A Theory of Narrow Thinking*

Chen Lian†

December 25, 2019

Abstract

I develop an approach, which I term narrow thinking, to break the decision-maker’s ability to perfectly coordinate her multiple decisions. For a narrow thinker, different decisions are based on different, non-nested, information. I recast this individual decision problem as multiple selves playing an incomplete-information game. The narrow thinker then makes each decision with an imperfect understanding of the others. The friction effectively attenuates the interaction across decisions. It then provides a model of narrow bracketing without directly letting each decision be made in isolation. Depending on the environment, narrow thinking can translate into either over- or under-reaction relative to the frictionless benchmark. The main application is that narrow thinking generates “smooth” mental accounting, without requiring the decision maker to have explicit budgets. The approach also generates predictions about what drives the degree of mental accounting behavior.

Keywords: bounded rationality, narrow bracketing, incomplete information, multiple selves, mental accounting

—I am extremely grateful to Marios Angeletos, Alp Simsek, Muhamet Yildiz and Ricardo Caballero for continuous guidance through the project. I thank Daron Acemoglu, Hassan Afrouzi, Emmanuel Farhi, Drew Fudenberg, Xavier Gabaix, Nicola Gennaioli, Robert Gibbons, Itzhak Gilboa, Jonathon Hazell, Bengt Holmstrom, Jennifer La’O, Yueran Ma, Filip Matejka, Stephen Morris, Harry Pei, Collin Raymond (discussant), Karthik Sastry, Frank Schilbach, Joshua Schwartzstein, Andrei Shleifer, Ludwig Straub, Juuso Toikka, Laura Veldkamp, Olivier Wang, Ivan Werning, four anonymous referees, and seminar participants at ASSA, Barcelona GSE, D-TEA, Harvard, and MIT for insightful comments. I acknowledge the financial support from Alfred P. Sloan Foundation Pre-doctoral Fellowship in Behavioral Macroeconomics, awarded through the NBER.

†MIT; lianchen@mit.edu; 77 Massachusetts Avenue, 02139, Cambridge, MA, USA
1 Introduction

Each decision maker faces multiple economic decisions. She purchases different goods, supplies labor, and chooses portfolios separately. In standard modeling practice, we nevertheless implicitly assume that an decision maker can perfectly coordinate these decisions. Consider a standard textbook consumer problem of demanding multiple goods. The classical demand function is derived imposing that, when the consumer purchases one good, she has perfect knowledge of all her other consumption decisions. The consumer can then fully incorporate other consumption decisions’ impact on this particular decision. It is as if the decision maker determines all her consumption decision together, and perfectly coordinates them. In the language of Read, Loewenstein and Rabin (1999), such a decision maker “broadly brackets” all her decisions. However in practice, as research in psychology and behavioral economics shows (Tversky and Kahneman, 1981; Rabin and Weizsacker, 2009), the decision maker often “narrowly brackets,” and makes each decision in isolation.

In this paper, I build a new theory of narrow bracketing behavior, which I term narrow thinking. The theory is based on a different psychological observation that is seemingly separate from narrow bracketing: the decision maker may not incorporate all relevant information when making each decision (Kahneman, 2011). In the rest of the paper, I will show how this observation can break the decision-maker’s ability to perfectly coordinate her multiple decisions, provide a “smooth” model of narrow bracketing, and explain related behavioral phenomena such as mental accounting.

Notion of narrow thinking. The notion of narrow thinking I use throughout the paper is that different decisions are based on different, non-nested, information. This notion can be motivated by the psychological evidence on inattention, bounded recall and selective retrieval from memory (Anderson, 2009; Kahana, 2012; Bordalo, Gennaioli and Shleifer, 2017). As an example of such a narrow thinker, consider the following consumer. When she purchases food, she knows the food price, but does not have the gasoline price at the front of her mind. When she purchases gasoline, she knows the gasoline price, but does not have the food price at the front of her mind. Because her two consumption decisions are based on different, non-nested, information, such a decision maker is a narrow thinker as defined above. As explained shortly, such a narrow thinker faces difficulty in coordinating her decisions.

More abstractly, consider the following general multiple-decision problem. The decision maker’s utility depends on her $N$ decisions $\{x_i\}_{i=1}^{N}$, and the fundamental $\tilde{\theta}$: $u \left( x_1, \ldots, x_N, \tilde{\theta} \right)$. Under narrow thinking, the decision maker is subject to a decision-specific information constraint: each decision $x_i$ needs to be a function of the decision-specific (potentially multi-dimensional) signal $\omega_i$, which captures the state of mind when the decision maker decides on $x_i$. The decision-maker can then be thought of as a team of multiple selves (Marschak and Radner, 1972). Each self is in
charge of one decision but different selves do not perfectly share their information.

I recast the decision problem under narrow thinking as multiple selves playing an incomplete-information, common interest, game. In the equilibrium of the game, as each self does not perfectly know other selves’ signals (states of mind), each self’s decision is made with an imperfect understanding of other selves decisions. In this sense, narrow thinking introduces intra-personal frictions in coordinating multiple decisions. This equivalence also helps me illustrate the methodological contribution: one can use the economic theoretic tools developed for inter-personal coordination frictions to study intra-personal decisions, and provide a new model of bounded rationality.

**A simple consumer theory example: narrow thinking implies narrow bracketing.**

I start with a simple consumer theory example to illustrate how narrow thinking effectively attenuates the interaction across decisions and provides a smooth model of narrow bracketing. In this example, the interaction across the two consumption decisions comes from the complementarity/substitutability embedded in the utility function (second-order cross-derivatives of the utility function), and there are no income effects. Each self $i \in \{1, 2\}$ of the narrow thinker perfectly knows the price of the good she buys $p_i$, but only receives a noisy signal about the other price.

The main question of interest is how the narrow thinker’s consumption responds to price changes. Consider a shock to the price $p_i$. The response of consumption $x_i$ can be decomposed into two parts. The first part captures the direct effect of $p_i$, holding the other consumption fixed. As self $i$ perfectly knows $p_i$, the size of such a direct effect is the same as that of the standard consumer theory. The second part captures the indirect effect, that is, the impact of the other consumption $x_{-i}$ through the complementarity/substitutability between the two goods. Under narrow thinking, the coordinated response of $x_{-i}$ is limited, and such an indirect effect will be dampened.

In the limit case in which each self’s signal about the other price is infinitely noisy, the indirect effect driven by the coordinated response of $x_{-i}$ is completely muted. The narrow thinker’s demand for $x_i$ is the same as that of a decision maker who directly neglects the influence from the other decision. In this sense, narrow thinking implies narrow bracketing. Away from the limit, the indirect effect driven by the coordinated response of $x_{-i}$ is dampened but not completely muted. The narrow thinker’s demand elasticity can then be written as a weighted average between that of standard consumer theory and that when the decision maker directly neglects the other decision. In this sense, narrow thinking provides a smooth model of narrow bracketing. This also highlights the difference from the existing models of narrow bracketing (Barberis, Huang and Thaler, 2006, Rabin and Weizsacker, 2009), which requires the decision maker to make each decision in isolation.

**Main application: a smooth model of mental accounting.** I then turn to the case of the consumer problem with income effects, in which the interaction across decisions comes from the budget constraint. In this environment, narrow thinking provides a smooth model of mental
accounting. A typical explanation of mental accounting is based on explicit budgets (Heath and Soll, 1996), i.e. the decision maker allocates a fixed budget to each good or spending category. In my smooth model, a narrow thinker does not need to have explicit budgets. The difficulty in coordinating her decisions nevertheless moves her demand elasticity closer to that of the explicit mental budgeting model, where there is no need to coordinate different decisions. By providing a smooth model of mental accounting, my approach also leads to new predictions about what drives the degree of mental accounting behavior.

Such a smooth model of mental accounting can then explain mental accounting-type behavior documented in the literature, such as excess sensitivity to own-price shocks (Hastings and Shapiro, 2013), excess smoothness to taste shocks (Heath and Soll, 1996), the label effect (Abeler and Marklein, 2016), and the small wage elasticity of labor supply (Camerer et al., 1997). Let me use excess sensitivity to own-price shocks as an example to explain the intuition. Consider an increase of the food price. Under standard consumer theory, the decision maker can coordinate all her decisions by decreasing other consumption to smooth out the drop in food consumption. Under narrow thinking, however, the coordinated decrease of other consumption is limited, and food consumption will decrease more.

**Under-reaction and over-reaction.** Depending on the environment, narrow thinking can translate into either under- or over-reaction relative to the frictionless benchmark. I introduce a general principle which helps predict whether the narrow thinker over- or under-reacts in a given environment. Similar to the consumer theory example, I first decompose each self’s optimal decision into two parts: the direct effect, driven by the movement of the fundamental while holding other decisions fixed; and the indirect effect, driven by the coordinated response of other decisions. When the indirect effect works in the same direction as the direct effect, a dampening of the indirect effect under narrow thinking leads to under-reaction. When the indirect effect works in the opposite direction to the direct effect, a dampening of the indirect effect under narrow thinking leads to over-reaction.

**Additional applications and extensions.** I also study how narrow thinking can help explain two other types of narrow bracketing behavior: the neglect of “adding-up” effects (Read, Loewenstein and Rabin, 1999) and myopic loss aversion (Barberis, Huang and Thaler, 2006). Finally, I provide a framework to endogeneize narrow thinking: in this problem, besides making multiple decisions, the decision maker also chooses what information each decision is based upon, subject to a cognitive constraint. As different decisions are based on different decision rules, each self is “interested in” different parts of the fundamental. For example, in the simple consumer theory example above, each self wants to know more about the price of the good she buys. In this sense, narrow thinking arises endogenously.
Related literature. The narrow thinking approach builds upon the rational inattention literature (e.g. Sims, 2003; Mackowiak and Wiederholt, 2009; Matejka and McKay, 2015; Mackowiak, Matejka and Wiederholt, 2018) by using noisy signals to capture that the decision maker may not incorporate all relevant information when making each decision. But there is a key difference. For static multiple-decision problems such as the demand of multiple goods (e.g. Koszegi and Matejka, 2019), the rational inattention approach lets different decision be based on the same, imperfect, information. The key friction is the decision maker’s imperfect perception of the fundamental. Each decision is nevertheless made with perfect knowledge about other decisions. By contrast, the narrow thinker’s different decisions are based on different information. Each decision is then made with an imperfect understanding of other decisions. In fact, a key contribution of my paper is to show limited information, seemingly separate from narrow bracketing, can actually provide a smooth model of narrow bracketing and explain related phenomenon such as mental accounting.

Koszegi and Matejka (2019) provide a rational-inattention based theory of mental accounting. Compared to my model of mental accounting, there are a few key differences. First, they stay within the rational inattention paradigm and let different decisions be based on the same, imperfect, information. Second, they focus on providing conditions about when the decision maker endogenously chooses to have explicit budgets. My approach instead focuses on providing a smooth model of mental accounting and generates new predictions about what influences the degree of mental accounting behavior. Third, my approach shows that mental accounting-type behavior arises because of the decision maker’s difficulty in coordinating her multiple decisions. This permits me to connect mental accounting more broadly with narrow bracketing behavior. On another note, Galperti (2019) provides a model of mental accounting based on self-control concerns.

Gabaix (2014, 2019) develops a “sparsity” method to model the decision maker’s imperfect perception of world. Similar to rational inattention, the sparse agent’s multiple decisions are made based on the same, imprecise, perception of the fundamental. For example, when purchasing food, a sparse agent may perfectly know the food price but have imperfect knowledge about the gasoline price. But this means that, when purchasing gasoline, this agent still perfectly knows the food price and has imperfect knowledge about the gasoline price. This is different from the narrow thinker studied here. In an extension of Gabaix (2014), the author also studies a “schizophrenic agent,” whose different decisions are based on different perception of the fundamental. My paper

\[1\] In another strand of literature, Caplin and Dean (2015), Caplin, Dean and Leahy (2017), and Oliveira et al. (2017) study the axiomatic foundations for rational inattention.

\[2\] When applying the rational inattention approach to dynamic problems (e.g. Steiner, Stewart and Matejka, 2017; Hébert and Woodford, 2019), similar to the standard sequential decision problem, the typical assumption is that the information of the earlier decision is perfectly nested in the information of the later decision. On the other hand, the narrow thinker’s different decisions are based on different, non-nested, information.

\[3\] Gabaix (2014)’s sparsity approach does not use noisy signals and the perception of the fundamental there is imperfect but deterministic.
shows that this friction can lead to the decision maker’s difficulty in coordinating her decisions, provide a “smooth” model of narrow bracketing, and explain related behavioral phenomena such as mental accounting.

