Trade Theory with Behavioral Agents*

Wisarut Suwanprasert†
Middle Tennessee State University

December 25, 2025

Abstract

In this paper, I develop a theoretical framework to analyze gains from trade and optimal tariffs in the presence of behavioral biases. The theoretical results center on a sufficient statistic—the behavioral wedge—which captures distortions induced by a broad class of behavioral biases. I present three main theoretical results. First, I show how behavioral biases can either dampen or amplify the welfare gains from trade and, in some cases, even generate welfare losses. Second, I characterize optimal tariffs in this context and show how they can be used to correct distortions arising from behavioral biases. Third, I characterize optimal behavioral nudges and show how they may be used to manipulate the world's terms of trade. Finally, I discuss the potential role of behavioral biases in shaping public support for the 2018 China–United States trade war and Brexit.

Keywords: trade theory, behavioral economics, gains from trade, optimal tariffs, nudges.
JEL classification numbers: D9, F1, H2

*I thank Carlos Góes, Pimwilai Kijjanapanich, Raymond Riezman, and seminar participants at the Midwest International Trade Conference (Spring 2025), and the PIER Research Workshop 2025 for their valuable comments and discussions. All remaining errors are our own.

†Department of Economics and Finance, Jennings A. Jones College of Business, Middle Tennessee State University, Tennessee, USA. Email: wisarut.suwanprasert@mtsu.edu.

1 Introduction

In recent years, pivotal events such as the 2018 China–United States trade war and the United Kingdom's exit from the European Union, commonly referred to as Brexit, have raised fundamental questions about the public's persistent skepticism toward international trade. Standard trade theory predicts that trade liberalization yields aggregate gains. Yet, opposition to trade remains widespread. Conventional explanations, such as the distributional effects of trade and pressures on the job market, certainly play a role, as highlighted by findings in Stantcheva (2022). However, this paper proposes that one possible explanation lies in behavioral economics: behavioral biases may distort consumers' perceptions of the benefits and costs of international trade and tariffs.

This paper argues that behavioral biases might significantly influence consumers' understanding and acceptance of welfare gains from trade and the welfare costs of tariffs. For instance, consumers who are habitually inclined towards certain consumption patterns, default to known products, or display a bias for domestic goods may not fully capitalize on the economic advantages of cheaper imported goods, thus failing to recognize the full extent of gains from trade. Similarly, consumers who are inattentive to the welfare losses from tariffs may disproportionately weigh perceived job risks, opting for trade policies that inefficiently protect domestic employment. Such choices, however, are often made without a full understanding of the welfare losses incurred through tariffs, which can lead to higher prices and reduced social welfare.

This paper extends standard international trade theory by incorporating insights from behavioral economics. Building on the unified framework of Farhi and Gabaix (2020), which accommodates a broad class of behavioral biases, I analyze the gains from trade and optimal tariffs in economies populated by behavioral agents. The analysis focuses on two broad categories of behavioral biases. The first involves utility misperceptions, where the agent's perceived utility differs from her true utility. This category includes phenomena such as consumption habits, home bias, internalities, addiction, and myopia. The second involves price misperceptions, in which agents misperceive or misprocess price information—for example, due to inattention, bounded memory, or left-digit bias.

The theoretical results center on a sufficient statistic—the behavioral wedge—which characterizes the extent to which behavioral agents depart from rational behavior, conditional on the same economic environment. I then derive a behavioral analogue of Roy's identity, which decomposes the welfare effect of a price change into two components: (i) a direct income effect, consistent with traditional Roy's identity, and (ii) an indirect substitution effect, unique to the behavioral agent and captured by the behavioral wedge. The latter reflects the substitution effect that arises from suboptimal consumption adjustments driven by behavioral biases. This effect is absent under rational behavior—an implication of the envelope theorem. It captures how the behavioral agent reallocates consumption differently from a rational agent in response to price changes.

Examining welfare gains from trade in the presence of behavioral biases reveals that such biases can either amplify or diminish those gains—and in some cases, may even generate welfare losses. For example, if a behavioral agent fails to internalize the health costs of excessive sugar consumption, the availability of cheaper imported sugar may exacerbate overconsumption, reducing or even reversing the welfare gains traditionally associated with trade. In such cases, the welfare effect of sugar imports could be smaller than predicted by standard models, or even negative.

Furthermore, I propose a theory of optimal tariffs for the behavioral agent by combining behavioral biases with the optimal tariff theory of Johnson (1953). In large open economies, tariffs can be used not only to manipulate the terms of trade but also to correct inefficiencies arising from behavioral biases. Even in small open economies—where terms-of-trade manipulation is not feasible—tariffs can serve as second-best policy instruments to mitigate welfare losses caused by behavioral biases.

The paper also considers "nudges" as policy instruments aimed at correcting behavioral biases and characterizes the joint design of optimal nudges and tariffs. This approach provides a more nuanced perspective on trade policy within a behavioral framework. I show that, even in the absence of behavioral distortions, nudges may remain welfare-improving by strategically influencing the world's terms of trade.

The main contribution of this paper is to integrate behavioral economics and international economics. Despite the growth of behavioral economics, the theory of behavioral agents has yet to be fully developed in the context of international trade. The field of international trade has been known for its rigorous and complex models, with the assumption of fully rational agents. These models typically depict consumers as capable of accurately determining their consumption and responding appropriately to price changes resulting from trade liberalization.

Some studies have indicated the presence of behavioral biases in this field. For instance, Freund and Ozden (2008) emphasize the importance of loss aversion and reference dependence in shaping people's preferences over trade policy. Tovar (2009) finds that loss aversion can explain the trade protections afforded to declining industries and the existence of an anti-trade bias in trade policy. Grossman and Helpman (2021) study equilibrium tariffs with social identification.

Furthermore, this paper extends the behavioral public finance framework to analyze import taxes, a relatively underexplored area compared to the more commonly studied income, sales, and sin taxes. Farhi and Gabaix (2020) investigate Ramsey and Pigou commodity taxes and Mirrlees non-linear income tax in the presence of behavioral biases. Chetty (2009) and Chetty et al. (2009) explore the salience effects of commodity taxes, while Mullainathan et al. (2012) introduce a reduced-form approach to evaluate optimal taxes in various settings. O'Donoghue and Rabin (2006) study optimal sin taxes in the presence of self-control problems. Gruber and Koszegi (2001) focus on optimal cigarette taxes in light of internalities. Allcott et al. (2019) and Dubois et al. (2020) study optimal soda tax in the presence of overconsumption. Allcott et al. (2014) analyze optimal energy tax with externalities and internalities. Baicker et al. (2015) study the optimal design problems in health insurance under behavioral biases and moral hazard problems.

The third aspect is related to perceptions related to the effects of trade liberalization. Hiscox (2006) shows that giving respondents information about job losses due to trade decreases their support for free trade; telling them that trade reduces prices does not change their views. Alfaro et al. (2023) show that telling respondents about research findings on the job losses or gains from trade or price effects of trade or tariffs can change people's views on trade. Di Tella and Rodrik (2020) find that trade-related shocks, especially when in the form of outsourcing to a developing country, generate more demand for protectionism. This sentiment played a significant role in both the US-China tariff war and Brexit, where economic nationalism and the desire to protect domestic industries and jobs from global competition were central themes driving these events. A survey by Stantcheva (2022) concludes that respondents perceive consumer gains from trade as vague and diffuse. Grossman and Helpman (2021) find that people's views on trade draw from the status of groups they identify with.

The rest of the paper is organized as follows. Section 2 develops a basic model of behavioral agents. Section 3 introduces a classical trade model with behavioral agents and explores gains from trade. Section 4 studies the government's optimal interventions, such as optimal tariffs and optimal nudges. Section 5 discusses implications of the findings in this paper. Section 6 concludes.

2 A Basic Model of Behavioral Agents

This section presents a detailed model outlining various behavioral biases and explores how these biases alter the standard outcomes in price theory.

2.1 Model Setup

The initial part of this analysis delineates between a rational agent (he/him) and a behavioral agent (she/her) within an economy.[1]

The economy consists of n tradeable goods, indexed by $1, \ldots, n$, with a corresponding $n \times 1$ price vector $\boldsymbol{p} \equiv \{p_i\}$. Both rational and behavioral agents share the same underlying preferences, captured by a "true" utility function $u(\boldsymbol{x})$, where $\boldsymbol{x} \equiv \{x_i\}$ denotes an $n \times 1$ vector of consumptions bundles.

