Trade Theory with Behavioral Agents

[Click here for the paper]

Wisarut Suwanprasert

Middle Tennessee State University

2026 ASSA Annual Meetings

Motivation

- Traditional economic theories suggest that trade liberalization generally yields net positive benefits.
- Why does a significant portion of the population remain opposed to international trade?
  - For example, the 2018 China–United States trade war, the Brexit, the 2025 US tariffs, etc.
- Standard explanations include the distributional effects of trade and pressures on the job markets.
- **This paper:**

What roles do **behavioral biases** play in distorting **consumers' perceptions** of the benefits and costs of international trade and tariffs?

What This Paper Does

- Extends standard trade models by incorporating **behavioral biases**
  - Captures various forms of behavioral biases
  - The interpretations focus on two broad types
    - ⋆ Utility misperceptions: consumption habit, home bias, externality, internality/overconsumption, addiction, myopia
    - ⋆ Price misperceptions: inattention, bounded memory, left-digit bias
- Reexamines the **gains from trade**
  - Behavioral biases can either dampen or amplify the welfare gains from trade and, in some cases, even generate welfare losses.
- Characterizes **optimal tariffs** as a **second-best** instrument
  - Small economies: tariffs as corrective tools to offset biased consumption
  - Large countries: trade-offs between terms-of-trade gains and behavioral distortions
- Characterizes **optimal behavioral nudges**
  - Nudges can be strategically designed to influence the terms of trade
  - Serve as an alternative or complement to tariffs

Related Literature

- **Behavioral economics and international economics**
  - Antràs (2025) highlights this as an underexplored area.
  - Loss aversion: Freund & Ozden (2008), Tovar (2009)
  - Social identity in trade preferences: Grossman & Helpman (2021)
  - Understanding of trade: Stantcheva (2022)
- **Perceptions, Beliefs, and Trade Support**
  - Information and trade attitudes: Hiscox (2006), Alfaro et al. (2023), Di Tella & Rodrik (2020)
  - Perceived gains as vague, group-based judgments: Stantcheva (2022), Grossman & Helpman (2021)
- **Behavioral Public Finance: Foundations for Optimal Taxation**
  - Ramsey and Pigou commodity taxes and Mirrlees nonlinear income tax: Farhi & Gabaix (2020)
  - Tax salience and reduced-form approaches: Chetty (2009), Chetty et al. (2009), Mullainathan et al. (2012)
  - Sin taxes and corrective taxation: Gruber & Koszegi (2001), O'Donoghue & Rabin (2006)
  - Applications to soda, energy, and health markets: Allcott et al. (2014, 2019), Dubois et al. (2020), Baicker et al. (2015)

A Basic Model of Behavioral Agents

- Based on Farhi and Gabaix (2020)
- An economy with n tradeable goods and the price vector $\boldsymbol{p} \equiv \{p_i\}$.
- Two types of agents: **Rational agent (he/him)** and **Behavioral agent (she/her)**

1. **Rational agent (he/him)**:

$$\boldsymbol{x}^r(\boldsymbol{p}, w) \text{ solves } \max u(\boldsymbol{x}) \text{ s.t. } \boldsymbol{p} \cdot \boldsymbol{x} = w.$$

   - FOC: $\nabla \boldsymbol{u}(\boldsymbol{x}^r) = \lambda \boldsymbol{p}$.

2. **Behavioral agent (she/her)**

$$\boldsymbol{x}^b(\boldsymbol{p}, w) \text{ satisfies } \boldsymbol{p} \cdot \boldsymbol{x}^b(\boldsymbol{p}, w) = w \text{ but does not necessarily maximize } u(\boldsymbol{x})$$

   - Two biases:
     - ⋆ **Utility misperception**: maximizes $u^b(\boldsymbol{x}) \neq u(\boldsymbol{x})$
     - ⋆ **Price misperception**: perceives prices as $\pi(\boldsymbol{p}, w) \neq \boldsymbol{p}$

Examples of Utility and Price Misperceptions

- **Utility Misperceptions**
  - Habit formation
  - Home bias
  - Externalities/Internalities
  - Addiction
  - Myopia
- **Price Misperceptions**
  - Inattention to true price/Reliance on defaults
  - Bounded memory
  - Inattention to true price changes
  - Inattention to taxes
  - Left-digit bias

Behavioral Wedge

- Behavioral wedge $\boldsymbol{\theta}(\boldsymbol{x}) \equiv \{\theta_i(\boldsymbol{x})\}$ is defined as

$$\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{p} - \frac{\nabla \boldsymbol{u}(\boldsymbol{x})}{v_w} \tag{1}$$

  - Each component $\theta_i(\boldsymbol{x})$ measures the normalized gap between the market price of good i and its true marginal utility, expressed in monetary terms (as normalized by v_w).