Bounded recall provides a psychological justification about why the narrow thinker’s different decisions are made based on different, non-nested, information (e.g. Kahana, 2012, Wilson, 2014). My paper then further studies how this friction leads to the decision maker’s difficulty in coordinating her multiple decisions. The bounded recall models in Gennaioli and Shleifer (2010) and Bordalo, Gennaioli and Shleifer (2017) instead focus on using representativeness heuristics to model what comes to the decision maker’s mind when she makes a given individual decision. Azeredo da Silveira and Woodford (2019) develop a bounded recall model of over-reaction of consumption to news in an intertemporal setting. In their model, because the consumer partially forgets her past income, her expected future income over-reacts to news in current income.

A methodological contribution of the paper is that one can use the economic theoretic tools developed for inter-personal coordination frictions to study intra-personal decisions, and provide a new model of bounded rationality. This methodological goal achieved by the equivalence between the decision problem under narrow thinking and the incomplete information game among multiple selves. A key insight from the existing literature is that incomplete information can attenuate the equilibrium interaction across different agents (Angeletos and Lian, 2016, 2018; Bergemann, Heumann and Morris, 2017). Within the context of single-agent, multiple-decision problems studied in the paper, the friction translates into an effective attention of interaction across decisions and a model of narrow bracketing. My approach is also reminiscent of Angeletos and Pavan (2007): they use team theory to solve the planning problem in a multiple-agent economy with dispersed information. Angeletos and Pavan (2009) and Angeletos and La’O (Forthcoming) then use the method to characterize optimal policy with dispersed information.

By viewing the decision maker as a collection of multiple selves, the paper also connects to the literature on multiple-selves (Piccione and Rubinstein, 1997; Benabou and Tirole, 2002, 2003, 2004; Gottlieb, 2014). In this literature, multiple selves have conflicted interests. The multiple selves of the narrow thinker, on the other hand, have common interests. Despite common interests, as different selves do not share their information, they have difficulty in coordinating their decisions in response to shocks to the fundamental.

**Layout.** Section 2 defines the notion of narrow thinking. Section 3 uses a simple consumer theory example to illustrate how narrow thinking implies narrow bracketing. Section 4 turns to the main application: how narrow thinking provides a smooth model of mental accounting. Section 5 provides more general results about the narrow thinker’s behavior. Section 6 studies additional applications and extensions. The Appendix contains proofs and additional results.
2 Narrow Thinking in a Multiple-Decision Problem

This section first introduces a general, unconstrained, multiple-decision problem and defines the notion of narrow thinking: different decisions are based on different, non-nested, information. I next show the solution to this single-agent problem is formally equivalent to an incomplete information, common interest, game among multiple selves: each self is in charge of one decision, but different selves do not perfectly share their information. This representation helps explain why a narrow thinker has difficulty in coordinating her multiple decisions. I finally discuss how to transform a constrained problem into the unconstrained problem introduced here.

Utility. The decision maker’s utility depends on $N$ decisions $\vec{x} = (x_1, \ldots, x_N) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ and the fundamental $\vec{\theta} = (\theta_1, \ldots, \theta_M) \in \Theta$:

$$u(\vec{x}, \vec{\theta}),$$

where $u : \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \times \Theta \to \mathbb{R}$ is a twice continuously differentiable and strictly concave over $\vec{x}$. For each $i$, $\mathcal{X}_i$, a convex set on $\mathbb{R}$, denotes the set of possible decision $x_i$. $\Theta \subseteq \mathbb{R}^M$ denotes the set of possible fundamental $\vec{\theta}$.

Information. I let $(S, \mathcal{F}, P)$ denote the probability (state) space. The fundamental $\vec{\theta}$ then should be viewed as the realization of a random vector on the probability space. To accommodate narrow thinking, I introduce decision-specific information. For each decision $i \in \{1, \ldots, N\}$, I use $\omega_i \in \Omega_i$ to denote the information, i.e. signal (potentially multi-dimensional), under which decision $i$ is made, where $\Omega_i$ denotes the set of possible signal realizations for decision. One can interpret $\omega_i$ as the state of mind when the decision maker decides on $x_i$. Here, each $\omega_i$ is the realization of an exogenously drawn random vector on the probability space. Later, in Section 6, I study a problem in which the decision maker chooses endogenously the information upon which each decision is based.

Decision problem. The decision maker chooses jointly all her decision rules $\{x_i(\cdot) : \Omega_i \to \mathcal{X}_i\}_{i=1}^N$ to maximize her expected utility

$$\max_{\{x_i(\cdot)\}_{i=1}^N} E \left[ u \left( x_1(\omega_1), \ldots, x_N(\omega_N), \vec{\theta} \right) \right].$$

The only restriction embedded in (2) is an information constraint: each decision $i$ needs to be a function of its signal $\omega_i$.

Mathematically, the problem set up in (2) is essentially a “team” problem in the sense of

\footnote{For notation simplicity, in the rest of the paper (except for Section 6), I use the same letter to denote a random variable and its realization.}
Marschak and Radner (1972). In Marschak and Radner (1972), the objective is the common payoff of the team, and the constraint is a team-member-specific information constraint. In the single-agent multiple-decision context studied here, one can think the decision-maker as a team of multiple selves (Piccione and Rubinstein, 1997). The common objective is the utility of the decision maker, and the constraint is a self-specific information constraint. Finally, it is worth noting that, as $u$ is strictly concave over $\vec{x}$, the optimum of (2), if exists, is unique.\footnote{Uniqueness is in the sense that, in any two optima, decision rules are the same almost surely.}\footnote{For generality, I do not restrict the potential set for each $x_i, X_i$, to be compact. As a result, the optimum of (2) may not exists. However, for all applications studied below, the existence of the optimum is guaranteed.}

**Narrow thinking.** Now, I introduce the notion of narrow thinking used throughout the paper: different decisions are made based on different, non-nested, information. Let me use $\mathcal{F}_i$ to denote the $\sigma-$algebra (on the probability space) generated by decision $i$’s signal $\omega_i$.\footnote{Specifically, I use $\mathcal{I}_i$ to denote the partition of the state space $S$ generated by decision $i$’s signal $\omega_i$. Each element of $\mathcal{I}_i$ is then given by $\omega_i^{-1}(y)$, where $y \in \Omega_i$ is a possible signal realization for decision $i$.}

**Definition 1** A decision maker is a narrow thinker if there exists a pair of $(i, j) \in \{1, \cdots, N\}$ such that $\mathcal{F}_i \not\subseteq \mathcal{F}_j$ and $\mathcal{F}_j \not\subseteq \mathcal{F}_i$.

The above condition means that there are at least two decisions $(i, j)$ such that, in the Blackwell’s sense, neither decision $i$’s signal is more informative than decision $j$’s signal nor decision $j$’s signal is more informative than decision $i$’s signal. Equivalently, Definition 1 means that, for the pair $(i, j)$, decision $i$’s corresponding partition is neither coarser nor finer than decision $j$’s corresponding partition.\footnote{Formally, it means, for all $i \neq j$, $\mathcal{F}_i = \mathcal{F}_j$.}

As an example of such a narrow thinker, consider a simple consumer theory example. When the decision maker purchases food, she perfectly knows the food price. However, she does not have the gasoline price at the front of her mind, i.e. she only receives a noisy signal about the gasoline price. When she purchases gasoline, she perfectly knows the gasoline price, but only receives a noisy signal about the food price. Such a decision maker is a narrow thinker, as her two consumption decisions are based on different, non-nested, information. In Section 3, I further study this example and discuss the psychological justifications for narrow thinking, that is, why different decisions are made based on different, non-nested, information.

**Broad thinking.** I then contrast the notion of narrow thinking with the notion of broad thinking. The latter lets the multiple-decisions be made based on the same information.

**Definition 2** A decision maker is a broad thinker if all decisions are made based on the same information.
In the context of multiple-decision problems that are traditionally treated as static, e.g. the standard consumer problem of demanding multiple goods, the notion of broad thinking nests both classical consumer theory and a few standard bounded rationality approaches (e.g. rational inattention and sparsity). When all decisions are based on the same, perfect, knowledge of the fundamental \( \vec{\theta} \), i.e. the price vector, the decision problem in (2) coincides with the standard consumer problem (Mas-Colell, Whinston and Green, 1995). In the case of rational inattention (Koszegi and Matejka, 2019) and sparsity (Gabaix, 2014), the decision maker has imperfect knowledge of the fundamental \( \vec{\theta} \), i.e. the price vector, when making each decision. Nevertheless, different decisions are based on the same, though imperfect, information.

Mapping to the game. The problem in (2) is a single-agent planning problem: the decision maker chooses all decisions jointly, subject to a decision-specific information constraint. To help better understand how narrow thinking captures a decision maker’s difficulty in coordinating her decisions, it is useful to provide an equivalent, game-theoretic, representation of (2). First notice, as the utility \( u(\cdot) \) is strictly concave over \( \vec{x} \), the following decision-by-decision optimality condition is a necessary and sufficient condition for the optimum in (2).

**Lemma 1** \( \{x_1^*, \ldots, x_N^*\} \) solves (2) if and only if

\[
x_i^*(\omega_i) = \arg \max_{x_i} E \left[ u \left( x_i, \vec{x}_{-i}, \vec{\theta} \right) \mid \omega_i \right] \quad \forall i, \omega_i \in \Omega_i.
\] (3)

Condition (3) means that, for each \( i \), the optimal decision \( x_i^*(\omega_i) \) maximizes the decision maker’s expected utility, given the signal realization \( \omega_i \) and the optimal decision rules of other decisions. Lemma 1 then points to the equivalence between the decision problem under narrow thinking and an incomplete information, common interest, game \( G \) among multiple selves. In this game, each player \( i \) corresponds to the self \( i \), who is in charge of decision \( i \). Condition (3) then characterizes the optimal strategy for each self \( i \). To define this game \( G \) formally:

1. The state space \((S, \mathcal{F}, P)\), the fundamental \( \vec{\theta} \), and signals \( \{\omega_i\}_{i=1}^N \) are as defined above.
2. There are \( N \) players. All players share the same payoff function \( u \left( \vec{x}, \vec{\theta} \right) \), where \( \vec{x} = (x_1, \ldots, x_N) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \) and \( x_i \) is player \( i \)’s action.
3. Each player \( i \)’s Harsanyi type is given by her signal \( \omega_i \).

**Proposition 1** The Bayesian Nash Equilibrium in the above defined incomplete information, common interest, game \( G \) among multiple selves coincides with the optimum in (2).

Narrow thinking captures a decision maker’s difficulty in coordinating her decisions. Under narrow thinking, different decisions are made based on different, non-nested, information. In the Bayesian Nash Equilibrium of the equivalent game, each self’s imperfect knowledge
about other selves’ information then translates into her imperfect knowledge about other selves’ decisions. This means that when the decision maker makes a particular decision, she has an imperfect perception of other decisions. In this sense, the narrow thinker faces frictions in coordinating her multiple decisions. As the later analysis shows, such friction effectively attenuates the interaction across decisions and provides a smooth model of narrow bracketing.

Under broad thinking, however, different decisions are made based on the same information. The game among multiple selves becomes a complete information game. Each self’s knowledge about other selves’ information then translates into her perfect knowledge about other selves’ decisions. The decision maker is then able to fully consider the impact of other decisions when making a decision. In this sense, she can perfectly coordinate her multiple decisions. It is as if all decisions are made together.