The rational agent selects a consumption bundle $\boldsymbol{x}^r(\boldsymbol{p}, w)$ to maximize utility $u(\boldsymbol{x})$ subject to the standard budget constraint $\boldsymbol{p} \cdot \boldsymbol{x}^r(\boldsymbol{p}, w) = w$, where w is income. The solution satisfies the first-order condition $\nabla \boldsymbol{u}(\boldsymbol{x}^r) = \lambda \boldsymbol{p}$ for some Lagrange multiplier $\lambda > 0$, where $\nabla \boldsymbol{u}(\boldsymbol{x})$ is the gradient vector of $u(\boldsymbol{x})$ evaluated at a consumption vector $\boldsymbol{x}$. The agent's welfare is captured by the indirect utility function $v^r(\boldsymbol{p}, w) = u(\boldsymbol{x}^r(\boldsymbol{p}, w))$.

In contrast, the behavioral agent suffers from behavioral biases. While her consumption bundle $\boldsymbol{x}^b(\boldsymbol{p}, w)$ satisfies the same budget constraint, $\boldsymbol{p} \cdot \boldsymbol{x}^b(\boldsymbol{p}, w) = w$, her choice does not necessarily maximize her true utility $u(\boldsymbol{x})$. In this framework, I consider two distinct types of behavioral biases: utility misperception and price misperception. In the first case, the behavioral agent maximizes the "perceived" utility function $u^b(\boldsymbol{x})$, instead of the true utility function $u(\boldsymbol{x})$. In the second case, the behavioral agent fails to keep track of the true prices $\boldsymbol{p}$ and misperceives the prices as $\boldsymbol{\pi}(\boldsymbol{p}, w)$, a function of the true prices $\boldsymbol{p}$ and her income w. That is, the behavioral agent maximizes

[1] My daughter deserves to be the protagonist in my model, and it is my honor to take the supporting role.

a misperceived utility function $u^b(\boldsymbol{x})$ based on misperceived prices $\boldsymbol{\pi}(\boldsymbol{p}, w)$, but the actual utility received is the "true" utility $u(\boldsymbol{x})$.

The behavioral agent's demand function $\boldsymbol{x}^b(\boldsymbol{p}, \boldsymbol{\pi}(\boldsymbol{p}, w), w)$ align with the condition $\nabla \boldsymbol{u}^b(\boldsymbol{x}^b) = \lambda^b \boldsymbol{\pi}$, where $\nabla \boldsymbol{u}^b(\boldsymbol{x})$ is the gradient vector of $u^b(\boldsymbol{x})$ at a consumption vector $\boldsymbol{x}$ and $\lambda^b > 0$ is the Lagrange multiplier associated with the budget constraint $\boldsymbol{p} \cdot \boldsymbol{x}^b(\boldsymbol{p}, \boldsymbol{\pi}(\boldsymbol{p}, w), w) = w$. This condition implies that the behavioral agent maximizes her behavioral utility functions by equating the "perceived" marginal rates of substitution with "perceived" relative prices. While the behavioral agent misperceives prices, the Lagrange multiplier λ^b adjusts to guarantee that her budget constraint is satisfied. The behavioral agent's perceived utility can be measured by the perceived indirect utility function $v^b(\boldsymbol{p}, \boldsymbol{\pi}(\boldsymbol{p}, w), w) = u^b(\boldsymbol{x}^b(\boldsymbol{p}, \boldsymbol{\pi}(\boldsymbol{p}, w), w))$.

2.2 Examples of Behavioral Biases

This section gives examples of various types of behavioral biases that can be considered in the framework in Section 2.1.

2.2.1 Utility Misperceptions

Consumption habit: A behavioral agent may have a habitual consumption pattern, deviating from which incurs a utility cost. An example of a perceived utility function for a behavioral agent with consumption habits is $u(\boldsymbol{x}) - c(\boldsymbol{x} - \boldsymbol{x}^{\text{habit}})$, where $u(\boldsymbol{x})$ is the true utility function and $c(\boldsymbol{x} - \boldsymbol{x}^{\text{habit}})$ represents the mental cost of deviating from habitual consumption $\boldsymbol{x}^{\text{habit}} \equiv \{x_i^{\text{habit}}\}$.

Home bias: A behavioral agent may display a preference for domestic goods, perceiving enhanced utility from domestic products over foreign ones. Suppose that the true utility function is $u(\boldsymbol{x}_{\text{home}}, \boldsymbol{x}_{\text{foreign}})$, where $\boldsymbol{x}_{\text{home}}$ and $\boldsymbol{x}_{\text{foreign}}$ denote vectors of domestic and foreign goods consumption, respectively. An example of a perceived utility function for a behavioral agent with home bias is $u(\alpha \boldsymbol{x}_{\text{home}}, \boldsymbol{x}_{\text{foreign}})$, where $\alpha > 1$ represents the degree of home bias.

Externality: A behavioral agent may neglect the environmental costs of their consumption, leading to a distorted perception of their utility. Suppose that the true utility function is $u(x_1, x_2) - e(x_1)$, where $e(x_1)$ is the environmental cost of consuming x_1. A behavioral agent may perceive the utility as $u(x_1, x_2)$, omitting the environmental cost.

Internality/overconsumption: A behavioral agent may overlook the health costs associated with certain goods, such as tobacco or alcohol. Suppose that the true utility function is $u(x_1, x_2) - h(x_1)$, where $h(x_1)$ is the health cost of consuming x_1. A behavioral agent may perceive the utility as $u(x_1, x_2)$, ignoring these health costs.

Addiction: A behavioral agent may feel compelled to consume a minimum quantity of a specific good, regardless of its utility impact. Suppose that the true utility function is $u(x_1, x_2)$. A behavioral agent with addiction perceives the utility as if it comes with the constraint $x_1 \geq \overline{x}_1$, where $\overline{x}_1$ is the minimum required consumption of the addictive good.

Myopia: A behavioral agent may be myopic, undervaluing long-term costs in favor of immediate utility. Suppose that the true utility function is $u(x_1, x_2) - \beta c(x_1)$, where $c(x_1)$ represents the future cost of consuming x_1 and β is the true discount rate. A behavioral agent may perceive the utility as $u(x_1, x_2) - \delta\beta c(x_1)$, where $\delta < 1$ represents the degree of myopia.

2.2.2 Price Misperceptions

Inattention to true price/Reliance on defaults: A behavioral agent may be limited by cognitive resources and rely on default or average prices instead of incorporating real-time price information. This behavioral tendency can be attributed to factors such as mental fatigue or limited memory capacity. As a result, a behavioral agent may make decisions under the assumption that prices align with their default or average levels despite potential deviations in the actual market.

For example, a behavioral agent may perceive price p_1 as $\pi_1 = \mathbb{E}(p_1)$.

Bounded memory: This cognitive limitation impedes behavioral agents from precisely tracking fluctuating prices. Wilson (2014) models bounded memory as behavioral agents using finite-state automation to summarize price information. That is, a behavioral agent may categorize price information into basic descriptors such as "low" or "high," and decision-making is then based on these simplified classifications.

For example, a behavioral agent may perceive price p_1 as $\pi_{1,\text{low}} = \mathbb{E}(p_1 | p_1 < \overline{p})$.

Inattention to true price changes: A behavioral agent may not fully perceive the extent of price reductions, underestimating the actual change due to partial attention.

For example, a behavioral agent may perceive price p_1 as $\pi_1 = \overline{p}_1 + (1-\delta)(p_1 - \overline{p}_1)$, where $\overline{p}_1$ is the initial price and $\delta \in [0, 1]$ represents the degree of inattention to price changes. In this case, $\delta = 0$ indicates full attention and 1 is complete inattention

Inattention to taxes: When tariffs affect the prices of goods, a behavioral agent may not fully recognize the price increase, perceiving a lower impact due to limited attention to the tax component (Chetty et al., 2009).