1. For a rational agent, $\boldsymbol{\theta}(\boldsymbol{x}^r) = \mathbf{0}$.
2. When $\theta_i > 0$, good i is overconsumed at the margin.
   - ⋆ The market price p_i exceeds the marginal utility per dollar, u_i / v_w
   - ⋆ The agent derives too little utility from an additional unit relative to its cost and should reduce consumption.
3. When $\theta_i < 0$, good i is underconsumed at the margin.

Behavioral Wedge

- Behavioral wedge in the presence of utility misperceptions and price misperceptions:

$$\boldsymbol{\theta} = \underbrace{\frac{\nabla u^b\left(\boldsymbol{x}^b\right)}{v_w^b} - \frac{\nabla u\left(\boldsymbol{x}^b\right)}{v_w}}_{\text{utility misperception}} + \underbrace{\boldsymbol{p} - \frac{\pi\left(\boldsymbol{p}, w\right)}{\pi\left(\boldsymbol{p}, w\right) \cdot \boldsymbol{x}_w^b}}_{\text{price misperception}}.$$

- When $\theta_i > 0$, i.e., good i is overconsumed at the margin
  - Utility misperception: $\frac{\nabla u^b\left(\boldsymbol{x}^b\right)}{v_w^b} > \frac{\nabla u\left(\boldsymbol{x}^b\right)}{v_w}$
    - ⋆ Misperceived the marginal utility to be larger than the true marginal utility
  - Price misperception: $\frac{\pi_i}{\pi(\boldsymbol{p},w)\cdot \boldsymbol{x}_w^b} < p_i$
    - ⋆ Misperceived the price to be cheaper than what it actually is

Behavioral Roy's Identity

- For **Rational** Agents:

$$\frac{\partial v^r(\boldsymbol{p}, w)}{\partial p_j} = -x_j^r v_w^r.$$

  - Pure income effect
  - No substitution effect (by the envelope theorem)

- For **Behavioral** Agents:

$$\frac{\partial v^b(\boldsymbol{p}, \boldsymbol{\pi}(\boldsymbol{p}, w), w)}{\partial p_j} = \left(-x_j^b - \boldsymbol{\theta} \cdot \boldsymbol{S}_j^b\right) v_w^b,$$

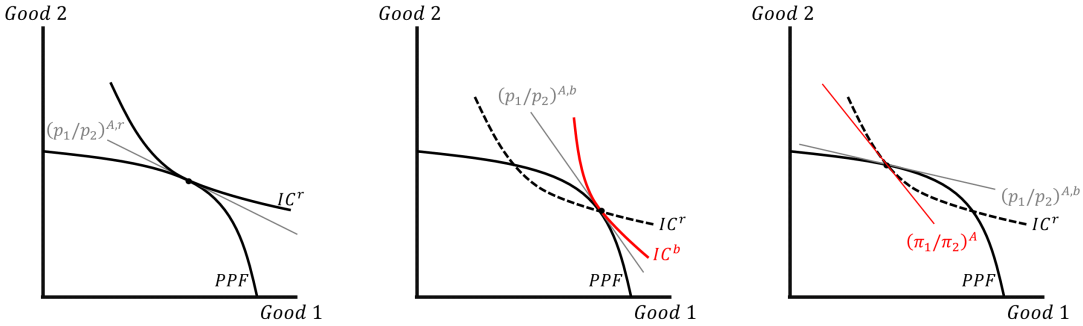
  where $\boldsymbol{S}_j^b$ is the j-th column of the behavioral income-compensated Slutsky matrix $\boldsymbol{S}^b$

  - Income effect: $-x_j^b v_w^b$
  - Substitution effect: $-\boldsymbol{\theta} \cdot \boldsymbol{S}_j^b v_w$
    - ⋆ a change in the price of good j changes consumption according to $\boldsymbol{S}_j^b$ and causes welfare losses by $\boldsymbol{\theta} \cdot \boldsymbol{S}_j^b v_w$.

A Classical Trade Model

Pattern of Trade

Figure: Autarkic equilibria under different scenarios



(a) a rational agent

(b) a behavioral agent with utility misperception

(c) a behavioral agent with price misperception

Gains from trade

Proposition

Welfare gains that arise from a change in world prices $\boldsymbol{Dp}^W$ *can be expressed as*

$$dv\left(\boldsymbol{p}^W, \boldsymbol{p}^W \cdot \boldsymbol{y}\right) = \underbrace{v_w\left(\boldsymbol{y} - \boldsymbol{x}\right) \cdot \boldsymbol{Dp}^W}_{\text{the traditional gains from trade}} + \underbrace{v_w\left(-\boldsymbol{\theta}^T \boldsymbol{S}^b\right) \cdot \boldsymbol{Dp}^W}_{\text{distortions from behavioral biases}} .$$