Transforming a constrained problem into the unconstrained problem. The problem considered above is an unconstrained optimization problem. In applications, one sometimes faces a constrained problem in which the fundamental and decisions need to satisfy

\[ B(\vec{x}, \theta) \leq 0, \]  

where \( B \) is twice continuously differentiable and convex over \( \vec{x} \).

Here I provide a simple and standard approach to transform a constrained problem under narrow thinking to the unconstrained problem in (2): I let the last decision \( x_{N+1} \) adjust automatically given the constraint and other boundedly rational decisions.\(^9\) For example, in the consumer theory setting (Section 3 and 4), one can interpret the last decision as saving or borrowing, and allow it to adjust given the budget constraint and the consumption decisions under narrow thinking. For this to be feasible, for any given \((x_1, \cdots, x_N) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \) and \( \theta \in \Theta \), I assume that there exists a \( x_{N+1} \in \mathcal{X}_{N+1} \) such that the constraint (4) is satisfied. In the consumer theory setting below, this means that I allow the last decision, saving or borrowing, to be negative. With mild conditions,\(^10\) the constraint in (4) always binds in the optimum. One can then use it to substitute \( x_{N+1} \) in the utility, and the problem becomes an unconstrained problem in (2) for the first \( N \) decisions.

\(^9\)For another example of a similar approach, Sims (2003) lets saving adjust automatically based on the budget constraint in the rationally inattentive consumption decision.

\(^10\)For example, the utility \( u(\vec{x}, \theta) \) in (1) and the constraint \( B(\vec{x}, \theta) \) in (4) increase in each decision \( x_i \),
3 A Simple Consumer Theory Example: Narrow Thinking Implies Narrow Bracketing

In this section, I use a simple consumer theory example to illustrate a few key insights: how narrow thinking effectively attenuates the interaction across decisions and provides a smooth model of narrow bracketing.

3.1 Narrow Thinking in a Simple Consumer Theory Example

Set up. The decision maker’s utility depends on her consumption of two goods $x_1 \in \mathbb{R}^+$ and $x_2 \in \mathbb{R}^+$ ("food" and "gasoline"), and is quasi-linear in the last good $y \in \mathbb{R}$ ("saving" or "borrowing"):

$$v(x_1, x_2) + y,$$

where $v$ is strictly concave, twice continuously differentiable, and increasing in $(x_1, x_2)$. As discussed above, one can substitute $y$ based on the budget constraint $p_1 x_1 + p_2 x_2 + y \leq w$, and express the decision maker’s utility as:

$$u(x_1, x_2, p_1, p_2) = v(x_1, x_2) + w - p_1 x_1 - p_2 x_2,$$

(5)

where $p_i$ is good $i$’s price and $w$ is the decision maker’s wealth (treated as a constant, as I am interested in response to price shocks here).

Here, the interaction across the two consumption decisions comes from the complementarity/substitutability embedded in the utility function, i.e. the second-order cross-derivatives of the utility function $\frac{\partial^2 v}{\partial x_1 \partial x_2}$. There are no income effects. In the next section about mental accounting, I will study the case with income effects. There, the interaction across decisions comes from the budget constraint.

Narrow thinking. In the consumer theory example here, it is natural to consider the following narrow thinker: she perfectly knows the price of the good she buys, but only receives a noisy signal about the other price. Specifically, I let prices and signals be log-normally distributed. This facilitates the analytical characterization of the narrow thinker’s behavior and makes sure that prices are always positive. Each self $i \in \{1, 2\}$ of the narrow thinker, who is in charge of purchasing good $i$, perfectly knows $p_i \sim \log \mathcal{N} (\log \bar{p}_i, \sigma^2_{p_i})$, but receives a noisy signal about each of the other $p_{-i}$: $s_{i,-i} = p_{i} \epsilon_{i,-i}$, with $\epsilon_{i,-i} \sim \log \mathcal{N} (0, \sigma^2_{\epsilon_{i,-i}})$ and $\sigma^2_{\epsilon_{i,-i}} > 0$. All $\epsilon$s and $p$s are independent from each other. That is, for $i \in \{1, 2\}$, $\omega_i = \{p_i, s_{i,-i}\}$. Finally, let me introduce $\lambda_{i,-i} \equiv \frac{\sigma^2_{\epsilon_{i,-i}}}{\sigma^2_{p_i} + \sigma^2_{\epsilon_{i,-i}}} \in [0, 1]$ to capture the precision of self $i$’s signal about $p_{-i}$.
Why are different decisions made based on different, non-nested, information? There can be multiple cognitive frictions justifying narrow thinking. First, a fundamental finding of cognitive psychology is that people have bounded recall (Kahana, 2012; Bordalo, Gennaioli and Shleifer, 2017). For example, the recency effect in psychology documents that a person often only has perfect recall of the last few items she encounters. Here, such bounded recall means that when the decision maker purchases food (gasoline), she may not perfectly remember the gasoline (food) price and consumption.

Second, narrow thinking can also arise because of selective retrieval from memory (Anderson, 2009). That is, when the decision maker makes a particular decision, she only evokes a very limited amount of information stored in her memory (Tversky and Kahneman, 1973; Gennaioli and Shleifer, 2010). In the consumer theory context here, a particular relevant observation is the “What You See Is All There Is” principle emphasized by Kahneman (2011) and Enke (2018): when the decision maker purchases food (gasoline), she only sees the food (gasoline) price, and does not have the gasoline (food) price at the front of her mind.

Finally, in Section 6, I also study a problem where the decision maker chooses optimally what information each decision is based upon, subject to a cognitive constraint. In the example here, each self endogenously wants to know more about the price of the good she buys.

Log-linearization. For analytical tractability, I will work with log-linearized optimal decision rules throughout. Such an approximation is standard in the applied literature, provides simple and interpretable formula for commonly used utility functions, and allows me to state main results below in terms of demand elasticities. Specifically, I log-linearize around the point where each price is fixed at $\bar{p}_i$ and each decision is made with perfect knowledge of all prices: $\{\hat{x}_i\}^2_{i=1} = \arg\max_{\{x_i\}^2_{i=1}} u(x_1, x_2, \bar{p}_1, \bar{p}_2)$. I then use a hat over a variable to denote its log-deviation from this point, e.g. $\hat{x}_i = \log \frac{x_i}{\bar{x}_i}$. One can also establish parallel results with linearized optimal decision rules. In this case, main results will be stated in terms of gradients instead of elasticities.

Optimal consumption decisions. The narrow thinker’s optimal consumption decision for each good $i$ is given by the decision-by-decision optimality in (3). Given the environment in (5), I take the first order condition of (3) and log-linearize it. I arrive at the following optimal consumption decision rule for the narrow thinker: for $i \in \{1, 2\}$,

$$\hat{x}_i^* (\omega_i) = -\psi_i \hat{p}_i + \gamma_{i,-i} E_i [\hat{x}_{-i}^* (\omega_{-i})],$$

where $E_i [\cdot] = E [\cdot | \omega_i]$, $\psi_i = -\frac{1}{v_{x_i,x_i}} > 0$, and $\gamma_{i,-i} = -\frac{v_{x_i,x_{-i}}}{v_{x_i,x_i}} \frac{\bar{x}_{-i}}{\bar{x}_i}$. \footnote{Here, $v_{x_i,x_{-i}} = \frac{\partial^2 v(\hat{x}_1, \hat{x}_2)}{\partial x_i \partial x_2}$ and $v_{x_i,x_i} = \frac{\partial^2 v(\hat{x}_1, \hat{x}_2)}{(\partial x_i)^2}$.}
The first term in (6) captures the direct effect of price changes on \( x_i \), that is, the effect of \( \vec{p} = (p_1, p_2) \) holding other decisions fixed. \( \psi_i \) then parametrizes the size of such an effect. As self \( i \) perfectly knows the price of the good she purchases, the size of such a direct effect is the same as the one in standard consumer theory. The second term in (6) captures the indirect effect on \( x_i \), that is, other consumption decisions’ impact on \( x_i \). A positive (negative) \( \gamma_{i,-i} \) means that each pair of goods are complements (substitutes), and that the optimal consumption of good \( i \) increases (decreases) with self \( i \)'s belief about each of the other consumption \( x_{-i} \). Under narrow thinking, as the coordinated response of \( x_{-i} \) is limited, such an indirect effect will be dampened.

**Narrow thinker’s demand.** The main question of interest is how the narrow thinker’s consumption responds to price changes. To facilitate comparison with the demand function in standard consumer theory, I define the narrow thinker’s (log) demand as a function of (log) prices: for \( i \in \{1, 2\} \),

\[
\hat{x}_i^{\text{Narrow}} (\hat{p}_1, \hat{p}_2) \equiv E [\hat{x}_i^* (\omega_i) | \hat{p}_1, \hat{p}_2],
\]

averaging over the realization of noises in signals. It can then be directly compared to \( \hat{x}_i^{\text{Standard}} (\hat{p}_1, \hat{p}_2) \), the (log) demand function in standard consumer theory, in which each consumption decision is made with perfect knowledge of all prices.

**The limit case: narrow thinking implies narrow bracketing.** We first consider a limit case: \( \sigma_{1,2}^2, \sigma_{2,1}^2 = +\infty (\lambda_{1,2}, \lambda_{2,1} = 0) \). That is, each self’s signal about the other price is infinitely noisy.

**Lemma 2** When \( \sigma_{1,2}^2, \sigma_{2,1}^2 = +\infty \), the narrow thinker’s demand is given by: for \( i \in \{1, 2\} \),

\[
\hat{x}_i^{\text{Narrow}} (\hat{p}_1, \hat{p}_2) = -\psi_i \hat{p}_i \equiv \hat{x}_i^{\text{Neglect}} (\hat{p}_1, \hat{p}_2).
\]

To illustrate the intuition behind the Lemma, let us consider the response of \( x_i \) to \( p_i \). In this limit case of narrow thinking, the other self \(-i\) will not perceive and respond to the shock. As a result, for \( x_i \), the indirect effect from the coordinated response of \( x_{-i} \) is completely muted. Self \( i \) then effectively neglects the interaction across decisions: the narrow thinker’s demand for \( x_i \) is the same as that of the decision maker who directly neglects the other decision \( (x_i^{\text{Neglect}}) \). In this sense, narrow thinking implies narrow bracketing.

**The general case: a smooth model of narrow bracketing.** We now turn to the general case that \( \sigma_{1,2}^2, \sigma_{2,1}^2 < +\infty \). In this case, the indirect effect in (6) is dampered but not completely muted. Narrow thinking then provides a smooth model of narrow bracketing.
Proposition 2  The narrow thinker’s own-price demand elasticity can be written as, for \( i \in \{1, 2\} \),

\[
\frac{\partial x_{i}^{\text{Narrow}}}{\partial p_i} = \omega_i \frac{\partial x_{i}^{\text{Neglect}}}{\partial p_i} + (1 - \omega_i) \frac{\partial x_{i}^{\text{Standard}}}{\partial p_i},
\]

where \( \omega_i = \frac{1 - \lambda_{-i,i}}{1 - \lambda_{-i,i} \gamma_{i_{-i},i_{-i}}} \in [0, 1] \) captures the weight on \( x_{i}^{\text{Mechanical}} \). Moreover,

(i) \( \omega_i \) increases when the other self – \( i \)’s signal about \( p_i \) is less precise (lower \( \lambda_{-i,i} \)).

(ii) \( \omega_i \) increases when the interaction across consumption decisions \( \gamma_{i_{-i},i_{-i}} \) is larger.

Proposition 2 shows, in response to shocks to each \( p_i \), the narrow thinker’s demand \( (x_{i}^{\text{Narrow}}) \) can be written as a weighted average between the demand of standard consumer theory \( (x_{i}^{\text{Standard}}) \) and the demand when the decision maker mechanically neglects the other decision \( (x_{i}^{\text{Neglect}}) \). In this sense, narrow thinking bridges the gap between these two cases and provides a smooth model of narrow bracketing. Proposition 2 also highlights the difference from the existing models of narrow bracketing (Barberis, Huang and Thaler, 2006; Rabin and Weizsacker, 2009), which let the decision maker make each decision in isolation (as in \( x_{i}^{\text{Neglect}} \)).