Suppose that a tariff τ raises the price from $\overline{p}_1$ to $\overline{p}_1 + \tau$. A behavioral agent may perceive the new price as $\pi_1 = \overline{p}_1 + (1-\delta)\tau$, where $\delta \in [0, 1]$ represents the degree of inattention to taxes. In this case, $\delta = 0$ indicates full attention and 1 is complete inattention

Left-digit bias: This behavioral bias illustrates a psychological tendency where a behavioral agent disproportionately focuses on the leftmost digits of a price, leading to an inaccurate perception of the overall price.

Suppose that a price is quoted as $p_1 = n + r$, where $n \in \mathbb{N}$ is the integer part of the price and $r \in [0, 1)$ is the fractional remainder. A behavioral agent may perceive this price as $\pi_1 = n + \delta r$, where $\delta \in [0, 1)$ represents the degree of left-digit bias.

2.3 Behavioral Wedge: A Sufficient Statistic for Behavioral Bias

To capture how behavioral biases distort economic decisions, it is useful to construct a single object that summarizes deviations from rational optimization. Motivated by the envelope theorem, a behavioral wedge is introduced as a sufficient statistic that quantifies the marginal distortion in consumption choices induced by behavioral biases. The wedge isolates the welfare-relevant differences between observed behavior and the rational benchmark in a tractable and interpretable way.

In the standard utility maximization problem, a rational agent chooses a consumption bundle $\boldsymbol{x}^r$ that satisfies the first-order condition $\nabla \boldsymbol{u} (\boldsymbol{x}^r) = \lambda \boldsymbol{p}$, where $\lambda > 0$ is the Lagrange multiplier associated with the budget constraint. This multiplier corresponds to the marginal value of income, given by $v_w = \nabla \boldsymbol{u} (\boldsymbol{x}^r) \cdot \boldsymbol{x}^r_w$.

Rewriting the first-order condition yields the normalized condition:

$$\frac{\nabla \boldsymbol{u} (\boldsymbol{x}^r)}{v_w} = \boldsymbol{p},$$

which equates the marginal utility vector—expressed in monetary terms by normalizing with the marginal value of income v_w—to the vector of market prices.

In contrast, a behavioral agent chooses consumption $\boldsymbol{x}^b$ that maximizes a possibly misperceived utility function under potentially misperceived prices. As a result, the normalized gradient of the true utility at $\boldsymbol{x}^b$, $\nabla \boldsymbol{u} \left(\boldsymbol{x}^b\right) / v_w$, may no longer coincide with the true price vector $\boldsymbol{p}$.

This deviation motivates the construction of the behavioral wedge, denoted by $\boldsymbol{\theta} (\boldsymbol{x}) \equiv \{\theta_i (\boldsymbol{x})\}$, and 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). For a rational agent who maximizes $u (\boldsymbol{x})$ subject to the budget constraint, the first-order condition implies that this wedge is identically zero: $\boldsymbol{\theta} (\boldsymbol{x}^r) = \boldsymbol{0}$.

When $\theta_i > 0$, the market price p_i exceeds the marginal utility per dollar, u_i / v_w. In this case, good i is overconsumed at the margin: the agent derives too little utility from an additional unit relative to its cost and should reduce consumption. Conversely, when $\theta_i < 0$, the marginal utility exceeds the price, indicating that good i is underconsumed; the agent should consume more, as the good offers more value than it costs. Thus, the behavioral wedge provides a directional measure of inefficiency in marginal consumption decisions.

In the model with misperceptions in both utility and price, the behavioral wedge defined in equation (1) can be explicitly decomposed as:

$$\boldsymbol{\theta} = \underbrace{\frac{\nabla u^b \left(\boldsymbol{x}^b \left(\boldsymbol{p}, \boldsymbol{\pi}\left(\boldsymbol{p}, w\right), w\right)\right)}{v_w^b \left(\boldsymbol{p}, \boldsymbol{\pi}\left(\boldsymbol{p}, w\right), w\right)} - \frac{\nabla u \left(\boldsymbol{x}^b \left(\boldsymbol{p}, \boldsymbol{\pi}\left(\boldsymbol{p}, w\right), w\right)\right)}{v_w \left(\boldsymbol{p}, w\right)}}_{\text{utility misperception}} + \underbrace{\boldsymbol{p} - \frac{\boldsymbol{\pi}\left(\boldsymbol{p}, w\right)}{\boldsymbol{\pi}\left(\boldsymbol{p}, w\right) \cdot \boldsymbol{x}_w^b \left(\boldsymbol{p}, w\right)}}_{\text{price misperception}}. \tag{2}$$

The first term captures utility misperception—the wedge between perceived and actual marginal utilities, each normalized by the money-metric value. This component is positive when the agent overweights a good's utility relative to its true contribution to welfare. The second term captures price misperception—the wedge between true prices and the agent's perceived prices. A positive value indicates that the agent perceives a good to be cheaper than it truly is, which leads to overconsumption at the margin.

The wedge integrates multiple forms of behavioral bias into a unified measure. This structure allows the behavioral wedge to serve as a sufficient statistic for welfare-relevant distortions.

In the next subsection, the behavioral wedge is used to derive a generalized version of Roy's identity.

2.4 Behavioral Roy's Identity

This section revisits Roy's identity to assess the welfare impacts of price changes in the context of behavioral biases.

Traditional Roy's identity provides a fundamental link between observed demand and welfare in response to price changes. For a rational agent with demand $\boldsymbol{x}^r$, the welfare impact of a small price change Δp_j is equivalent to an income loss of $x_j^r \Delta p_j$. The rational agent experiences no welfare change from substitution effects, as dictated by the envelope theorem, leading to a welfare change due to a change in price p_j as

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

This logic breaks down for a behavioral agent: violations of rationality invalidate the envelope theorem, allowing price changes to affect welfare not only through income but also via suboptimal substitution driven by misperceptions of prices or marginal utilities.

In particular, the behavioral agent misperceives price changes as $\Delta \boldsymbol{\pi}$ and responding based on incorrect marginal rates of substitution. As a result, the response deviates from the optimal path, introducing a substitution effect on welfare. The envelope theorem no longer holds, and the welfare effect of a change in p_j reflects both substitution and income effects.

That is, for the behavioral agent, the welfare change due to a small change in price p_j is

$$\frac{\partial v^b\left(\boldsymbol{p}, \boldsymbol{\pi}\left(\boldsymbol{p}, w\right), w\right)}{\partial p_j} = \left(-x_j^b - \boldsymbol{\theta} \cdot \boldsymbol{S}_j^b\right) v_w^b. \tag{3}$$

where $\boldsymbol{S}_j^b$ is the j-th column of the behavioral income-compensated Slutsky matrix $\boldsymbol{S}^b$, corresponding to the consumption response to a compensated change in the price p_j defined as

$$\boldsymbol{S}_j^b = \boldsymbol{x}_{p_j}^b\left(\boldsymbol{p}, w\right) + x_j^b\left(\boldsymbol{p}, w\right) \boldsymbol{x}_w^b\left(\boldsymbol{p}, w\right).$$

The term $\boldsymbol{\theta} \cdot \boldsymbol{S}_j^b v_w^b$ in equation (3) is a new term that is specific to the behavioral agent. It captures how behavioral biases influence welfare change from a price change. The first term is a standard income effect: an increase in the price of good j appears as if income changes by $-x_j^b$, which is converted to welfare change by the money metric v_w^b as $-x_j^b v_w^b$. The second term is the substitution effect: a change in the price of good j alters consumption according to $\boldsymbol{S}_j^b$. With non-zero behavioral wedge $\boldsymbol{\theta}$, the substitution effect affects welfare as $\boldsymbol{\theta} \cdot \boldsymbol{S}_j^b v_w$.

3 A Classical Trade Model

This section describes a classical trade model and then analyzes how behavioral biases affect gains from trade and shape the pattern of trade.

3.1 Setup

This section explores a classical international trade model involving a country, called "Home," transitioning from autarky to global trade.

The economy consists of n goods with the $n \times 1$ price vector $\boldsymbol{p}$, and a single production factor, labor. Consumption and output vectors are denoted as $\boldsymbol{x} \equiv \{x_i\}$ and $\boldsymbol{y} \equiv \{y_i\}$, respectively, where x_i and y_i are the consumption and output of good i. With a fixed labor supply L, the production functions adhere to standard economic assumptions: twice differentiability, diminishing marginal returns, and constant returns to scale. Assume that the production possibility frontier is represented by $F\left(\boldsymbol{y}; L\right) = 0$. It is assumed that equilibrium always exists and is unique.

