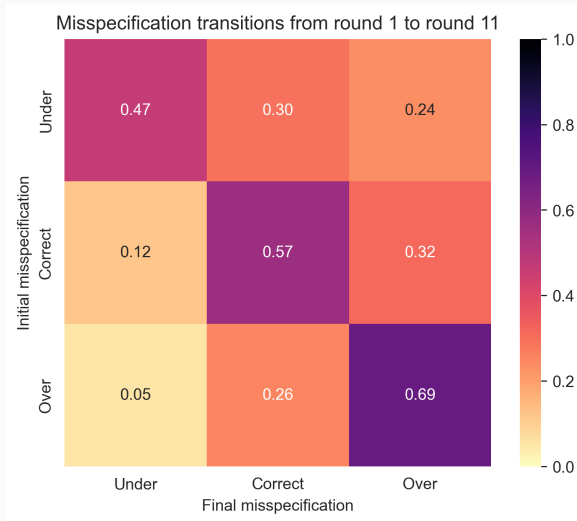


Misattributions

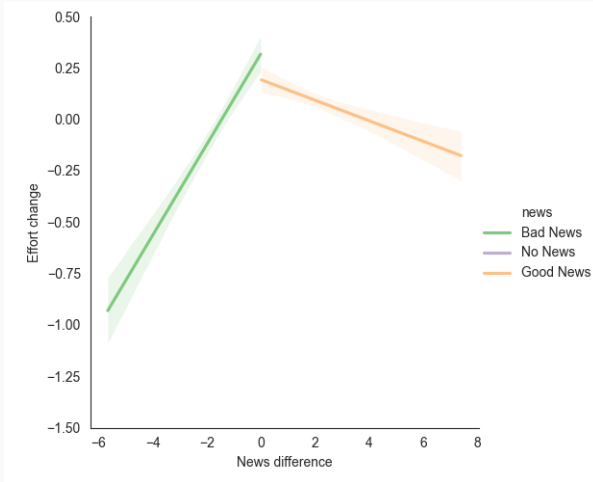
Initial Misspecifications



Transition Matrix

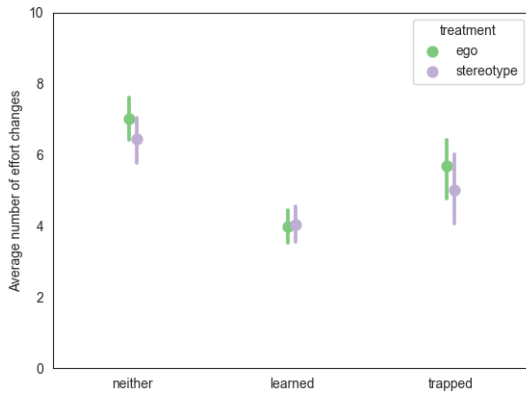


Good News v. Bad News



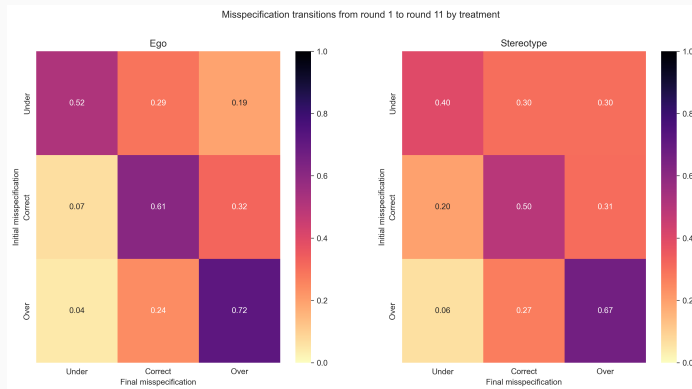
Other Reasons

Excessive Switching

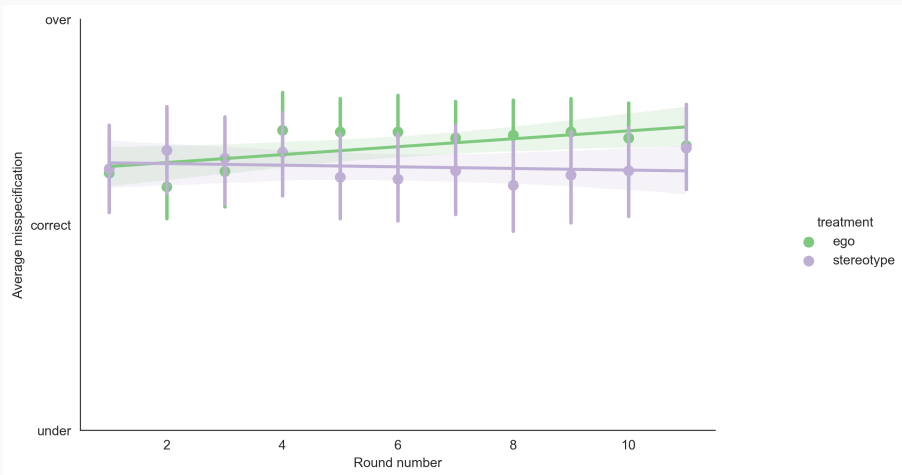


Stereotypes

Do misspecifications persist more in the Ego condition?



Is the bias more pronounced in the Ego condition?



Differences across treatments

Very slight differences across treatments

- Less stickiness in initial beliefs in Stereotype
- More attribution bias in Ego condition