The intuition behind Proposition 2 is similar to that behind Lemma 2: as each self \( i \) knows that the other self – \( i \) will not perfectly perceive and respond to shocks to \( p_i \), the response of \( x_i \) to \( p_i \) is less influenced by the indirect effect from the coordinated response of \( x_{-i} \).

The second part of Proposition 2 further studies what drives the weight \( \omega_i \), which captures the deviation from the standard consumer theory. Naturally, if the other self’s signal about \( p_i \) is more noisy, \( \omega_i \) becomes larger and the demand elasticity of the narrow thinker deviates more from that of the standard consumer theory. Moreover, a larger interaction across decisions means that the effective attenuation of interaction under narrow thinking is more important. This also leads to a larger \( \omega_i \) and moves the demand elasticity of the narrow thinker away from that of the standard consumer theory.

Predictions about demand elasticities. Based on the lessons above, I now directly compare demand elasticities of the narrow thinker with those of standard consumer theory.\(^{12}\)

Corollary 1  (i) The narrow thinker’s cross-price demand elasticities are attenuated: for \( i \in \{1, 2\} \),

\[
\left| \frac{\partial x_{-i}^{\text{Narrow}}}{\partial p_i} \right| \leq \left| \frac{\partial x_{-i}^{\text{Standard}}}{\partial p_i} \right|.
\]

(ii) The narrow thinker’s own-price demand elasticities are attenuated: for \( i \in \{1, 2\} \),

\[
\frac{\partial x_{i}^{\text{Standard}}}{\partial p_i} \leq \frac{\partial x_{i}^{\text{Narrow}}}{\partial p_i} < 0.
\]

\(^{12}\)Corollary 1 can be easily extended to a \( N \)-goods, symmetric, case. See Appendix B for details.
The attenuation of cross-price demand elasticities under narrow thinking mostly comes from “what comes mind,” that is, the fact that each self only receives a noisy signal about the other price. More interestingly, narrow thinking also attenuates the own-price demand elasticities. To understand (10), let us first consider the complements case with $\gamma > 0$. From (6), we know an increase in $p_i$ will have a negative direct effect on $x_i$, captured by $\frac{\partial \hat{x}_N}{\partial p_i} < 0$. As different goods are complements, the consumption of other goods $x_{-i}$ will also decrease. Such a decrease will further decrease $x_i$, generating a negative indirect effect on $x_i$. This means $\frac{\partial \hat{x}_{\text{Standard}}}{\partial p_i} \leq \frac{\partial \hat{x}_N}{\partial p_i}$. Under narrow thinking, the indirect effect from the coordinated response of $x_{-i}$ is dampened. From the weighted average expression in (9), we then know $x_i$ decreases less in response to an increase in $p_i$.

We now turn to the substitutes case with $\gamma < 0$. Similarly, an increase in $p_i$ has a negative direct effect on $x_i$. As different goods are substitutes, the consumption of other goods $x_{-i}$ will now increase. Such an increase will then further decrease $x_i$, again generating a negative indirect effect on $x_i$. That is, we still have $\frac{\partial \hat{x}_{\text{Standard}}}{\partial p_i} \leq \frac{\partial \hat{x}_N}{\partial p_i} < 0$. Under narrow thinking, the indirect effect from the coordinated response of $x_{-i}$ is dampened, and $x_i$ decreases less in response to an increase in $p_i$.

In sum, as the indirect effect of $p_i$ on $x_i$ comes from a second degree interaction, the indirect effect from the coordinated response of $x_{-i}$ is always in the same direction as the direct effect. As a result, the dampening of indirect effects under narrow thinking attenuates own-price demand elasticities.

**Frictional response to shocks and unbiasedness on average.** The above frictional behavior under narrow thinking is about the response to price shocks. As the narrow thinker’s prior about prices coincides with their statistical mean, the narrow thinker’s behavior is unbiased on average.14

**Proposition 3** On average, each narrow thinker’s decision coincides with the frictionless one:

$$E \left[ \hat{x}_{\text{Narrow}} \right] = E \left[ \hat{x}_{\text{Standard}} \right] \quad \forall i,$$

(11)

where $E \left[ \cdot \right]$ averages over the realization of all fundamental and signals.

Based on this prediction, the narrow thinker’s demand elasticity estimated based on temporary

---

13In fact, we have, for $i \in \{1, 2\}$, $\frac{\partial \hat{x}_{\text{Narrow}}}{\partial p_i} = \lambda_{-i} \left[ \omega_i \frac{\partial \hat{x}_{\text{Neglect}}}{\partial p_i} + (1 - \omega_i) \frac{\partial \hat{x}_{\text{Standard}}}{\partial p_i} \right]$. That is, beyond “what comes to mind” (captured by $\lambda_{-i}$), “narrow bracketing” in Proposition 2 also contributes to the attenuation of cross-demand elasticities.

14Proposition 3 is only true within the context of the (log-)-linearized decision rules in (6). However, the observation the narrow thinker’s demand elasticity estimated based on temporary price shocks can then differ from the one estimated based on persistent price differences holds more generally.
price shocks can then differ from the one estimated based on persistent price differences. This is different from the standard consumer theory and can be viewed as a testable prediction of narrow thinking. Appendix B provides further discussion along this line.

**Comparison with rational inattention.** Here, as each self $i$ knows the price of the good she buys, the direct effect of $p_i$ on $x_i$ is the same as of standard consumer theory, and the friction comes solely from the dampening of indirect effects. On the other hand, when applying rational inattention and sparsity to the consumer theory problem here (e.g. Gabaix, 2014 and Koszegi and Matejka, 2019), different decisions are based on the same imperfect knowledge about the prices. The key friction is the frictional direct effects. The decision maker nevertheless perfectly knows her other decisions when making a particular decision. In fact, a form of certainty equivalence emerges for the rationally inattentive decision maker: one can use the standard frictionless decision function, $x_i^{\text{Standard}}(\cdot)$, to characterize her decision. Her consumption for good $i$ is given by $x_i^{\text{Standard}}(E[\bar{p}[\omega]])$, where $\omega$ is her imperfect signal about prices (shared by all selves). Narrow thinking, on the other hand, breaks such certainty equivalence: $x_i^*(\omega_i) \neq x_i^{\text{Standard}}(E[\bar{p}[\omega_i]])$.

**Slutsky asymmetry.** Under narrow thinking, as long as $\lambda_{1,2} \neq \lambda_{2,1}$, that is, when the signal-to-noise ratio of self 1’s signal about $p_2$ differs from that of self 2’s signal about $p_1$, the Slutsky matrix becomes asymmetric.\textsuperscript{15} It is worth noting that rational inattention and sparsity can also lead to Slutsky asymmetry (Gabaix, 2014; Abaluck and Adams, 2017), but for a different reason. There, different decisions are based on the same imperfect knowledge about the prices. Slutsky asymmetry arises if the decision maker’s knowledge about $p_1$ and $p_2$ have different precision.

**The role of sophistication.** As discussed after Proposition 2, each self $i$ here is “sophisticated” as she understands that the other self does not perfectly perceive $p_i$. Such sophistication is naturally embedded in the standard solution concept of the Bayesian Nash Equilibrium in the incomplete information game among multiple selves: each self understands that the other self faces incomplete information. However, such a discipline based on the standard solution concept does not mean that the narrow thinker needs to incur sophisticated game theoretic thinking. She can learn over time based on her past behavior in similar situation that her consumption of a good may not be very sensitive to movements in prices of other goods. In fact, game theorists have developed such learning foundations for the Bayesian Nash Equilibrium (Fudenberg and Levine, 2009). Moreover, as Lemma 2 shows, such sophistication can also lead to simple behavior: the decision maker effectively ignore the impact from the other decision.

\textsuperscript{15} Note that the Slutsky matrix is about demand gradients instead of demand elasticities. To derive demand gradients from demand elasticities (at the point of log-linearization), we have, for all $i, k$, $\frac{\partial x^{\text{Narrow}}}{\partial p_k} = \frac{\partial x^{\text{Narrow}}}{\partial \bar{p}_k} \frac{\bar{p}_k}{p_k}$. 

15
3.2 Interpretation in a Sequential Setting

The above analysis of the narrow thinker’s demand is through the lens of a static, incomplete-information, game among multiple selves. This is aligned with the standard consumer theory, which treats the demand of multiple goods as a static problem. One may naturally wonder how to interpret the above analysis about the narrow thinker’s behavior in an explicit sequential setting.

In fact, in a sequential setting, the narrow thinker’s information structure introduced above can be interpreted as a particular form of bounded recall (or selective retrieval from memory). For example, consider the case that consumption $x_1$ is decided before consumption $x_2$. For the above narrow thinker, when her self 1 decides on $x_1$, she perfectly knows the price $p_1$, but only receives a noisy signal about the price $p_2$. When her self 2 decides on $x_2$, she perfectly knows the price $p_2$, but cannot perfectly recall her past decision $x_1$. In particular, self 2 recalls her past decision $x_1$ only through a noisy signal about $p_1 (s_{2,1})$ and $p_2$. In other words, the narrow thinker’s information structure introduced above imposes a particular form of bounded recall: self 2’s limited memory is captured by her noisy signal about the past price (but not directly from a noisy signal about her past endogenous decision). In Appendix B, I also show that the above narrow thinker is observationally equivalent to a decision maker whose bounded recall is captured by a noisy signal about the past endogenous decision. I focus on the narrow thinker above as the analysis is much more tractable when there are $N \geq 3$ decisions, as Sections 4 - 5 below.

A particular feature of the above analysis is that, as in the standard consumer theory, the narrow thinker’s behavior is insensitive to the exact order of the decisions. No matter whether the decision 1 or the decision 2 comes first, two selves’ information is always given by $\omega_1 = \{p_1, s_{1,2} = p_2e_{1,2}\}$ and $\omega_2 = \{p_2, s_{2,1} = p_1e_{2,1}\}$. To characterize the narrow thinker’s behavior, the analyst does not need the additional information about the order of decisions. This property is particularly when the available data (e.g. the demand data) do not involve the order of decisions.

In Appendix B, I also compare the narrow thinker’s behavior with the behavior of a decision maker who faces standard uncertainty but has perfect recall. Using the language in Section 2, this means the information of the earlier decision is nested by the information of the later decision, i.e. $F_1 \subseteq F_2$. There are three properties of this decision maker’s behavior worth mentioning. First, to characterize such a decision maker’s optimal behavior, the analyst still needs the additional information about the exact order of decisions. Second, given the order of decisions, a form of certainty equivalence emerges for the earlier decision, i.e. $\hat{x}_1^* (\omega_1) = x_1^{\text{Standard}} (\hat{p}_1, E_1 [\hat{p}_2])$. In other words, there is no coordination friction among the two selves, and the only friction comes from the earlier self’s uncertainty about the future fundamental. Third, the total extent of frictional

---

16For self 1 who decides first, the noise in her signal about $p_2$ allows for two interpretations. The noise can come from uncertainty. Alternatively, even if the uncertainty about $p_2$ is resolved before she decides on $x_1$, she could forget about the exact $p_2$ due to bounded recall. Both interpretations will lead to the same analysis.
behavior under perfect recall, driven by uncertainty about fundamentals alone, is limited. The own- and cross-sensitivities under perfect recall deviate less from their frictionless counterparts than those under narrow thinking.

In sum, one can view my approach based on the information structure used throughout the paper as a reduced form method to study the impact of bounded recall and selective retrieval of memory on otherwise standard multiple decision problems. Though it certainly cannot capture all the delicacies of the sequential decision making, the approach is tractable and does not need the additional information about the exact order of decisions.

4 A Smooth Model of Mental Accounting

In this Section, I turn to the main application: how narrow thinking provides a smooth model of mental accounting. For this purpose, I turn to the case of the consumer problem with income effects, where the interaction across different decisions comes from the budget constraint. A typical explanation of mental accounting is based on explicit mental budgets (Heath and Soll, 1996). i.e. the decision maker allocates a fixed budget to each good or spending category. In my smooth model, a narrow thinker does not have an explicit mental budget. The difficulty in coordinating her decisions nevertheless moves her demand elasticity closer to that of the explicit mental budgeting model, where there is no need for different decisions to coordinate. The narrow thinking approach also leads to new predictions about what drives the degree of mental accounting behavior.