Home's income, represented by $R\left(\boldsymbol{p}\right) = \boldsymbol{p} \cdot \boldsymbol{y}\left(\boldsymbol{p}\right)$ is derived from the production under a given price vector $\boldsymbol{p}$. Home's welfare, $v\left(\boldsymbol{p}, R\left(\boldsymbol{p}\right)\right)$, is evaluated under price vector $\boldsymbol{p}$ and income $R\left(\boldsymbol{p}\right)$, with $v_w \equiv \partial v / \partial R$ denoting the marginal utility of income.

The autarky and world price vectors are denoted by $\boldsymbol{p}^A$ and $\boldsymbol{p}^W$. In autarky, the market clearing conditions determine the autarky prices $\boldsymbol{p}^A$ such that that

$$\boldsymbol{p}^A \cdot \left(\boldsymbol{y}\left(\boldsymbol{p}^A\right) - \boldsymbol{x}\left(\boldsymbol{p}^A\right)\right) \equiv \mathbf{0}. \tag{4}$$

The condition in equation (4) allows the equilibrium to have excess supply if the price of that good is zero.

Gains from trade are defined as the welfare difference between these two prices:

$$G(\boldsymbol{p}_1, \boldsymbol{p}_2) = v(\boldsymbol{p}_2, R(\boldsymbol{p}_2)) - v(\boldsymbol{p}_1, R(\boldsymbol{p}_1)).$$

3.2 Pattern of Trade

This section explores the implications of behavioral biases for the pattern of trade by comparing two otherwise identical countries: one populated by a rational agent, and the other by a behavioral agent. Although the countries share the same fundamentals—technologies, endowments, and true preferences—their trade patterns could diverge due to differences in decision-making.

Since the rational agent and the behavioral agent choose different consumption bundles, the equilibrium conditions under autarky differ. As a result, the autarky prices for the two agents also diverge.

The autarky prices for the rational agent and the behavioral agent, denoted by $\boldsymbol{p}^{A,r}$ and $\boldsymbol{p}^{A,b}$, respectively, satisfy the conditions

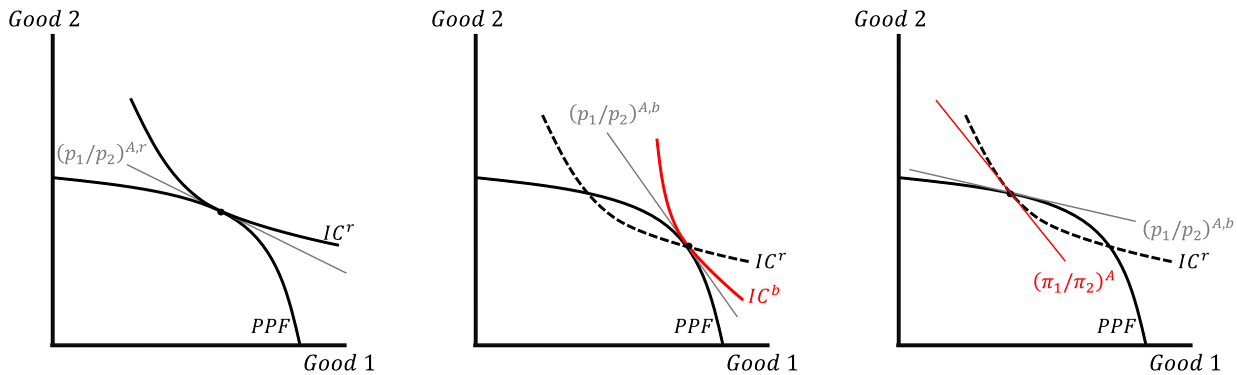
$$\boldsymbol{p}^{A,r} \cdot \left[\boldsymbol{y}\left(\boldsymbol{p}^{A,r}\right) - \boldsymbol{x}^r\left(\boldsymbol{p}^{A,r}\right)\right] \equiv \boldsymbol{0}$$
$$\boldsymbol{p}^{A,b} \cdot \left[\boldsymbol{y}\left(\boldsymbol{p}^{A,b}\right) - \boldsymbol{x}^b\left(\boldsymbol{p}^{A,b}, \boldsymbol{\pi}\left(\boldsymbol{p}^{A,b}, R\left(\boldsymbol{p}^{A,b}\right)\right), R\left(\boldsymbol{p}^{A,b}\right)\right)\right] \equiv \boldsymbol{0}.$$

These equilibrium conditions highlight how behavioral biases can alter autarky outcomes. In autarky, behavioral biases distort consumption choices, which in turn induce a movement along the production possibilities frontier (PPF). As a result, production must adjust to meet the distorted demand, leading to different production bundles: $\boldsymbol{y}\left(\boldsymbol{p}^{A,r}\right)$ for the rational economy and $\boldsymbol{y}\left(\boldsymbol{p}^{A,b}\right)$ for the behavioral one. Since autarky prices must clear markets domestically, the change in production and consumption implies that autarky prices $\boldsymbol{p}^{A,r}$ and $\boldsymbol{p}^{A,b}$ must also differ. These differences, in turn, reshape the pattern of comparative advantage and trade flows, even though fundamentals such as technology and endowments are identical across countries.

Figure 1 illustrates possibilities of different patterns of trade. Consider a two-good economy to visualize the impact of these biases on trade patterns. Home's production technology is described by the production possibility frontier (PPF).

A rational agent (Figure 1a) Home's welfare is maximized, given its production constraint, by choosing the production point and consumption point where the indifference curve IC^r is tangent to the PPF. The autarky price is $(p_1/p_2)^{A,r}$.

A behavioral agent with utility misperception (Figure 1b) The behavioral agent misperceives her preferences and aims to maximize her utility using the behavioral indifference curve IC^b, which results in the utility level according to the true indifference curve IC^r. This preference bias for good 1 adjusts the autarky price to $(p_1/p_2)^{A,b}$, moving production away from good 2 towards



(a) a rational agent

(b) a behavioral agent with utility misperception

(c) a behavioral agent with price misperception

Figure 1: Autarkic equilibria under different scenarios

good 1. As a result, the welfare of the behavioral agent in Figure 1b is smaller than the welfare of the rational agent in Figure 1a.

A behavioral agent with price misperception (Figure 1c) The behavioral agent misperceives the price ratio as $(\pi_1/\pi_2)^A$, instead of the true price ratio $(p_1/p_2)^{A,b}$. This misperception leads her to underconsume good 1 and overconsume good 2, based on her perceived prices. Despite production aligning with true prices, consumption based on misperceived prices results in a suboptimal allocation in autarky, where the behavioral agent's true utility, indicated by the true indifference curve IC^r, is smaller than the utility level the rational agent achieves.

In short, these three examples illustrate how behavioral biases in utility and price misperception can distort trade patterns and welfare outcomes in autarky.

3.3 Gains from trade

This section evaluates the welfare gains from trade under two scenarios: (i) a change in world prices and (ii) a deviation from the autarky.

3.3.1 A Change in World Prices

This subsection compares welfare gains from trade of the rational and behavioral agents when world prices change due to trade liberalization.

For the rational agent, a change in price p_j^W leads to welfare gains from trade approximated by

$$\frac{\partial v\left(\boldsymbol{p}^W, \boldsymbol{p}^W \cdot \boldsymbol{y}\right)}{\partial p_j^W} = v_w\left(y_j - x_j\right).$$

The derivation is directly from Roy's identity and Hotelling's lemma. This equation states that

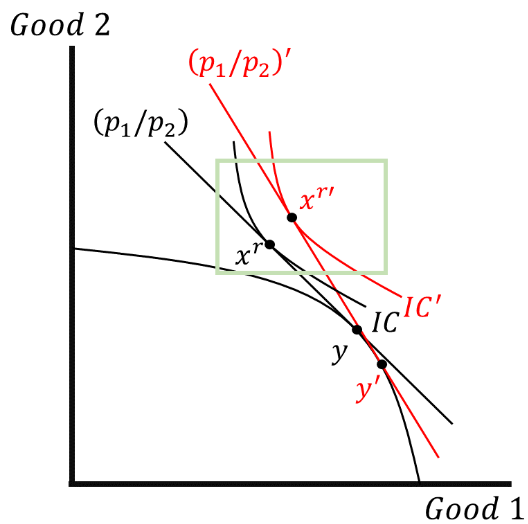


Figure 2: A change in market equilibrium under the rational agent when the price ratio increases from (p_1/p_2) to $(p_1/p_2)'$.

the welfare gain from a change in price p_j^W depends on $(y_j - x_j) \triangle p_j^W$, where the term $(y_j - x_j)$ is the net export of good j and $\triangle p_j^W$ is the magnitude of the price change. Home experiences welfare gains if (i) Home is an exporter of good j and price p_j increases, or (ii) Home is an importer of good j and price p_j^W decreases. In contrast, Home receives welfare losses if (i) Home is an exporter of good j and price p_j decreases, or (ii) Home is an importer of good j and price p_j^W increases. The size of the welfare change in terms of money depends on the magnitudes of net export/import and the price change. This welfare change is converted from money value to utility units using a money metric denoted by v_w.