Concluding Remarks

Summary

Overall:

- Traps don't seem to be the main reason for lack of learning
- Evidence pointing to misattributions
- Ego-relevance seems to play a minor role

In the presence of traps:

- 44% of subjects learn the true state
- About 20% of the subjects fall into self-defeating equilibria when they exist
- 36% of the subjects don't learn correctly and don't fall into traps

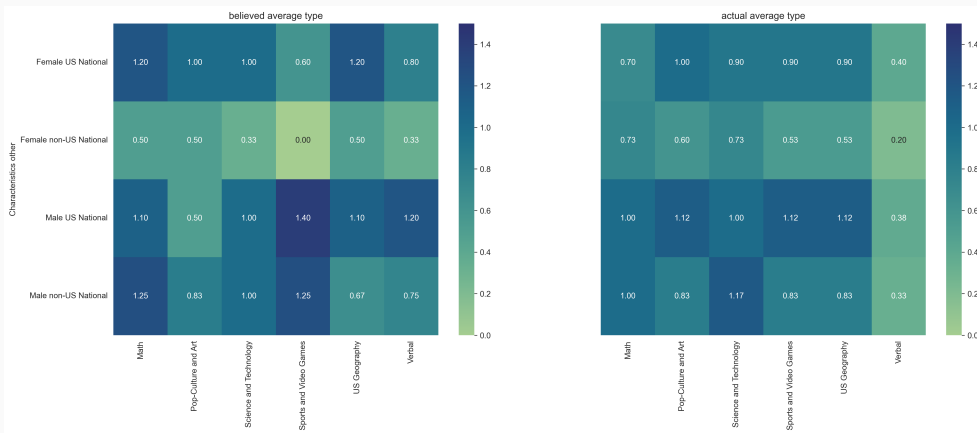
Stereotypes:

- Subjects might be self-censoring their beliefs
- Trying to correct initial biases can look like missattribution bias
- No confirmation bias