4.1 The Main Results

Environment. As the discussion about mental accounting behavior is inherently connected to the budget constraint (Thaler, 1985, 1999), I consider a consumer theory environment with income effects: the interaction across different decisions now come from the budget constraint. To isolate the channel of interest, I also let the consumer have separable utilities. That is, the channel studied in the previous section driven by the second cross-derivates of the utility function is muted here. Specifically, the consumer’s utility is given by

$$\sum_{i=1}^{N} v_i(x_i) + h(y),$$  \hspace{1cm} (12)$$

where $v_i(x_i) = \frac{x_i^{1-\kappa_i}}{1-\kappa_i}$ captures the consumer’s utility from consuming good $i$, and $\kappa_i > 0$ parametrizes the rate at which the marginal utility of consuming good $i$ moves with respect to $x_i$. A higher $\kappa_i$
means the demand for good \( i \) is less elastic. In fact, \(-1/\kappa_i\) can be viewed as the “Frisch” elasticity of demand for good \( i \), that is, the elasticity holding the marginal utility of money fixed.

Moreover, \( h(y) \), a strictly concave function on \( \mathbb{R} \), captures the consumer’s utility from the last decision, which can be interpreted as utility from saving (borrowing) or the value of money. The “residual decision” \( y \) is allowed to be negative and this guarantees that the budget constraint, \( \sum_{i=1}^{N} p_i x_i + y \leq w \), will always be satisfied. As discussed above, one can substitute \( y \) based on the budget and transform the problem into an unconstrained problem, with \( u(x_1, \cdots, x_N, \tilde{p}) = \sum_{i=1}^{N} v_i(x_i) + h \left( w - \sum_{i=1}^{N} p_i x_i \right) \).

Similar to Section 3 above, I consider the following narrow thinker: each self \( i \in \{1, \cdots, N\} \) of the narrow thinker, who is in charge of purchasing good \( i \), perfectly knows \( p_i = \log N \left( \log \tilde{p}_i, \sigma_{p_i}^2 \right) \), but receives a noisy signal about each of the other \( p_k: s_{i,k} = p_k \epsilon_{i,k} \), with \( \epsilon_{i,k} \sim \log N \left( 0, \sigma_{\epsilon_{i,k}}^2 \right) \) and \( \sigma_{\epsilon_{i,k}}^2 > 0 \). All \( \epsilon \)s and \( p \)s are independent from each other. Finally, for all \( i, k \in \{1, \cdots, N\} \), let me also introduce \( \lambda_{i,k} \equiv \frac{\sigma_{\epsilon_{i,k}}^2}{\sigma_{p_i}^2 + \sigma_{\epsilon_{i,k}}^2} \in [0, 1] \) to capture the precision of self \( i \)'s signal about \( p_k \).

**Optimal decisions.** Similar to Section 3, I work with log-linearization. I use a hat over a variable to denote its log-deviation from the point where each price is fixed at \( \tilde{p}_i \) and each decision is made with perfect knowledge of all prices: \( \{\tilde{x}_i\}_{i=1}^{N} = \arg\max_{\{x_i\}_{i=1}^{N}} u(x_1, \cdots, x_N, \tilde{p}_1, \cdots, \tilde{p}_N) \).

The optimal consumption for each self \( i \in \{1, \cdots, N\} \) is given by:

\[
\begin{pmatrix}
\frac{-\kappa_i \hat{x}_i^* (\omega_i)}{\text{marginal utility of consuming good } i} \\
\frac{-\kappa_y \hat{y}^*}{\text{marginal value of money}}
\end{pmatrix} = \hat{p}_i + \hat{E}_i \begin{pmatrix}
\frac{-\kappa_i \hat{x}_i^* (\omega_i)}{\text{marginal utility of consuming good } i} \\
\frac{-\kappa_y \hat{y}^*}{\text{marginal value of money}}
\end{pmatrix},
\tag{13}
\]

where \( -\kappa_i \hat{x}_i^* (\omega_i) \) captures the marginal utility of consuming good \( i \), \( -\kappa_y \hat{y}^* \) captures self \( i \)'s belief about the marginal value of money, and \( \kappa_y = \frac{h''(y) \hat{y}}{h'(y)} \) captures the rate at which the marginal value of money \( h'(y) \) moves with respect to \( y \). Condition (13) means that, from each self \( i \)'s perspective, her expected marginal rate of substitution between the consumption \( x_i \) and the consumption \( y \) should equal \( p_i \). It holds because the last decision \( y \) will adjust based on \( x_i \), and the standard perturbation argument holds between the consumption \( x_i \) and \( y \).

The log-linearized budget constraint is given by

\[
\sum_{i=1}^{N} \mu_i (\hat{x}_i^* (\omega_i) + \hat{p}_i) + \mu_y \hat{y}^* = 0,
\tag{14}
\]

where \( \mu_i = \frac{\tilde{p}_i \hat{p}_i}{w} \) is the spending share of good \( i \) and \( \mu_y = \frac{\tilde{y}}{w} \) is the saving share at the point of log-linearization.

**Violation of the fungibility principle.** From (13), one can see directly that the fungibility

\[\footnote{For notation simplicity, \( \sigma_{p,i}^2 = 0 \) and \( \lambda_{i,i} = 1 \), i.e. each self \( i \) has a perfect signal about \( p_i \).}
principle is violated under narrow thinking. By fungibility principle (Thaler, 1985, 1999), I mean the prediction in standard consumer theory that the marginal value of spending an additional unit of money on each good is the same:

\[ -\kappa_i \hat{x}_i^{\text{Standard}} (\hat{p}_1, \cdots, \hat{p}_N) - \hat{p}_i = -\kappa_j \hat{x}_j^{\text{Standard}} (\hat{p}_1, \cdots, \hat{p}_N) - \hat{p}_j \quad \forall i \neq j. \] (15)

For a narrow thinker, as different selves have different information, they hold different beliefs about the marginal value of money, \(-\kappa_y E_i [\hat{y}^*] \). The marginal value of spending an additional unit of money on each good \(i\), \(-\kappa_i \hat{x}_i^* (\omega_i) - \hat{p}_i = -\kappa_y E_i [\hat{y}^*] \), then also differs.

**Lemma 3** Under narrow thinking, for a pair of decision \((i, j)\), the marginal value of spending an additional unit of money can differ:

\[ -\kappa_i \hat{x}_i^* (\omega_i) - \hat{p}_i \neq -\kappa_j \hat{x}_j^* (\omega_j) - \hat{p}_j. \] (16)

A “smooth” model of mental accounting. I now show how narrow thinking provides a smooth model of mental accounting. A typical explanation of mental accounting is based on explicit mental budgets (Heath and Soll, 1996), i.e. the decision maker allocates an explicit budget \(w_i\) for each good \(i\),

\[ p_i x_i^{\text{Explicit}} = w_i. \] (17)

In this explicit budgeting model, each decision can be made in isolation and there is no need for different decisions to coordinate. Here, I will show, even though the narrow thinker does not have explicit budgets as in (17), her difficulty in coordinating her decisions moves her demand elasticity closer to that of the explicit mental budgeting model. This is similar to the lesson in the previous sections about how the narrow thinker’s difficulty in coordinating her decisions leads to a smooth model of narrow bracketing.

Specifically, substitute \(\hat{y}^*\) based on the budget (14) into (13), we have, for \(i \in \{1, \cdots, N\}\),

\[ \hat{x}_i^* (\omega_i) = -\frac{1 + \mu_i \frac{\kappa_y}{\mu_y} \hat{p}_i}{\kappa_i + \mu_i \frac{\kappa_y}{\mu_y}} \hat{p}_i - \frac{\kappa_y}{\mu_y} E_i \left[ \sum_{j \neq i} \mu_j \left( \hat{x}_j^* (\omega_j) + \hat{p}_j \right) \right]. \] (18)

We can see directly that the response of consumption \(x_i\) to \(p_i\) depends crucially on the indirect effect from the coordinated responses of other \(x_j\). Under narrow thinking, other selves will not perfectly perceive and respond to shocks to \(p_i\). The response of \(x_i\) to \(p_i\) is then less influenced by the indirect effects from the coordinated response of other \(x_j\). This moves the response of \(x_i\) to \(p_i\) closer to the case of explicit budgeting in (17), in which only \(x_i\) responds to \(p_i\).

To formalize this intuition, similar to condition (7), for each \(i\), I define the narrow thinker’s (log)
demand function as \( \hat{x}_{i}^{\text{Narrow}}(\hat{p}_1, \ldots, \hat{p}_N) \equiv E[\hat{x}_{i}^*(\omega_i)|\hat{p}_1, \ldots, \hat{p}_N] \), averaging over the realization of noises in signals. I can then establish:

**Proposition 4** The narrow thinker’s own-price demand elasticity is given by: for \( i \in \{1, \ldots, N\} \),

\[
\frac{\partial \hat{x}_{i}^{\text{Narrow}}}{\partial \hat{p}_i} = \omega_i \frac{\partial \hat{x}_{i}^{\text{Explicit}}}{\partial \hat{p}_i} + (1 - \omega_i) \frac{\partial \hat{x}_{i}^{\text{Standard}}}{\partial \hat{p}_i},
\]

where

\[
\omega_i = 1 - \left( \frac{1}{\sum_{j \neq i} \mu_j \kappa_j \frac{\mu^y_j \lambda_{j,i}}{\kappa_y} + \frac{\mu^y_j}{\kappa_y}} \left( \frac{1}{\sum_{j \neq i} \mu_j \kappa_j + \frac{1}{\kappa_y}} \right) \right) \in [0, 1]
\]

captures the weight on \( \hat{x}_{i}^{\text{Explicit}} \).

Proposition 4 shows, in response to shocks to \( p_i \), the narrow thinker’s demand \( \hat{x}_{i}^{\text{Narrow}} \) is given by a weighted average between demand consumer theory (\( \hat{x}_{i}^{\text{Standard}} \)) and demand with explicit budgeting (\( \hat{x}_{i}^{\text{Explicit}} \)). In this sense, narrow thinking bridges the gap between these two cases and provides a smooth model of mental accounting.

**What drives the degree of mental accounting behavior.** In (20), \( \omega_i \), the weight on \( \hat{x}_{i}^{\text{Explicit}} \), captures the deviation from the standard consumer theory. It can be interpreted as the degree of mental accounting behavior. The next Proposition studies what drives the weight.

**Proposition 5** For \( i \in \{1, \ldots, N\} \), the degree of mental accounting behavior \( \omega_i \in [0, 1] \) has the following properties:

(i) \( \omega_i \) increases when each self \( j \)’s signal about \( p_i \) is less precise (lower \( \lambda_{j,i} \)), for all \( j \neq i \).

(ii) \( \omega_i \) increases when the saving share \( \mu_y \) is smaller or the curvature of \( h(\kappa_y) \), \( \kappa_y \), is larger.

(iii) \( \omega_i \) increases when the expenditure share of good \( i \), \( \mu_i \), is larger or the curvature \( v_i(x_i) \), \( \kappa_i \), is smaller.

(iv) \( \omega_i \to 1 \), when \( \lambda_{j,i} \to 0 \) for all \( j \neq i \) and \( \frac{\mu^y_i}{\kappa_y} \to 0 \). In this limit, the narrow thinker’s behavior proxies that of the explicit budgeting model in (17).