In general, the welfare gains that arise from a change in world prices $\boldsymbol{Dp}^W \equiv \{dp_i^W\}$ can be calculated by

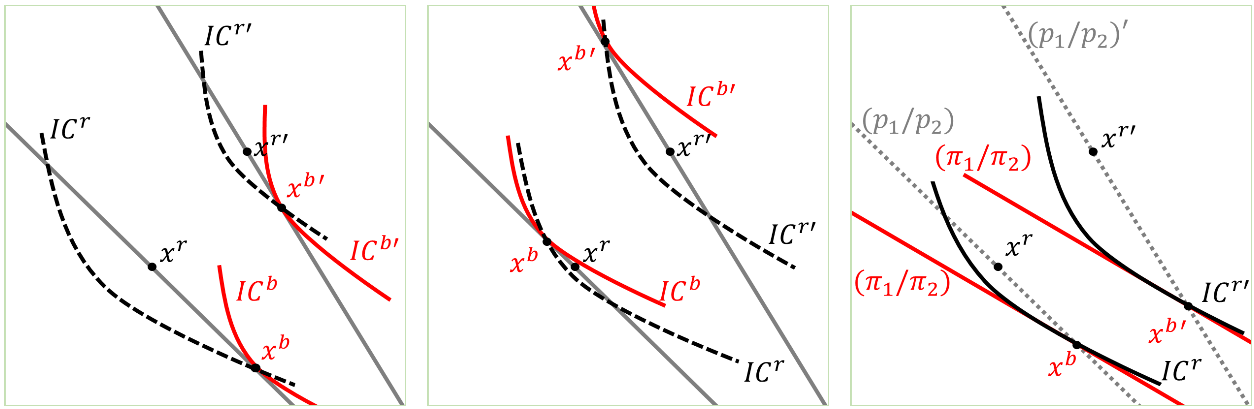
$$dv\left(\boldsymbol{p}^W, \boldsymbol{p}^W \cdot \boldsymbol{y}\right) = v_w\left(\boldsymbol{y} - \boldsymbol{x}\right) \cdot \boldsymbol{Dp}^W \tag{5}$$

Figure 2 illustrates gains from trade under the rational agent. When the price ratio increases from (p_1/p_2) to $(p_1/p_2)'$, production moves along the PPF from point y to point y', and consumption moves from point x^r on the black budget line to point $x^{r'}$ on the (new) red budget line. Welfare gains from the price change are captured by the movement from the indifference curve IC to the indifference curve IC'.

For the behavioral agent, the welfare change from a change in price p_j^W is

$$\frac{\partial v\left(\boldsymbol{p}^W, \boldsymbol{p}^W \cdot \boldsymbol{y}\right)}{\partial p_j^W} = v_w\left(y_j - x_j - \boldsymbol{\theta}^T \boldsymbol{S}_j^b\right) = \left(y_j - x_j\right) v_w + \left(-\boldsymbol{\theta}^T \boldsymbol{S}_j^b\right) v_w.$$

This equation is directly from the behavioral version of Roy's identity in equation (3) and



(a) a behavioral agent with utility misperception

(b) a behavioral agent with utility misperception

(c) a behavioral agent with price misperception

Figure 3: A change in market equilibrium when the price ratio increases from (p_1/p_2) to $(p_1/p_2)'$ under behavioral agents.

Hotelling's lemma. The welfare gains from trade under the behavioral agent comprise not only the traditional gains but also the distortions from behavioral biases. This latter component depends on two factors: (i) the extent to which price changes affect consumption, as quantified by the Slutsky matrix $\boldsymbol{S}_j^b$, and (ii) the degree of inefficiency, measured by the behavioral wedge $\boldsymbol{\theta}$. Consequently, the behavioral agent might realize either amplified or diminished welfare gains from trade, contingent on the nature and extent of her behavioral biases.

The welfare gains from trade that arise from a change in world prices $\boldsymbol{Dp}^W \equiv \{dp_i^W\}$ is summarized by Proposition 1.

Proposition 1. *Welfare gains that arise from a change in world prices $\boldsymbol{Dp}^W$ can be calculated by*

$$dv\left(\boldsymbol{p}^W, \boldsymbol{p}^W \cdot \boldsymbol{y}\right) = \underbrace{v_w\left(\boldsymbol{y} - \boldsymbol{x}\right) \cdot \boldsymbol{Dp}^W}_{\text{the traditional gains from trade}} + \underbrace{v_w\left(-\boldsymbol{\theta}^T \boldsymbol{S}^b\right) \cdot \boldsymbol{Dp}^W}_{\text{distortions from behavioral biases}} . \quad (6)$$

Figure 3 provides three examples of how behavioral biases impact welfare gains from trade. These scenarios are based on the equilibrium in Figure 2. Because behavioral biases do not affect production under the world prices, in all three examples in Figure 3, the production points y and y', the budget lines, and the consumption points x^r and $x^{r'}$ are consistent with the equilibrium under the rational agent in Figure 2. For visual clarity, Figure 3 concentrates exclusively on consumption decisions.

Figure 3a portrays a scenario in which a behavioral agent with utility misperception may get amplified gains from trade, Figure 3b illustrates a situation where utility misperception could lead to diminished trade gains, and Figure 3c shows a case that a behavioral agent with price misperception may get diminished gains.

Utility misperception and amplified gains from trade (Figure 3a) The behavioral agent misperceives her preferences and maximizes her utility based on her misperceived utility function, leading her to choose the consumption point x^b where her perceived indifference curve IC^b is tangent to her budget line. However, her true utility level aligns with the indifference curve IC^r. After the price change, she chooses her new consumption point $x^{b'}$ where her perceived indifference curve $IC^{b'}$ is tangent to the new budget line and yields her true utility level based on the indifference curve $IC^{r'}$. In this case, the magnitude of the distortions from behavioral bias diminishes after the price change, as evidenced by $\left\|x^{r'} - x^{b'}\right\| < \|x^r - x^b\|$. Consequently, the behavioral agent's welfare gains from trade surpass those of the rational agent.

For instance, consider the gains from trade when imported avocados become cheaper. The rational agent correctly controls their consumption in response to health considerations and appropriately increases his avocado consumption, reaping welfare gains from trade. The behavioral agent, initially prone to underconsuming healthy foods, might benefit more significantly from the price change by increasing their healthy food consumption, thus reducing the distortions stemming from the original behavioral bias.

Utility misperception and reduced gains from trade (Figure 3b) Contrary to scenario 3a, The behavioral agent misperceives her preferences, but her new consumption point $x^{b'}$ on the perceived indifference curve $IC^{b'}$ amplifies the behavioral bias distortions after the price change, as suggested by $\left\|x^{r'} - x^{b'}\right\| > \|x^r - x^b\|$. This results in the behavioral agent's welfare gains from trade being smaller than those of the rational agent.

To contextualize, in the avocado example, while the rational agent adapts their consumption towards cheaper avocados, the behavioral agent who has habitual consumption fails to adjust optimally, potentially incurring smaller gains or even welfare losses from trade due to their consumption habits.