The end

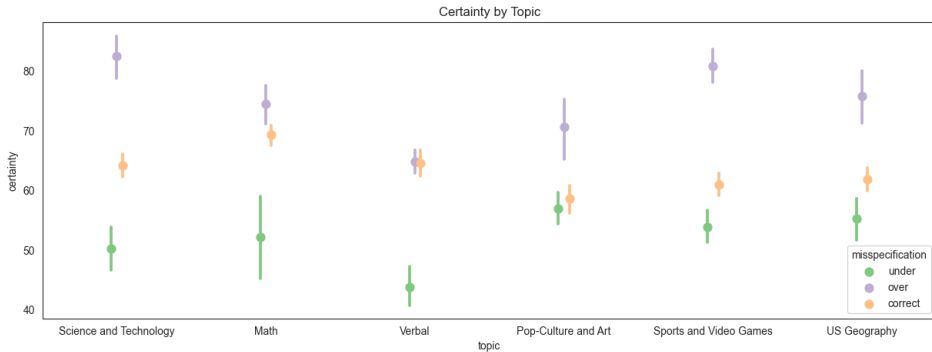
Thank you!

Misspecifications



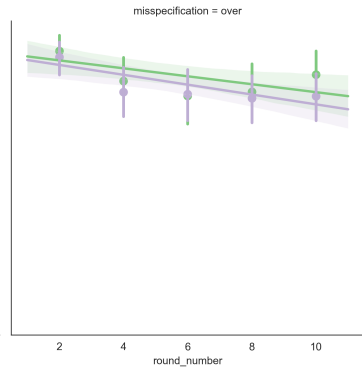
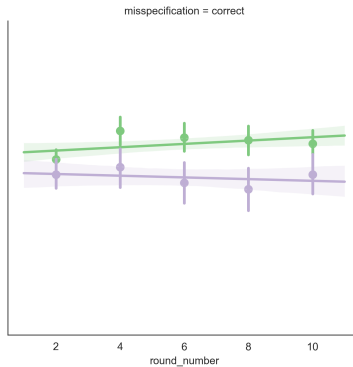
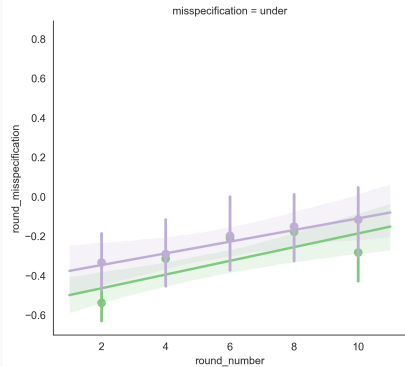
[Back](#)