Part (i) of Proposition 5 shows, if other selves’ signals about \( p_i \) are more noisy, the degree of mental accounting behavior \( \omega_i \) becomes larger. Part (ii) of Proposition 5 shows that a smaller saving share (\( \mu_y \)) or a larger curvature \( h(\kappa_y) \) increase the degree of mental accounting behavior \( \omega_i \). Intuitively, a smaller \( \mu_y \) or a larger \( \kappa_y \) means that there is less room for \( y \) to absorb the error made by other selves’ of the narrow thinker. The friction then moves the response of \( x_i \) to \( p_i \) closer to that of the explicit mental budgeting model. Specifically, this result means that a decision maker who does not have much saving will exert more mental accounting-type behavior.
Part (iii) of Proposition 5 shows a larger expenditure share of good $i$ ($\mu_i$) or a smaller curvature $\nu_i(x_i)(\kappa_i)$ increase the degree of mental accounting behavior $\omega_i$. Recall that $\kappa_i$ is inversely related to the “Frisch” elasticity of demand for good $i$. A larger $\mu_i$ or a smaller $\kappa_i$ then means the response of $x_i$ to $p_i$ is more important in determining the marginal value of money. Other selves’ imperfect perception of the response of $x_i$ to $p_i$ then leads more inefficient responses of their decisions. This again moves the response of $x_i$ to $p_i$ closer to that of the explicit mental budgeting model.

Part (iv) of Proposition 5 establishes a limit result about when the narrow thinker’s behavior proxies that of the explicit budgeting model. This is similar to Lemma 2 in Section 3. Specifically, the condition $\{\lambda_{j,i} \to 0\}_{j \neq i}$ means that other selves’ signals about $p_i$ are infinitely noisy and other consumption $\{x_j\}_{j \neq i}$ does not respond to $p_i$. The condition means $\frac{\mu_j}{\kappa_y} \to 0$ means that the adjustment of $y$ is also irrelevant in response to $p_i$. Together, only $x_i$ responds to $p_i$, and the narrow thinker’s behavior proxies that with explicit budgeting, in which also only $x_i$ responds to $p_i$.

Magnitude of the degree of mental accounting behavior. I now provide a first look at the quantitative magnitude of the degree of mental accounting behavior $\omega_i$ in (20). The formula for $\omega_i$ in (20) is rather complicated. The following corollary provides a simpler lower bound for $\omega_i$.

**Corollary 2** Assume common $\kappa$ and $\lambda$, that is, for all $i \neq k \in \{1, \cdots, N\}$, $\kappa_i = \kappa_y = \kappa$ and $\lambda_{i,k} = \lambda$. We have:

$$\omega_i \geq 1 - \frac{\left(\lambda \frac{\mu_x}{\mu_y} \left(1 - \frac{\mu_i}{\mu_x}\right) + 1\right) \left(\frac{\mu_x}{\mu_y} + 1\right)}{\left(1 - \frac{\mu_i}{\mu_x}\right) + \lambda \frac{\mu_x}{\mu_y} + 1} \omega_i,$$

where $\mu_x = \sum_{i=1}^{N} \mu_i$ captures the total expenditure share.

(21) provides a lower bound for the degree of mental accounting $\omega_i$. The lower bound is given by a function of three sufficient statistics: the degree of narrow thinking $\lambda$, the expenditure-saving ratio $\frac{\mu_x}{\mu_y}$, and the share of spending on good $i$ in total expenditure $\frac{\mu_i}{\mu_x}$.

I calibrate the expenditure-saving ratio $\frac{\mu_x}{\mu_y} \approx 5.8$ based on the ratio between the median household’s consumption and net wealth from Baker (2018).\(^{18}\) Consider a good $i$ with spending share $\frac{\mu_i}{\mu_x} = 0.2$. In Figure 1, I plot the lower bound $\omega_i$ as a function of $\lambda$. We can see that the narrow thinker has a sizable degree of mental accounting behavior.

One remaining challenge is to calibrate the degree of narrow thinking, $\lambda$. It is not an easy problem, but let me provide three different directions that might help. First, as discussed above,\(^{18}\) Baker (2018) has high-quality measures based on consumption and net wealth from Linked Financial Account Data. He also re-weights his data to match the distribution of household characteristics in the United States. Based on Table 2 of Baker (2018), the median household’s spending is $57480. The median household’s net wealth (excluding housing) is at most $9990. Together, it implies the expenditure-saving ratio $\frac{\mu_x}{\mu_y} \approx 5.8$.\(^{21}\)
one can interpret $\lambda$ as how much each self $i$ can recall about the price of other goods.\footnote{Note that, for $i \neq k$, $E \left[ \hat{E}_i [\hat{p}_k] | \hat{p}_k \right] = \lambda \hat{p}_k$. As a result, one interpretation of $\lambda$ is how much self $i$ can recall about shocks to $p_k$.} I can then use the psychological evidence on the forgetting curve to calibrate $\lambda$. For example, based on the famous Ebbinghaus forgetting curve (Ebbinghaus, 2013), a decision maker retains around 25\% of the information after a five-days interval. Of course, there is an obvious concern about external validity. But if I take the forgetting curve seriously in the setting here, it implies $\lambda = 25\%$. Second, one can leverage the findings of Hastings and Shapiro (2013) about the violation of the fungibility principle: they find that, from the perspective of gasoline decisions, the marginal utility of money increases 15 times more in response to an increase in the gasoline price than in response to an equivalent decrease in income. In Appendix C, I show how I can extend my analysis to their setting, how to infer $\lambda$ from the above ratio about the differential impact on the marginal utility of wealth, and how their estimates imply a sizable friction. Finally, as mentioned above, one can infer $\lambda$ based on the fact the narrow thinker’s demand elasticity estimated based on temporary price shocks can differ from the one estimated based on persistent price differences.\footnote{Hastings and Shapiro (2013) find that, when gasoline prices rise, consumers substitute to lower grade gasoline and decrease their total gasoline expenditure, to an extent that cannot be explained by neoclassical effects. It is worth noting that their setting is slightly different from mine: they study the impact of a common price shock to three different types of gasoline. In Appendix C, I show how one could re-interpret my analysis to accommodate

**Excess sensitivity to own-price changes.** I now directly compare demand elasticities of the narrow thinker with those of standard consumer theory. This helps show how the smooth model of mental accounting in Proposition 4 can help explain mental accounting-type behavior documented in the literature, such as excess sensitivity to own-price changes (Hastings and Shapiro, 2013).\footnote{Note that, for $i \neq k$, $E \left[ \hat{E}_i [\hat{p}_k] | \hat{p}_k \right] = \lambda \hat{p}_k$. As a result, one interpretation of $\lambda$ is how much self $i$ can recall about shocks to $p_k$.}
Corollary 3 For each good $i$ such that $\kappa_i > 1$, the narrow thinker’s consumption $x_i$ decreases (increases) more in response to positive (negative) shocks to $p_i$:
\[
\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} < \frac{\partial x_i^{\text{Standard}}}{\partial p_i} < 0.
\]

To see the mechanism behind the excess sensitivity, note that, for a good with a relatively inelastic demand ($\kappa_i > 1$), an increase in $p_i$ will decrease the consumption of other goods, $x_j$ (both in standard consumer theory and under narrow thinking). This is because, when $\kappa_i > 1$, the income effect of $p_i$ on $x_j$ (negative) will dominate the substitution effect of $p_i$ on $x_j$ (positive). The indirect effect through the decrease of $x_j$ then positively influences $x_i$. This positive indirect effect then works in the opposite direction to the negative direct effect of $p_i$ on $x_i$. Under narrow thinking, the indirect effect from the coordinated response of $x_j$ is dampened. The narrow thinker then exhibits excess sensitivity to own-price changes.

One can also directly see the result through the lens of Proposition 4. For a good with a relatively inelastic demand ($\kappa_i > 1$), $-1 < \frac{\partial x_i^{\text{Standard}}}{\partial p_i}$. On the other hand, from (17), it is always the case that $\frac{\partial x_i^{\text{Explicit}}}{\partial p_i} = -1$. As the narrow thinker’s own-price demand elasticity is weighted average between the two, narrow thinking leads to excess sensitivity to own-price changes. In fact, narrow thinking can significantly increase the own-price demand elasticity of a good with an inelastic demand. For example, consider a good with $\frac{\partial x_i^{\text{Standard}}}{\partial p_i} = -0.1$. Based on the calibration of (21) above with $\lambda = 25\%$, we have $\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} = -0.3$, three times of $\frac{\partial x_i^{\text{Standard}}}{\partial p_i}$.

For a good with a relatively elastic demand ($\kappa_i < 1$),\(^\text{21}\) instead, the narrow thinker’s consumption $x_i$ decreases less in response to an increase in $p_i$. That is, for all $i$, $\frac{\partial x_i^{\text{Standard}}}{\partial p_i} < \frac{\partial x_i^{\text{Narrow}}}{\partial p_i}$. This is because an increase in $p_i$ now increases the consumption of other goods, $x_j$, as the substitution effect of $p_i$ on $x_j$ (positive) now dominates the income effect of $p_i$ on $x_j$ (negative). A increase in $x_j$ will then further decrease $x_i$. In this case, the indirect effect (negative) works in the same direction as the direct effect (negative). A dampening of the indirect effect under narrow thinking then leads to under-reaction, i.e. excess smoothness to own-price changes. Interestingly, in a recent paper (Hirshman, Pope and Song, 2018), the authors find that consumers exhibit excess sensitivity in response to gasoline price changes, but not in response to price changes of pens, glass cleaner and paper clips. It seems possible that the consumer’s demand with respect to gasoline is less elastic their setting: one can interpret $x_i$ as the consumption of a composite gasoline good, and $p_i$ as the consumption of a composite gasoline good. Consistent with their finding, my model predicts that the total gasoline consumption decreases excessively when the gasoline prices rise. Moreover, in Appendix C, I also explain how I can extend my setting to accommodate discrete choices among different grades of gasoline, as in Hastings and Shapiro (2013).

\(^\text{21}\)When $\kappa_i = 1$, the utility for good $i$ becomes the log utility. In this case, $\frac{\partial x_i^{\text{Standard}}}{\partial p_i} = \frac{\partial x_i^{\text{Narrow}}}{\partial p_i} = \frac{\partial x_i^{\text{Explicit}}}{\partial p_i} = 1$. That is, even in standard consumer theory, the consumer behaves as she has an explicit budget for good $i$. So does the narrow thinker.
(has a higher $\kappa_i$). As a result, the empirical finding is line with the prediction in Corollary 3.

**The boundary of a self.** In the main analysis, I interpret $x_i$ as the consumption of a good $i$, which is decided by self $i$. An alternative interpretation is that each $x_i$ captures the composite consumption for a spending category. Each self $i$ is then in charge of deciding consumption of all goods in that spending category. The above analysis can then be re-interpreted as a smooth model of mental accounting at the spending category level. See Appendix C for details.

This interpretation also begs the question about what defines the boundary of a self. As we interpret each self’s signal as capturing her state of mind (as discussed in Section 2), one should define the boundary of a self as a group of decisions made together based on the same state of mind. Such a definition also connects to the notion of “cognitive inertia” in Read, Loewenstein and Rabin (1999): if multiple decisions come to the decision maker one at a time, she will bracket them narrowly; if multiple decisions come to the decision maker collectively, she will bracket them broadly.

**Differences from other mental accounting models.** The existing formalization of mental accounting behavior mostly centers around the explicit budgeting model in (17), e.g. Heath and Soll (1996). Compared to this explicit budgeting model, my approach has four main differences. First, as discussed above, my approach provides a smooth model of mental accounting. Second, I do not need to model explicitly about how much money the decision maker assigns for each good or spending category. Third, my model also generates new predictions about what influences the degree of mental accounting behavior, as in Proposition 5. Fourth, my approach shows that mental accounting-type behavior arises because of the intra-personal difficulty in coordinating decisions. This connects mental accounting more broadly with narrow bracketing behavior. In Hastings and Shapiro (2013), Hastings and Shapiro (2018), and Farhi and Gabaix (2019), they provide an extension of the explicit budgeting model by allowing the consumer to deviate the budget subject to a cost. The last three points above about my approach also apply to this extension.