Price misperception and distorted gains from trade (Figure 3c) This example focuses on a behavioral agent with price misperception; she does not recognize the true price ratio $(p_1/p_2)'$ and uses the perceived price ratio (π_1/π_2). Her consumption choice aligns with the point on the budget line where her true marginal rate of substitution equals the perceived price ratio (π_1/π_2), leading to the consumption point on the indifference curve IC^r. Despite the price change, her inattentiveness to the updated price ratio results in her choosing the new consumption point $x^{b'}$ on the indifference curve $IC^{r'}$, where her marginal rate of substitution remains aligned with the perceived price ratio (π_1/π_2).

Revisiting the avocado example, if the behavioral agent fails to account for the new, lower avocado prices, her consumption remains suboptimal. Her actions, governed by outdated price perceptions, result in welfare gains that are less than what could be achieved with accurate price information.

3.3.2 A Deviation from the Autarky

This subsection contrasts welfare gains from trade of rational and behavioral agents when the initial prices are the autarky prices.

For the rational agent, evaluating welfare change from a change in world prices at the autarky prices yields

$$dv\left(\boldsymbol{p}^W, \boldsymbol{p}^W \cdot \boldsymbol{y}\right)\Big|_{\boldsymbol{p}^W=\boldsymbol{p}^A} = 0. \tag{7}$$

This result is directly from equation (5) and the autarky market equilibrium condition that the domestic production of good i is equal to the consumption of good i. The fact that the derivative is zero implies that Home's welfare is minimized at the autarky prices and any deviation from these prices should theoretically result in welfare gains from trade.

In contrast to the rational agent, the behavioral agent's welfare changes from trade are influenced by behavioral biases, as shown in Proposition 2.

Proposition 2. *For the behavioral agent, welfare gains from trade evaluated at the autarky prices are calculated by*

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

Proposition 2 states that the derivative is not necessarily zero, which means that the behavioral agent' welfare does not necessarily reach its global minimum value at the autarky prices due to the presence of behavioral biases.

Given that the welfare function is convex with respect to prices, the fact that welfare is not minimized at the autarky prices suggests the existence of a set of world prices where a country's welfare could be less than that achieved under autarky conditions. In essence, a country may experience trade-induced welfare losses if world prices move into an unfavorable range.

Proposition 3. *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.*

4 Government's Optimal Interventions

4.1 Model Setup

Consider a world economy with 2 countries, called Home (without *) and Foreign (with *). There are 2 goods, denoted by $i = 1, 2$, which are produced under perfect competition. Home's consumption and production of good i are denoted by x_i and y_i, respectively. Good 2 is the numeraire good, and its price is normalized to one. Let p^w and p be the relative price of good 1 in the world and Home, respectively.

Home's income, represented by $R = py_1(p) + y_2(p)$ is derived from the production given the domestic prices. Suppose that Home is a natural importer of good 1 and a natural exporter of good 2. The balance of trade requires that $p^w(x_1 - y_1) = y_2 - x_2$. The market clearing condition is that $x_1 - y_1 = y_1^* - x_1^*$.

Tariffs τ and τ^* create a wedge between the domestic price and world price as $p = (1+\tau)p^w$ and $p^* = p^w/(1+\tau^*)$. Following the standard assumptions, the welfare function is assumed to be twice differentiable in τ and a global maximum exists and is unique. Home's government maximizes the welfare of its representative agent:

$$\max_{\tau} = v\left(p, R + (p - p^w)(x_1 - y_1)\right). \tag{8}$$

4.2 Optimal Tariffs

When Home's government chooses tariff τ to maximize the indirect welfare function in equation (8), the tariff is set according to the following first-order condition

$$\frac{\partial v}{\partial p}\frac{dp}{d\tau} + v_w\left[\frac{dR}{dp}\frac{dp}{d\tau} + \left(\frac{dp}{d\tau} - \frac{dp^w}{d\tau}\right)(x_1 - y_1) + \tau p^w \frac{d(x_1 - y_1)}{d\tau}\right] = 0. \tag{9}$$

The term $\frac{\partial v(p,w)}{\partial p}$ can be simplified using Roy's identity, while $\frac{dR}{dp} = y_1$ is directly from Hotelling's lemma. For the rational agent, the optimal tariff in equation (9) is reduced to

$$\underbrace{\tau^* p^w \frac{dm_1}{d\tau}}_{\text{distortions}} + \underbrace{\left(-m_1 \frac{dp^w}{d\tau}\right)}_{\text{terms-of-trade manipulation}} = 0, \tag{10}$$

where m_1 denotes Home's net import of good 1.

The first term represents the domestic distortion caused by the impact of tariffs on production and consumption levels. The second term represents the terms-of-trade effect of using tariffs to lower the world prices of imports. Equation (10) states that at the optimal tariff levels, the marginal benefit of tariffs via the terms-of-trade effect should be equal to its marginal cost in terms of domestic distortions.

For the rational agent, the optimal tariff τ^* in equation 10 is expressed explicitly as

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

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

This expression aligns with Johnson's (1953) optimal tariffs. The idea is that the efficacy of tariffs depends on the foreign elasticity of export supply. Countries with significant market power (where foreign supply elasticity is finite and notably less than infinity) can receive net welfare gains from positive tariffs. In contrast, small economies, characterized by infinite foreign supply elasticity, find that their optimal tariffs are zero.

For the behavioral agent, the optimal tariff balances not only between domestic distortions and the terms-of-trade effect but also accounts for how tariffs influence consumption behaviors distorted by behavioral biases.

Substituting the behavioral version of Roy's identity in equation (3) into the condition for the optimal tariff in equation (9) leads to the equation in Proposition 4.

Proposition 4. *In the case of the behavioral agent, Home's optimal tariff* τ^* *satisfies*

$$\underbrace{\left(-\boldsymbol{\theta}\cdot \boldsymbol{S}_1^b\right)\frac{dp}{d\tau}}_{\text{behavioral biases}}+\underbrace{\tau^* p^w \frac{dm_1}{d\tau}}_{\text{distortions}}+\underbrace{\left(-m_1\frac{dp^w}{d\tau}\right)}_{\text{terms-of-trade manipulation}}=0. \tag{11}$$

This condition highlights that, in addition to the tariff revenue and the terms-of-trade manipulation, optimal tariffs must also mitigate or exploit the effects of behavioral biases on consumption patterns.

Analyzing the impact of an increase in tariff τ on a particular good i involves understanding two key factors: (i) the cross-price elasticity, based on the Slutsky matrix $\boldsymbol{S}_1^b$ and (ii) whether good i is overconsumed, underconsumed, or consumed at an optimal level, based on the behavioral wedge $\boldsymbol{\theta}$.

An increase in tariff τ magnifies welfare loss from behavioral biases if (i) the price increase leads to higher consumption of good i when it is already overconsumed ($\theta_i > 0$), or (ii) the price increase reduces the consumption of good i, when it is already underconsumed ($\theta_i < 0$). Conversely, an increase in tariff τ can reduce the welfare loss from behavioral biases if (i) the price increase leads to higher consumption of good i when it is already underconsumed ($\theta_i < 0$), or (ii) the price increase lowers consumption of good i, when it is already overconsumed ($\theta_i > 0$).

The optimal tariffs in equation (11) can be solved explicitly as in Proposition 5.

Proposition 5. *In the case of the behavioral agent, Home's optimal tariff* τ *is*

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

In contrast to the rational agent's optimal tariffs, which are primarily a function of foreign elasticity of export supply, the behavioral agent's optimal tariffs incorporate an additional term that corrects welfare losses attributable to behavioral biases.

This additional term consists of three factors: (i) the effect on consumption, (ii) the magnitude of the behavioral biases, and (iii) the direct welfare cost of tariffs. First, the impact of tariffs on consumption is captured by the Slutsky matrix. In scenarios where consumers are inattentive to tariffs, these tariffs cannot effectively address behavioral biases. Consequently, under such circumstances, tariffs should not be employed as a means for correcting behavioral biases.

Second, the behavioral wedge quantifies the extent of welfare losses due to behavioral biases. If the magnitude of the behavioral wedge is large, optimal tariffs may aim to correct the biases.

Under the rational agent, the behavioral wedge is zero, and the optimal tariffs are reduced to a function of the foreign elasticity of export supply.

Third, the denominator in this term captures the direct welfare cost of tariffs. It is represented by the change in the value of net imports, evaluated at the world price. The intuition is that tariffs should aim to mitigate these welfare losses, particularly when they are substantial in comparison to the welfare costs incurred by the distortive effects of the tariff intervention.