Certainties



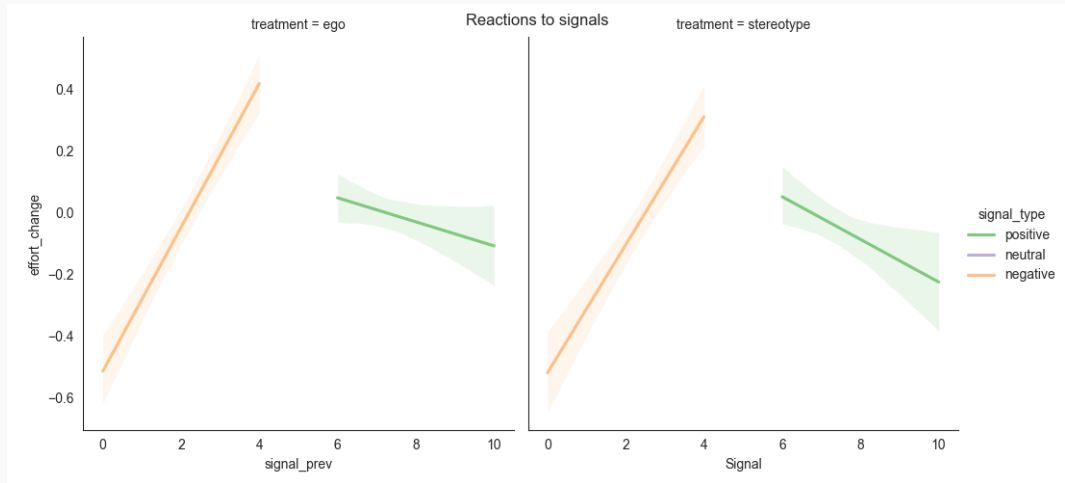
Back

Misspecification changes by treatment

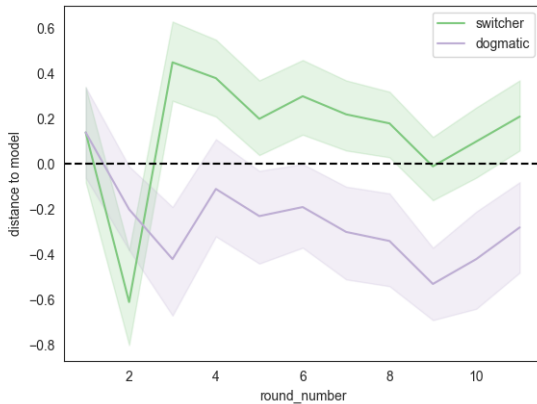


[Back](#)

Positive Signals v. Negative Signals

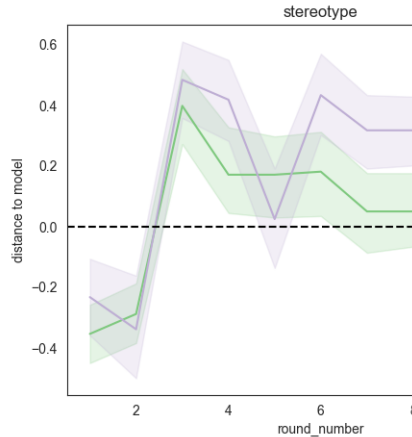
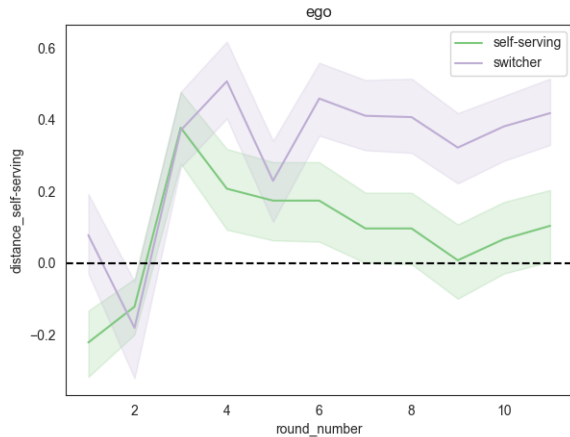


Dogmatic v. Switcher

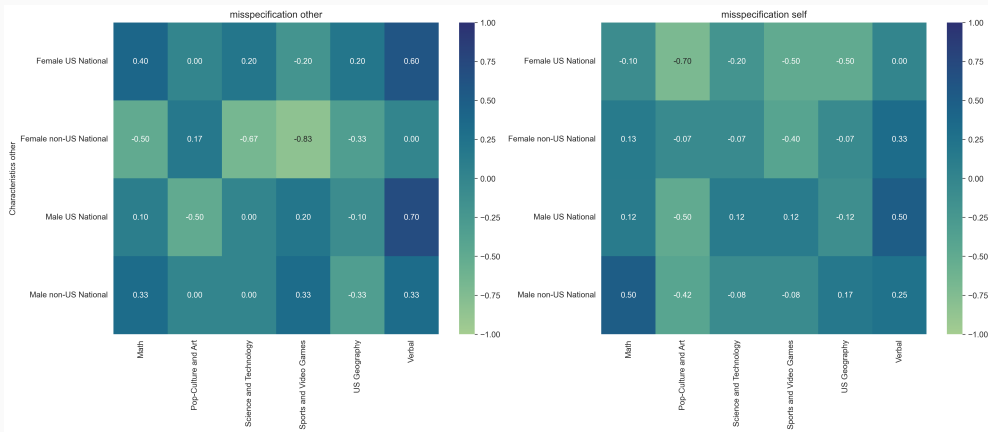


[Back](#)