Gains from Trade

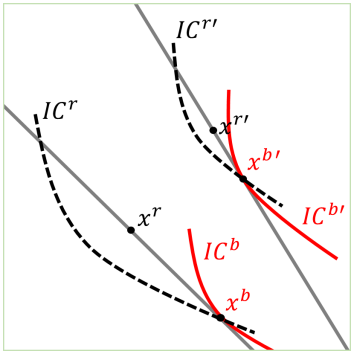
1. Traditional gains from trade
   - gains from specialization
   - gains from exchange
2. Distortions from behavioral biases

Lesson

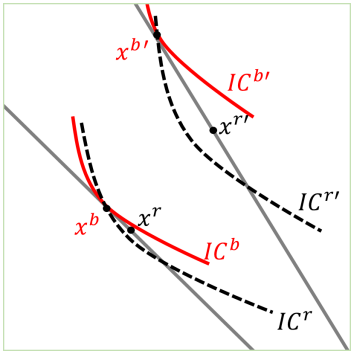
- Behavioral agents may realize larger or smaller welfare gains depending on the nature and extent of their biases.

Gains from trade

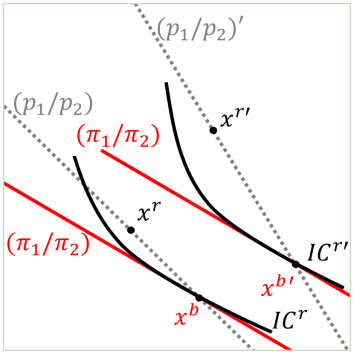
Figure: A change in market equilibrium when the price ratio increases from (p_1/p_2) to $(p_1/p_2)'$ under behavioral agents.



(a) a behavioral agent with utility misperception



(b) a behavioral agent with utility misperception



(c) a behavioral agent with price misperception

Deviating from Autarky

Rational Agent:

$$dv\left(\boldsymbol{p}^W, \boldsymbol{p}^W \cdot \boldsymbol{y}\right)\Big|_{\boldsymbol{p}^W=\boldsymbol{p}^A} = 0.$$

- Autarky represents a welfare minimum: Any deviation $\rightarrow$ gains from trade.

Behavioral Agent:

$$dv\left(\boldsymbol{p}^W, \boldsymbol{p}^W \cdot \boldsymbol{y}\right)\Big|_{\boldsymbol{p}^W=\boldsymbol{p}^A} = -v_w \boldsymbol{\theta}^T \boldsymbol{S}^b \cdot \boldsymbol{D}\boldsymbol{p}^W.$$

- Welfare may not be minimized at autarky
- Possibility of welfare losses from trade

Proposition

If $v_w \boldsymbol{\theta}^T \boldsymbol{S}^b \cdot \boldsymbol{D}\boldsymbol{p}^W \neq 0$, there is a non-empty set of world price vectors such that a country that has a behavioral agent experiences welfare losses from trade when the world prices fall into that set.

Optimal Tariffs

Model Setup:

- Two-country, two-good world
- Good 2 is the numeraire
- Home is a natural importer of good 1
- Domestic price with tariff: $p = (1 + \tau)\, p^w$
- Welfare: $v\left(p, R + \left(p - p^w\right)\left(x_1 - y_1\right)\right)$

Rational Agent:

$$\tau^* = \frac{1}{\varepsilon_1^*}$$

where ε_1^* is the elasticity of foreign export supply.

- Johnson's (1953) optimal tariffs
- Balances domestic distortion with terms-of-trade gain

Optimal Tariffs with Behavioral Agents

Behavioral Agent:

$$\underbrace{\left(-\boldsymbol{\theta}\cdot\boldsymbol{S}_1^b\right)\frac{dp}{d\tau}}_{\text{behavioral biases}}+\underbrace{\tau^* p^w\frac{dm_1}{d\tau}}_{\text{distortions}}+\underbrace{\left(-m_1\frac{dp^w}{d\tau}\right)}_{\text{terms-of-trade manipulation}}=0.$$

Implications:

- Optimal tariffs now also correct behavioral inefficiencies
- Tariffs can reduce (or worsen) over-/underconsumption

Optimal tariffs:

$$\tau^*=\frac{1}{\varepsilon_1^*}+\frac{\boldsymbol{\theta}^b\cdot\boldsymbol{S}_1^b}{p^w\left(dm/dp\right)}$$

- The optimal tariff of a small country with behavioral agents is non-zero.

Optimal Tariffs with Behavioral Agents

Interpretations – Large Countries:

1. Policymakers may use tariffs to correct behavioral biases, even at the expense of traditional welfare considerations.
2. Policymakers must recognize an additional welfare consequence introduced by behavioral biases when attempting to manipulate the terms of trade through tariffs.