In general, we can interpret this result as that (i) the policymakers use the optimal tariffs to correct the behavioral bias at the expense of the usual welfare, or (ii) the policymakers need to be aware of another welfare consequence via the behavioral biases when it aims to use tariffs to manipulate the terms of trade.

Proposition 6. *The optimal tariff of a small open economy characterized by a behavioral representative consumer is*

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

Proposition 6 establishes a novel insight that in the presence of behavioral biases, trade policy—specifically tariffs—serves as a second-best policy aiming at welfare losses attributable to behavioral biases. Traditional economic theories have offered various explanations for the use of trade policies by small open economies, including the protection-for-sale motive by Grossman and Helpman (1994) and the labor-market motive as discussed by Costinot (2009) and Suwanprasert (2017, 2020, 2024).

4.3 Optimal Nudges

In this section, I focus on nudges, which were introduced by Thaler and Sunstein (2008). Nudges come in many forms, such as reducing food waste in hotels through plate size adjustments (Kallbekken and Salen, 2013), increasing loan demand with attractive advertisements (Bertrand et al., 2010), and raising organ donation rates by making it a default option (Johnson and Goldstein, 2003).

To introduce nudges into the model in a general way, I assume that nudges are different from tariffs in that tariffs influence both the choice set and consumption decisions, but nudges subtly change consumption choices without altering the choice set. This distinction is important for understanding how governments can employ nudges alongside tariffs to optimize their welfare.

Nudges are modeled as a continuous policy variable $\eta \in [0, 1]$ that potentially affects consumption decision vector $\boldsymbol{x}(\eta)$. While nudge policies often manifest in discrete forms—such as images in advertisements or default settings—their essence can be modeled on a continuum, reflecting variations in intensity.

Following the standard assumptions, the welfare function is assumed to be twice differentiable with respect to both τ and η, and a global maximum exists and is unique.

Home's government aims to maximize the utility of its representative agent by selecting tariff τ and nudge η:

$$\max_{\tau,\eta} = v\left(p, R + (p - p^{w})\left(x_1(\eta) - y_1\right)\right) - c(\eta),$$

where $c(\eta)$ represents the cost associated with implementing the nudge policy η, subject to $c(0) = c'(0) = 0$, $c'(\eta) > 0$, and $c''(\eta) > 0$.

The optimal tariffs and nudges are characterized in Proposition 7.

Proposition 7. *At an interior optimum, the optimal nudge and tariff of an economy characterized by a behavioral representative consumer satisfy*

$$\underbrace{-\left(\theta_1 \frac{\partial x_1}{\partial \eta} + \theta_2 \frac{\partial x_2}{\partial \eta}\right)}_{\text{behavioral biases}} + \underbrace{\tau^* p^{w} \frac{\partial m_1}{\partial \eta}}_{\text{distortions}} + \underbrace{\left(-m_1 \frac{dp^{w}}{d\eta}\right)}_{\text{terms-of-trade manipulation}} = c'(\eta), \tag{12}$$

and

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

Equation (12) states that the optimal nudge policy balances its impact on welfare through three main mechanisms. The first mechanism is how the nudge addresses behavioral biases. Nudges can 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 $(\theta_i > 0)$.

The second mechanism captures how the demand for imported goods is distorted and, consequently, influences tariff revenues. The third mechanism is the terms-of-trade manipulation, representing how nudges indirectly affect global demand patterns, thereby affecting world prices. The changes in world prices can benefit or hurt Home, depending on how nudge affects consumption. Proposition 7 states that the marginal benefit of nudge η is equal to the marginal cost of implementing the nudge policy.

The optimal tariff follows the same trade-off in equation (11) Proposition 4.

Proposition 8. *At an interior optimum, the optimal nudge and tariff of an economy characterized by a rational representative consumer satisfy*

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

$$\frac{1}{\varepsilon_1^*} p^{w} \frac{\partial m_1}{\partial \eta} - m_1 \frac{dp^{w}}{d\eta} = c'(\eta).$$

Proposition 8 highlights the role of nudges in a rational agent framework, revealing that even when tariffs are optimally set as the inverse of the elasticity of foreign export supply (ε_1^*), there exists a potential for welfare enhancement through the use of nudges. Nudges emerge as a complementary policy tool, balancing between altering distortions and manipulating terms of trade.

5 Discussions

5.1 Why do people not feel the gains from trade?

The perception of gains from trade is often not uniform among individuals. This discrepancy can be attributed to a variety of factors that influence consumer preferences and choices. I will discuss how several behavioral biases may shape how people perceive gains from trade as smaller than they truly receive.

Utility misperception First, consider the role of consumption habits. Take, for instance, a consumer who has a longstanding preference for a particular coffee brand. Even when international trade introduces a superior coffee brand that offers higher quality at a lower cost, this consumer might still opt for the usual brand. This habit leads to a failure to perceive the benefits of trade.

Second, home bias can significantly affect consumer choices. Consumers may favor domestically produced goods, such as electronics, due to perceived superior quality or a desire to support local businesses. This preference persists even when imported products offer advanced features at a more competitive price. This bias towards domestic products obscures the benefits that trade can offer in terms of variety and cost-effectiveness.

Third, environmental ignorance can play a part. If consumers are not aware of the environmental benefits of certain imported products, such as biodegradable cleaning agents, they might continue to use less eco-friendly domestic alternatives. This lack of awareness prevents them from recognizing the broader benefits of trade, particularly in terms of environmental sustainability.

Fourth, the concept of internalities and overconsumption highlights how personal preferences can overshadow potential benefits from trade. A person might prefer domestically produced fast food due to its taste or convenience, despite the availability of healthier imported organic options. This preference for immediate gratification can lead to overlooking the health benefits that trade can bring.

Lastly, myopia can hinder the perception of long-term gains from trade. A consumer might choose a cheaper domestic car over a more expensive but fuel-efficient imported car, not taking into account the long-term savings on fuel and maintenance. This short-term perspective limits the ability to perceive the economic benefits that trade can offer in the long run.

Price misperception. Misconception in prices encompasses a range of behaviors, including inattention to the true price and reliance on defaults, bounded memory, partial attention to true price changes, inattention to taxes, and the left-digit bias.

First, the phenomenon of inattention to true price and reliance on defaults is a common occurrence. For instance, consider a consumer who regularly purchases a domestic brand of rice at $5 per bag. If an imported brand reduces its price from $6 to $4.50, this consumer, accustomed to the default price, may not notice the change. This lack of awareness results in the consumer missing out on an opportunity to save money despite the price advantage offered by international trade.

Second, bounded memory also plays a crucial role in consumer perception. For example, after a tariff reduction, the price of imported cheese may drop from $10 to $8. However, consumers with limited recall might continue to regard the cheese as 'expensive' based on their memory of past prices. This failure to recognize current savings due to tariff changes limits their ability to benefit from reduced prices.

Third, inattention to true price changes is another factor that can distort consumer perception. When tariffs on imported goods, such as cars, are removed, leading to significant price reductions, consumers might only perceive the price as marginally lower due to partial attention to the actual price change. For example, a price drop from $25,000 to $22,000 may be perceived as a reduction to just $24,000, minimizing the perceived impact of the tariff removal.

Fourth, the issue of inattention to taxes is similarly impactful. If a tariff on an imported good is eliminated, resulting in a lower price, consumers might overlook the reduced tax component of the price. This oversight prevents them from fully appreciating the decrease in price, and thus, the direct benefit of the tariff removal remains unrealized.

Lastly, the left-digit bias significantly influences consumer perception. This bias occurs when consumers focus primarily on the left-most digit of a price. For example, a price drop from $3.99 to $3.59 might still be perceived as around $3 due to the focus on the initial '3'. This bias can lead to misperceptions about the extent of price reductions, potentially affecting consumer decisions.

Conclusion In summary, these biases create a complex psychological landscape that can lead consumers to fail to perceive gains from trade fully.

5.2 The 2018 China–United States trade war and Brexit.

The 2018 China–United States trade war and the United Kingdom's decision to leave the European Union (Brexit) stand as two of the most significant geopolitical events of recent times. Both these events garnered substantial public support, a phenomenon that can be partly explained by behavioral biases, especially in terms of misperceptions in utility and prices. These biases may have influenced the public's ability to assess the benefits of trade accurately.