Bayesian v. Self-Attribution



The Stereotypes



types

Back

Dogmatic Overconfident: Simulated

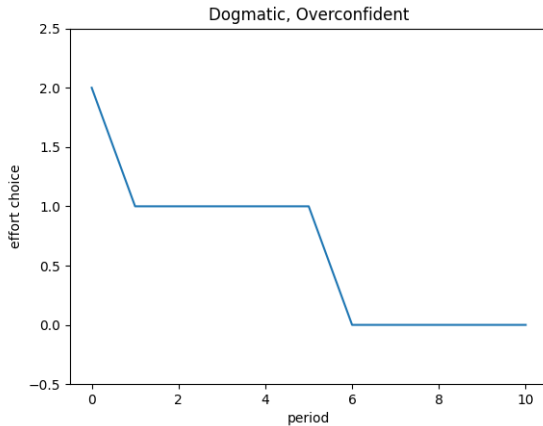


Figure 1: $\theta^* = \theta_M$, $\hat{\theta} = \theta_H$, $\omega^* = \omega_M$

Switcher Overconfident: Simulation

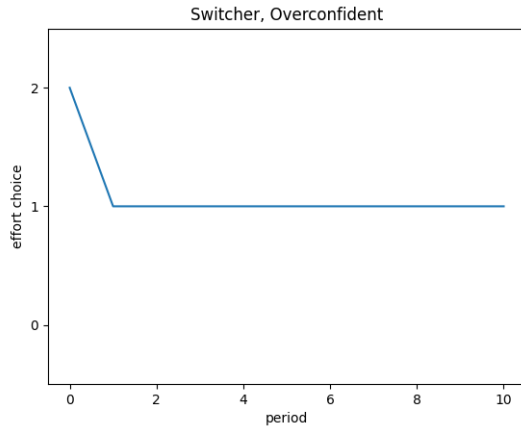
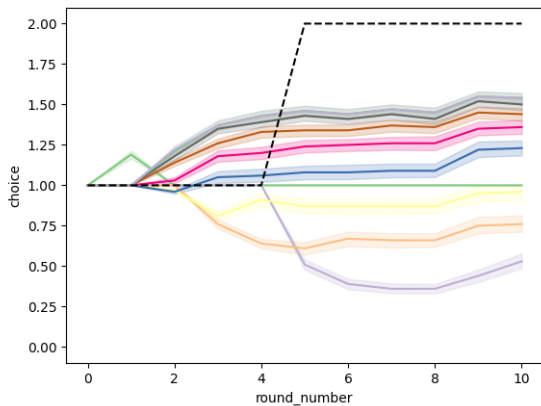


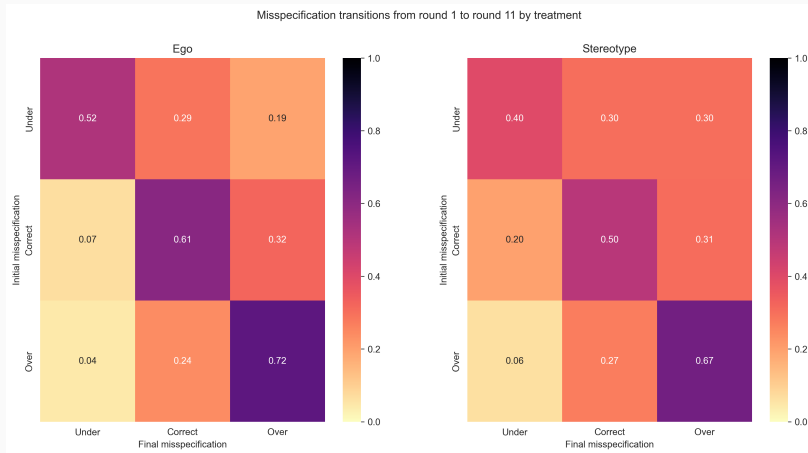
Figure 2: $\theta^* = \theta_M$, $\hat{\theta} = \theta_H$, $\omega^* = \omega_M$, $\alpha = 1.1$

Self-Attribution: Simulation



[Back](#)

Transitions treatments



Back

Subject categorization