Koszegi and Matejka (2019) also provide an information-based theory of mental accounting. Their approach is complementary to mine, but there are also four main differences. First, they stay within the rational inattention paradigm: different decisions are based on the same, imperfect, information. Under narrow thinking, different decisions are based on different, non-nested, information. Second, they focus on providing conditions about when the decision maker endogenously chooses to behave as the explicit budgeting model in (17). My approach instead focuses on providing a smooth model of mental accounting. I also generate new predictions about what influences the degree of mental accounting behavior, as in Proposition 5. Third, in their approach, the interaction across different consumption decisions comes from the complementarity/substitutability embedded in the utility function, i.e. the second-order cross-derivatives of the utility function.
Here, the interaction across different decisions instead comes from the budget constraint. This setting helps me provide a more direct formalization about the violation of the fungibility principle, as in Lemma 3. Fourth, more broadly, I connect mental accounting with narrow bracketing behavior.

4.2 Other Mental Accounting Phenomena

Now, I show how the above smooth model of mental accounting can help explain other mental accounting-type behavior documented in the literature, such as excess smoothness to taste shocks, the label effect, and the small wage elasticity to labor supply. Through these applications, I also want to illustrate another point: depending on the environment, narrow thinking can translate into either over- or under-reaction relative to the frictionless benchmark. A rule of thumb is: when the indirect effect from the coordinated response of other decisions works in the same direction as the direct effect, a dampening of the indirect effect under narrow thinking leads to under-reaction. When the indirect effect works in the opposite direction of the direct effect, a dampening of the indirect effect under narrow thinking leads to over-reaction. This rule has already been applied in the discussion after Corollary 3 above, and will be formalized in Proposition 9 below.

**Excess smoothness to taste shocks.** Another behavior connected to mental accounting is excess smoothness to taste shocks. Consider an example in Heath and Soll (1996). A consumer goes to a store, wanting to buy a pair of trousers. She realizes that she does not like any trousers in the store (a negative taste shock), but still chooses to buy a pair. To see how such behavior is connected to mental accounting, notice that with explicit budgets in (17), the consumption of a good is insensitive to its taste.

Narrow thinking can provide a smooth model for such behavior. To formalize, let me add taste shocks in the environment in Section 4.1. The decision maker’s utility is given by

$$
\sum_{i=1}^{N} \varphi_i v_i(x_i) + h(y),
$$

where $v_i(x)$ and $h(y)$ are the same as in Section 4.1. Here I introduce taste shocks, and $\varphi_i \sim \log \mathcal{N} \left( \log \varphi_i, \sigma^2_{\varphi_i} \right)$ parametrizes the taste for good $i$. The decision maker still needs to satisfy the budget constraint, $\sum_{i=1}^{N} p_i x_i + y \leq w$. As I am interested in response to taste shocks, I treat $w$ and $p$s as constants.

Similar to Section 4.1, I consider the following narrow thinker: each self $i \in \{1, \cdots, N\}$ of the narrow thinker, who is in charge of purchasing good $i$, perfectly knows its taste $\varphi_i$, but receives a noisy signal about each of the other $\varphi_k$: $s_{i,k} = \varphi_k \epsilon_{i,k}$, with $\epsilon_{i,k} \sim \log \mathcal{N} \left( 0, \sigma^2_{i,k} \right)$ and $\sigma^2_{i,k} > 0$. 

25
Finally, similar to (7), I define the narrow thinker’s (log) demand function as $\hat{x}_i^{\text{Narrow}}(\hat{\varphi}_1, \cdots, \hat{\varphi}_N) = E[\hat{x}_i^*(\omega_i) \mid \hat{\varphi}_1, \cdots, \hat{\varphi}_N]$.

**Proposition 6** For each good $i$, the narrow thinker’s consumption $x_i$ increases (decreases) less in response to positive (negative) taste shocks to $\varphi_i$: for $i \in \{1, \cdots, N\}$,

$$\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \varphi_i} > \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \varphi_i} = \omega_i \frac{\partial \hat{x}_i^{\text{Explicit}}}{\partial \varphi_i} + (1 - \omega_i) \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \varphi_i} > 0,$$

where the demand with explicit budgeting $x_i^{\text{Explicit}}$ is still given by (17) and the weight $\omega_i \in [0, 1]$ is still given by (20).

Proposition 4 shows, in response to shocks to $\varphi_i$, the narrow thinker’s demand $x_i^{\text{Narrow}}$ is given by a weighted average between demand of standard consumer theory ($x_i^{\text{Standard}}$) and demand with explicit budgeting ($x_i^{\text{Explicit}}$). The intuition is similar to that in Subsection 4.1: in standard consumer theory, the response of $x_i$ to $\varphi_i$ depends crucially on the indirect effects from the coordinated response of other $x_j$; under narrow thinking, other selves will not respond much to $\varphi_i$; the response of $x_i$ to $\varphi_i$ is then closer that with explicit budgeting, which effectively treats each decision in isolation. Remarkably, the weight on $x_i^{\text{Explicit}}$, which captures the deviation from the standard consumer theory, shares the the *same* formula of (20). In other words, the discussion about what drives the degree of mental accounting in Proposition 5 still applies here.

Proposition 4 further establishes that the narrow thinker exhibits excess smoothness to taste shocks. To further understand the intuition behind the excess smoothness, consider a positive taste shock to $\varphi_i$. Under standard consumer theory, an increase in $\varphi_i$ not only increases $x_i$ (a positive direct effect) but also decreases the consumption of other goods $x_j$. The coordinated decrease of other consumption $x_j$ further increases $x_i$ (a positive indirect effect). The indirect effect then works in the same direction as the direct effect. Under narrow thinking, the indirect effect from the coordinated response of $x_j$ is dampened. The narrow thinker then exhibits excess smoothness to taste shocks (under-reaction).

**The label effect.** The same mechanism can also explain another behavior connected to mental accounting, “the label effect” (Beatty et al., 2014; Benhassine et al., 2015; Abeler and Marklein, 2016; Hastings and Shapiro, 2018). For example, Beatty et al. (2014) study the UK Winter Fuel Payment program. Despite its label, the fuel payment is in fact a mere cash transfer and there is no obligation to spend any of the payment on fuel despite the label. Beatty et al. (2014) nevertheless find that households increase their fuel consumption excessively after receiving the Winter Fuel Payment. To see how such behavior is connected to mental accounting, notice that, in the explicit budgeting model in (17), if the decision maker views the fuel payment as part of the budget $w_i$ allocated for fuel, she should spend all the payment on fuel consumption.
Narrow thinking can provide a smooth model for such behavior. To formalize, consider the environment in Section 4.1. That is, the consumer’s utility is given by (12). The decision maker is subject to the budget constraint: \( \sum_{i=1}^{N} x_i + y \leq w + \sum_{i=1}^{N} w_i \), where \( w \) is the decision maker’s generic wealth (treat as a constant) and \( w_i \) captures the part of the budget that is labelled for the consumption of good \( i \), e.g. the winter fuel payment in Beatty et al. (2014) and the beverage voucher in Abeler and Marklein (2016). For the narrow thinker, each self \( i \in \{1, \ldots, N\} \), who is in charge of purchasing good \( i \), perfectly knows \( w_i \sim \mathcal{N}(\bar{w}_i, \sigma^2_{w_i}) \), but receives a noisy signal about each of the other \( w_k \): \( s_{i,k} = w_k \epsilon_{i,k} \), with \( \epsilon_{i,k} \sim \log \mathcal{N}(0, \sigma^2_{\epsilon_{i,k}}) \) and \( \sigma^2_{\epsilon_{i,k}} > 0 \). This information structure captures the idea that decision maker has the winter Fuel Payment at the front of her mind when purchasing fuel, but not necessarily when making other purchases.

Now I will show how the narrow thinker’s consumption for each good \( i \), \( x_i \), is excessively sensitive to \( w_i \). Similar to (7), I define the narrow thinker’s (log) demand function as \( \hat{x}_i^{\text{Narrow}}(\hat{w}_1, \cdots, \hat{w}_N) \equiv E[\hat{x}_i^*(\omega_i) | \hat{w}_1, \cdots, \hat{w}_N] \).

**Proposition 7** The narrow thinker’s consumption \( x_i \) increases (decreases) more in response to positive (negative) shocks to \( w_i \):

\[
\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{w}_i} = \omega_i \frac{\partial \hat{x}_i^{\text{Explicit}}}{\partial \hat{w}_i} + (1 - \omega_i) \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{w}_i} > \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{w}_i} > 0 \quad \forall i,
\]

where the weight \( \omega_i \in [0, 1] \) is still given by (20).

Similar to Propositions 4 and 6, Proposition 7 shows, in response to shocks to \( w_i \), the narrow thinker’s demand \( \hat{x}_i^{\text{Narrow}} \) is given by a weighted average between the demand consumer theory (\( \hat{x}_i^{\text{Standard}} \)) and the demand with explicit budgeting (\( \hat{x}_i^{\text{Explicit}} \)). Here, by explicit budgeting, I mean the consumer uses the entirety of \( w_i \) on the consumption of good \( i \). Such a smooth model of mental accounting then leads to the excess sensitivity of the consumption \( x_i \) to \( w_i \). To further understand the intuition behind the excess sensitivity, note that in standard consumer theory, an increase in \( w_i \) will increase the consumption of both \( x_i \) (a positive direct effect) and other consumption \( x_j \). The increase in other consumption \( x_j \) then decreases \( x_i \) (a negative indirect effect). The indirect effect then works in the opposite direction of the direct effect, and the dampening of the indirect effect under narrow thinking leads to under-reaction.

**The small wage elasticity of daily labor supply.** Another behavior connected to mental accounting is the small wage elasticity of daily labor supply (Camerer et al., 1997, Farber, 2015, Thakral and To, Forthcoming). In the standard labor supply theory, when the wage on a particular day increases, the decision maker not only increases her labor supply on the day of wage increase and but also decreases her labor supply on other days. Such a coordinated response then further
increases the labor supply on the day of wage increase and generates large elasticity of daily labor supply. On the other hand, in an explicit income targeting model (Camerer et al., 1997), the decision maker assigns an income target for each day. In this case, the decision maker exhibits a negative wage elasticity of daily labor supply. Under narrow thinking, the decision’s wage elasticity of daily labor supply is a weighted average between those two cases. This prediction is consistent with the empirically documented positive, but small, wage elasticity of daily labor supply Farber (2015). Moreover, in line with Proposition 3, the smaller wage elasticity of labor supply under narrow thinking is about the response to temporary daily wage shocks. In fact, based on wage variations at longer frequencies, Fehr and Goette (2007) and Angrist, Caldwell and Hall (2017) find a larger wage elasticity of labor supply. See Appendix C for details.

5 General Properties

I now provide a few more general results about the optimal behavior under narrow thinking: narrow thinking smoothly attenuates the interaction across decisions; depending on the environment, it can lead to either over- or under-reaction.

5.1 Effective Attenuation of Interaction

From utility to the narrow thinker’s optimal decision rules. Consider the general environment in Section 2. That is, the consumer’s utility is given by is given by $u\left(\bar{x}, \bar{\theta}\right)$ in (1), with $\bar{x} = (x_1, \ldots, x_N)$ and $\bar{\theta} = (\theta_1, \ldots, \theta_M)$.

Motivated by the examples studied in previous sections, I will use the following information structure for the narrow thinker. Each self $i \in \{1, \ldots, N\}$, who is in charge of decision $i$, receives a noisy signal about each $\theta_k \sim \log \mathcal{N}(\log \bar{\theta}_k, \sigma^2_{\theta_k})$: $s_{i,k} = \theta_k \epsilon_{i,k}, \epsilon_{i,k} \sim \log \mathcal{N}(0, \sigma^2_{\epsilon_{i,k}})$, for $k \in \{1, \ldots, M\}$. That is, $\omega_i = \left\{s_{i,k}\right\}_{k \in \{1, \ldots, M\}}$. All fundamentals and noises are independent from each other. For all $i, k$, let me also introduce $\lambda_{i,k} \equiv \frac{\sigma^2_{\epsilon_{i,k}}}{\sigma^2_{\theta_k} + \sigma^2_{\epsilon_{i,k}}} \in [0, 1]$ to capture the precision of self $i$’s signal about $\theta_k$.