Interpretations – Small Countries:

- The optimal tariff of a small open economy with a behavioral representative consumer is

$$\tau^* = \frac{\boldsymbol{\theta}^b \cdot \boldsymbol{S}_1^b}{p^w \left(dm/dp \right)} \neq 0.$$

- In the presence of behavioral biases, trade policy may function as a second-best instrument aiming at welfare losses attributable to these biases.

Nudges

Nudges

- Introduced by Thaler and Sunstein (2008).
- Examples:
  - Reducing food waste through plate size adjustments (Kallbekken and Salen, 2013)
  - Increasing loan demand with attractive advertisements (Bertrand et al., 2010)
  - Raising organ donation rates by making it a default option (Johnson and Goldstein, 2003)

Nudges in the Model:

- Modeled by continuous variable called **nudge intensity** $\eta \in [0, 1]$
- Nudges affect behavior without altering the feasible set of choices.
- Welfare: $v\left(p, R + \left(p - p^w\right)\left(x_1\left(\eta\right) - y_1\right)\right) - c\left(\eta\right)$

Optimal Nudges

Proposition

At an interior optimum, the optimal nudge and tariff of an economy characterized by a behavioral representative consumer satisfy

$$\underbrace{-\left(\theta_1 \frac{\partial x_1}{\partial \eta} + \theta_2 \frac{\partial x_2}{\partial \eta}\right)}_{\text{behavioral biases}} + \underbrace{\tau^* p^w \frac{\partial m_1}{\partial \eta}}_{\text{distortions}} + \underbrace{\left(-m_1 \frac{dp^w}{d\eta}\right)}_{\text{terms-of-trade manipulation}} = c'(\eta),$$

$$\underbrace{-\left(\theta_1 \frac{\partial x_1}{\partial p} + \theta_2 \frac{\partial x_2}{\partial p}\right) \frac{dp}{d\tau}}_{\text{behavioral biases}} + \underbrace{\tau^* p^w \frac{dm_1}{d\tau}}_{\text{distortions}} + \underbrace{\left(-m_1 \frac{dp^w}{d\tau}\right)}_{\text{terms-of-trade manipulation}} = 0$$

What do nudges do?

1. Increase consumption ($\partial x_i / \partial \eta > 0$) of underconsumed goods ($\theta_i < 0$), or decrease consumption ($\partial x_i / \partial \eta < 0$) of overconsumed goods ($\theta_i > 0$).
2. Influence tariff revenue by distorting the demand for imported goods.
3. Influence global demand patterns, thereby affecting world prices.

Discussion 1: Why Don't People Feel the Gains from Trade?

1. **Utility Misperceptions**:
   - **Habit formation**: Preference for familiar brands prevents switching to better imported options.
   - **Home bias**: Emotional or patriotic attachment to domestic goods overshadows superior imports.
   - **Neglected externalities/internalities**: Consumers overlook health or environmental benefits from trade.
   - **Myopia**: Short-term cost concerns may obscure long-term savings from high-quality imports.
2. **Price Misperceptions**:
   - **Inattention to true prices**: Consumers miss actual price decreases.
   - **Bounded memory**: Outdated price beliefs persist even after tariff cuts.
   - **Partial attention to price changes**: Consumers underestimate actual savings from trade liberalization.
   - **Inattention to taxes**: Tariff reductions go unnoticed.
   - **Left-digit bias**: Small price drops are perceived as negligible.

Summary: Even when trade delivers real benefits, behavioral biases prevent individuals from fully perceiving them.

Discussion 2: Behavioral Biases and Public Support for Protectionism

Behavioral economics helps explain public support for protectionism.

Possible explanation:

- Misperceived utility gains: Consumers do not associate everyday improvements with trade.
- Salience of costs: Gains are diffuse and delayed, while costs like job losses are salient and immediate.
- Present bias: Short-term disruptions outweigh long-term efficiency gains.

Examples:

1. **Case 1: The 2018 U.S.–China Trade War**
   - Americans undervalued benefits from cheap Chinese imports.
   - Focused on job losses, not lower cost of living.
2. **Case 2: Brexit**
   - UK consumers underappreciated price and variety gains from EU trade.
   - Emotional narratives overpowered economic logic.

Conclusion

- This study extends traditional trade theory by incorporating behavioral biases—such as utility and price misperceptions—into a unified framework.
- It shows that gains from trade can be amplified or diminished depending on behavioral distortions.
- Optimal tariffs are reinterpreted not only as tools for terms-of-trade manipulation but also as corrective instruments for welfare losses arising from behavioral inefficiencies.
- A joint framework of optimal tariffs and nudges is developed.
- This study offers an explanation for why individuals often fail to recognize the benefits of trade, drawing on examples such as the 2018 U.S.–China trade war, Brexit, and the 2025 U.S. tariffs.