One of the key behavioral biases at play involves a misperception in gains from trade. In the case of the trade war, many Americans did not fully appreciate how their standard of living benefited from access to cheaper imports from China. These imports kept the cost of living lower, a benefit that was not always visibly connected to the trade policies. Similarly, in the Brexit scenario, many UK residents failed to recognize the price benefits accruing from the seamless import of goods from the EU. This lack of recognition of the benefits of trade can be attributed to a cognitive disconnect between the everyday experiences of individuals and the larger economic processes at play.

Another aspect of behavioral bias relates to the overestimation of the costs of trade liberalization. Behavioral biases often lead to an emphasis on the more visible and immediate consequences of economic policies, such as job losses in certain sectors due to increased competition from trade.

This visibility creates a skewed perception that magnifies these costs, overshadowing the more dispersed and less tangible benefits of trade. This focus on the immediate and visible impacts significantly influenced public opinion, tilting it towards protectionist policies such as the trade war and Brexit. It demonstrates a common psychological tendency to prioritize immediate, concrete costs over abstract, long-term benefits, leading to a preference for policies that might, in reality, result in overall utility losses in the long run.

6 Conclusion

This study has extended the traditional analysis of gains from trade and optimal tariffs by incorporating the complexities of human behavior. The unified framework is able to capture many types of behavioral bias, such as utility misperceptions, bounded rationality, and inattention. I explore how behavioral biases influence gains from trade. Behavioral biases have the potential to either amplify or diminish these gains, with the specific outcome contingent upon the direction of price changes and the nature of the prevailing behavioral biases.

Next, I study the optimal tariffs in the presence of behavioral biases. I show that optimal tariffs may mitigate the welfare losses attributable to behavioral biases. I then consider a joint characterization of optimal tariffs and nudges. Furthermore, this paper discusses implications of understanding why the public may not perceive the gains from trade, particularly in the backdrop of the 2018 China–United States trade war and Brexit. These discussions provide valuable insights into the intersection of international trade and public perception, influenced by underlying behavioral biases.

Overall, this research contributes to the field of international trade, particularly through the lens of behavioral economics. It opens new avenues for research and policy, encouraging a deeper examination of the role of human behavior in the field of international trade. Future studies could further explore the applications of this framework in various trade contexts. This study stands as a pioneering effort in understanding and integrating behavioral economics into the realm of international trade and trade policy formulation.

References

Alfaro, Laura, Maggie Chen, and Davin Chor (2023) "Can Evidence-Based Information Shift Preferences Towards Trade Policy?" Working Paper 31240, National Bureau of Economic Research, https://doi.org/10.3386/w31240.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky (2019) "Regressive Sin Taxes, with an Application to the Optimal Soda Tax*," *The Quarterly Journal of Economics*, 134 (3), 1557–1626, https://doi.org/10.1093/qje/qjz017.

Allcott, Hunt, Sendhil Mullainathan, and Dmitry Taubinsky (2014) "Energy policy with externalities and internalities," *Journal of Public Economics*, 112, 72–88, https://doi.org/10.1016/j.jpubeco.2014.01.004.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein (2015) "Behavioral Hazard in Health Insurance *," *The Quarterly Journal of Economics*, 130 (4), 1623–1667, https://doi.org/10.1093/qje/qjv029.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman (2010) "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment*," *The Quarterly Journal of Economics*, 125 (1), 263–306, https://doi.org/10.1162/qjec.2010.125.1.263.

Chetty, Raj (2009) "The Simple Economics of Salience and Taxation," Working Paper 15246, National Bureau of Economic Research, https://doi.org/10.3386/w15246.

Chetty, Raj, Adam Looney, and Kory Kroft (2009) "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99 (4), 1145–77, https://doi.org/10.1257/aer.99.4.1145.

Costinot, Arnaud (2009) "Jobs, Jobs, Jobs: A "New" Perspective on Protectionism," *Journal of the European Economic Association*, 7 (5), 1011–1041, https://doi.org/10.1162/JEEA.2009.7.5.1011.

Di Tella, Rafael and Dani Rodrik (2020) "Labour Market Shocks and the Demand for Trade Protection: Evidence from Online Surveys," *The Economic Journal*, 130 (628), 1008–1030, https://doi.org/10.1093/ej/ueaa006.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell (2020) "How Well Targeted Are Soda Taxes?" *American Economic Review*, 110 (11), 3661–3704, https://doi.org/10.1257/aer.20171898.

Farhi, Emmanuel and Xavier Gabaix (2020) "Optimal Taxation with Behavioral Agents," *American Economic Review*, 110 (1), 298–336, https://doi.org/10.1257/aer.20151079.

Freund, Caroline and Caglar Ozden (2008) "Trade Policy and Loss Aversion," *American Economic Review*, 98 (4), 1675–1691, https://doi.org/10.1257/aer.98.4.1675.

Grossman, Gene M. and Elhanan Helpman (1994) "Protection for Sale," *The American Economic Review*, 84 (4), 833–850, `http://www.jstor.org/stable/2118033`.

Grossman, Gene M and Elhanan Helpman (2021) "Identity Politics and Trade Policy," *The Review of Economic Studies*, 88 (3), 1101–1126, https://doi.org/10.1093/restud/rdaa031.

Gruber, Jonathan and Botond Koszegi (2001) "Is Addiction "Rational"? Theory and Evidence," *The Quarterly Journal of Economics*, 116 (4), 1261–1303, https://doi.org/10.1162/003355301753265570.

Hiscox, Michael J. (2006) "Through a Glass and Darkly: Attitudes Toward International Trade and the Curious Effects of Issue Framing," *International Organization*, 60 (3), 755–780, 10.1017/S0020818306060255.

Johnson, Eric J. and Daniel Goldstein (2003) "Do Defaults Save Lives?" *Science*, 302 (5649), 1338–1339, https://doi.org/10.1126/science.1091721.

Johnson, Harry G. (1953) "Optimum Tariffs and Retaliation," *The Review of Economic Studies*, 21 (2), 142–153, https://doi.org/10.2307/2296006.

Kallbekken, Steffen and Hakon Salen (2013) "'Nudging' hotel guests to reduce food waste as a win-win environmental measure," *Economics Letters*, 119 (3), 325–327, https://doi.org/10.1016/j.econlet.2013.03.019.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon (2012) "A Reduced-Form Approach to Behavioral Public Finance," *Annual Review of Economics*, 4 (Volume 4, 2012), 511–540, https://doi.org/10.1146/annurev-economics-111809-125033.

O'Donoghue, Ted and Matthew Rabin (2006) "Optimal sin taxes," *Journal of Public Economics*, 90 (10), 1825–1849, https://doi.org/10.1016/j.jpubeco.2006.03.001.

Stantcheva, Stefanie (2022) "Understanding of Trade," Working Paper 30040, National Bureau of Economic Research, https://doi.org/10.3386/w30040.

Suwanprasert, Wisarut (2017) "A note on 'Jobs, Jobs, Jobs: A "New" Perspective on Protectionism' of Costinot (2009)," *Economics Letters*, 157, 163–166, https://doi.org/10.1016/j.econlet.2017.05.036.

——— (2020) "Optimal trade policy, equilibrium unemployment, and labor market inefficiency," *Review of International Economics*, 28 (5), 1232–1268, https://doi.org/10.1111/roie.12492.

——— (2024) "Optimal Trade Policy in a Ricardian Model with Labor-Market Search-and-Matching Frictions," *The B.E. Journal of Theoretical Economics*, 24 (2), 621–659, https://doi.org/10.1515/bejte-2023-0120.

Thaler, Richard H. and Cass R. Sunstein (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness*: Yale University Press, 304, https://doi.org/10.1017/S1474747209990175.

Tovar, Patricia (2009) "The effects of loss aversion on trade policy: Theory and evidence," *Journal of International Economics*, 78 (1), 154–167, https://doi.org/10.1016/j.jinteco.2009.01.012.

Wilson, Andrea (2014) "Bounded Memory and Biases in Information Processing," *Econometrica*, 82 (6), 2257–2294, https://doi.org/10.3982/ECTA12188.