To derive the optimal decision rule for each self $i$ of the narrow thinker, I start from the decision-specific optimality in (3), take a first order condition, and log-linearize it. Similar to above, I use a hat over a variable to denote its log-deviation from the point where each fundamental $\theta_k$ is fixed at $\bar{\theta}_k$ and each decision is made with perfect knowledge of all fundamentals: $\{\hat{x}_i\}_{i=1}^N = \arg\max_{\{x_i\}_{i=1}^N} u\left(x_1, \ldots, x_N, \hat{\theta}_1, \ldots, \hat{\theta}_M\right)$.

---

22As discussed above, constrained problems, such as consumer theory examples studied above, can be transformed into the unconstrained problem here.
Lemma 4 For each self \( i \in \{1, \ldots, N\} \), her (log-linearized) optimal decision is given by

\[
\hat{x}_i^*(\omega_i) = E_i \left[ \sum_{1 \leq k \leq M} \psi_{i,k} \hat{\theta}_k \right] + E_i \left[ \sum_{j \neq i, 1 \leq j \leq N} \gamma_{i,j} \hat{x}_j^*(\omega_j) \right],
\]

where \( E_i [\cdot] = E[\cdot|\omega_i] \) denotes self \( i \)'s belief, \( \psi_{i,k} = -\frac{u_{x_i,\theta_k}}{u_{x_i,\hat{x}_i}} \hat{\theta}_k > 0 \), and \( \gamma_{i,j} = -\frac{u_{x_i,\hat{x}_j}}{u_{x_i,\hat{x}_i}} \).

The first term in (24), which I call the direct effect, summarizes the fundamental \( \hat{\theta} \)'s direct influence on self \( i \)'s decision, holding other selves' decisions fixed. For each \( i \in \{1, \ldots, N\} \) and \( k \in \{1, \ldots, M\} \), \( \psi_{i,k} \) captures how the fundamental \( \theta_k \) directly influences the optimal decision \( i \).

The second term in (24), which I call the indirect effect, summarizes how other selves’ decisions influence self \( i \)'s decision. For each \( i \neq j \in \{1, \ldots, N\} \), \( \gamma_{i,j} \) summarizes how decision \( i \) is influenced by decision \( j \). A positive (negative) \( \gamma_{i,j} \) means that optimal decision \( i \) increases (decreases) with decision \( j \). In fact, one can think of (24) as each self \( i \)'s best response function in the equivalent game among multiple selves. It is akin to the best response function in a linear network game (Bergemann, Heumann and Morris, 2017; Golub and Morris, 2017), and \( \Gamma = \{\gamma_{i,j}\}_{1 \leq i,j \leq N} \) can be interpreted as the interaction matrix across different decisions.\(^{24}\)

Effective attenuation of interaction. I now show that, in response to shocks to the fundamentals, narrow thinking smoothly attenuates the interaction across decisions \( i \). Similar to (7), for each \( i \in \{1, \ldots, N\} \), I define the narrow thinker’s (log-)decision function as

\[
\hat{x}_i^{\text{Narrow}}(\hat{\theta}_1, \ldots, \hat{\theta}_M) \equiv E \left[ \tilde{x}_i^*(\omega_i) | \hat{\theta}_1, \ldots, \hat{\theta}_M \right],
\]

averaging over the realization of noises in signals. Now, I show

Proposition 8 In response to shocks to \( \theta_k \), the narrow thinker’s decisions can be characterized by

\[
\left( \frac{\partial^2 \hat{x}_1^{\text{Narrow}}}{\partial \theta_k \partial \theta_l} \frac{\partial^2 \hat{x}_2^{\text{Narrow}}}{\partial \theta_k \partial \theta_l} \cdots \frac{\partial^2 \hat{x}_N^{\text{Narrow}}}{\partial \theta_k \partial \theta_l} \right)' = (\mathbb{I}_N - \hat{\Gamma})^{-1} \Psi_k,
\]

\(^{23}\)Here, \( u_{x_i,\hat{x}_k} = \frac{\partial^2 u(\bar{x}_1, \ldots, \bar{x}_1, \theta_1, \ldots, \theta_M)}{\partial x_i \partial \theta_k} \), \( u_{x_i,\hat{x}_i} = \frac{\partial^2 u(\bar{x}_1, \ldots, \bar{x}_1, \theta_1, \ldots, \theta_M)}{\partial (x_i)^2} \), and \( u_{x_i,\hat{x}_j} = \frac{\partial^2 u(\bar{x}_1, \ldots, \bar{x}_1, \theta_1, \ldots, \theta_M)}{\partial x_i \partial x_j} \).

\(^{24}\)For notation simplicity, I also set \( \gamma_{i,i} = 0 \) for all \( i \).
where $\tilde{\Gamma}_k$ captures the effective interaction matrix and $\Psi_k$ captures the direct effect

\[
\tilde{\Gamma}_k = \begin{pmatrix}
1 & \lambda_{1,k} & \cdots & \lambda_{1,k} \\
\lambda_{2,k} & 1 & \cdots & \lambda_{2,k} \\
& & \ddots & \vdots \\
\lambda_{N,k} & \lambda_{N,k} & \cdots & 1
\end{pmatrix} \circ \Gamma \quad \text{and} \quad \Psi_k = \begin{pmatrix}
\lambda_{1,k} \psi_{1,k} \\
\lambda_{2,k} \psi_{2,k} \\
& & \ddots \\
\lambda_{N,k} \psi_{N,k}
\end{pmatrix},
\]

where $\Gamma$ is the original interaction matrix above and $\circ$ is the element by element product.

The matrix $\tilde{\Gamma}_k$ captures the effective interaction across decisions in response to shocks to $\theta_k$. For each pair of decisions $(i, j)$, the effective degree of interaction from decision $j$ to decision $i$, $\tilde{\Gamma}_k(i, j)$, is smoothly attenuated by the factor $\lambda_{i,k} = \frac{\sigma_{i,k}^2}{\sigma_{i,k}^2 + \sigma_{i,k}^2} \in [0, 1]$, depending on the precision of the narrow thinker’s information. That is, in response to shocks to $\theta_k$, because self $i$ has an imperfect perception of self $j$’s decision, an one unit increase (decrease) in $x_j$ only effectively increases (decreases) $x_i$ by $\lambda_{i,k} \gamma_{i,j}$. It is as if self $i$ cares less about the influence of other decisions, and she “thinks narrowly.”

The vector $\Psi_k$ in (26) captures the direct effect of $\theta_k$ on each decision. As each self $i$ may not perfectly know $\theta_k$, The direct effect of $\theta_k$ on $x_i$ can also be dampened. To isolate the attenuation of interaction across decisions (the friction of interest) from this dampening of direct effects, we can consider the case in which $\theta_k$ only directly influences $x_k$ ($\psi_{i,k} = 0$ for $i \neq k$), and self $k$ perfectly knows $\theta_k$ ($\lambda_{k,k} = 1$), e.g. the consumer theory example in Section 3. Then, the direct effects of $\theta_k$ on decisions are maintained, and the sole friction comes from the effective attenuation of interaction above and the accompanying dampening of indirect effects.

It is true that a large number of parameters $\{\lambda_{i,k}\}_{i \neq k}$ govern the effective attenuation of interaction. This reflects the complexity of each self’s information environment, once we take into consideration bounded recall and selective retrieval from memory. In applications, one can use the following method to minimize the degree of freedom. One can let each self have perfect knowledge about one fundamental, i.e. $p_i$ for each self $i$ in the consumer theory example above, and set all other $\lambda_{i,k}$ to be the same $\lambda$. In fact, this is what I did in the calibration exercise in Corollary 2.

### 5.2 Over- and Under-reaction

I now provide a more general result about when narrow thinking leads to over-reaction and when it leads to under-reaction. It nests all applications in Sections 3 and 4.

---

25The result is also reminiscent of Bergemann, Heumann and Morris (2017): in a multiple-agent network game setting, they find that incomplete information attenuates the interaction across players.

26In this case, $\Psi_k = (0, \cdots, \psi_{k,k}, \cdots, 0)'$.  

30
Motivated by the above applications, I equate the number of fundamentals with the number of decisions \( M = N \), and impose that each self \( i \) perfectly knows fundamental \( \theta_i \) \( (\sigma^2_{i,i} = 0) \). In the consumer theory examples in Sections 3 and 4.1, this information structure just means that each self \( i \) perfectly knows the price of the good she buys, \( p_i \). For notation simplicity, I also assume \( \psi_{i,k} \geq 0 \) for all \( i \) and \( k \).\(^{27}\)

Here, I study how each of the narrow thinker’s decision \( x_i \) responds to shocks to \( \theta_i \), which she perfectly knows, e.g., the own-price demand elasticities in consumer theory. In Appendix D, I also study how each \( x_i \) responds to shocks to other \( \theta_k \).

From (24), we know the response of \( x_i \) to \( \theta_i \) can be decomposed into a direct and an indirect effect.

\[
\frac{\partial \tilde{x}_i^\text{Narrow}}{\partial \hat{\theta}_i} = \psi_{i,j} \frac{\partial E_i [\hat{x}_j] | \hat{\theta}_1, \ldots, \hat{\theta}_N}{\partial \hat{\theta}_i} \quad \text{Direct} + \sum_{j \neq i} \gamma_{i,j} \frac{\partial E_i [\hat{x}_j] | \hat{\theta}_1, \ldots, \hat{\theta}_N}{\partial \hat{\theta}_i} \quad \text{Indirect}
\]

As each self \( i \) knows \( \theta_i \), the direct effect is maintained under narrow thinking. On the other hand, as the above applications, the indirect effect is dampened. When the indirect effect works in the same direction as the direct effect, a dampening of the indirect effect then leads to under-reaction under narrow thinking. When the indirect effect works in the opposite direction to the direct effect, a dampening of the indirect effect then leads to over-reaction under narrow thinking.

To formalize, we use \( \tilde{x}_i^\text{Standard,Ind} \) to denote the indirect effect on decision \( i \) in (24), when each decision is made with perfect knowledge of all the fundamentals. As we normalize the direct effect of \( \theta_i \) on \( x_i \) to be positive, the indirect effect of \( \theta_i \) on \( x_i \) works in the same direction as the direct effect when \( \frac{\partial \tilde{x}_i^\text{Standard,Ind}}{\partial \theta_i} > 0 \). On the other hand, the indirect effect works in the opposition direction as the direct effect when \( \frac{\partial \tilde{x}_i^\text{Standard,Ind}}{\partial \theta_i} < 0 \).

**Assumption 1** At least one of the following condition is satisfied:

1) **Symmetry**, i.e. there exists \( \psi, \Psi > 0 \), \( \lambda \in (0, 1) \), and \( \gamma \in \left(-1, \frac{1}{N-1}\right) \), such that \( \psi_{i,i} = \psi \), \( \psi_{i,k} = \Psi \), \( \gamma_{i,j} = \gamma \), and \( \lambda_{i,j} = \lambda \) for all \( j, k \neq i \);

2) **Complements**, i.e. \( \gamma_{i,j} \geq 0 \) for all \( i \neq j \) and \( \sum_{j \neq i} \gamma_{i,j} < 1 \) for all \( i \);

3) **Substitutes with a single factor structure**, i.e. there exists non-negative scalars \( \{\rho_i, \Gamma_i, \Delta_i\}_{i=1}^N \) such that \( \gamma_{i,j} = -\rho_i \Gamma_j \) and \( \psi_{i,k} = \rho_i \Delta_k \) for all \( j, k \neq i \) and \( \rho_i \Gamma_i < 1 \) for all \( i \).\(^{28}\)\(^{29}\)

\(^{27}\)The analysis is the same when \( \psi_{i,k} \leq 0 \) for all \( i \) and \( k \), e.g. in the context of consumer theory. See the proof of Proposition 9.

\(^{28}\)To guarantee \( I - \Gamma \) is invertible, we impose \( \gamma \in \left(-1, \frac{1}{N-1}\right) \) in case 1), \( \sum_{j \neq i} \gamma_{i,j} < 1 \) for all \( i \) in case 2), and \( \rho_i \Gamma_i < 1 \) for all \( i \) in case 3).

\(^{29}\)For case 3), note that the “single factor structure” only restricts \( \psi_{i,k} \) for \( i \neq k \). On the other hand, \( \psi_{i,i} \) can be
Proposition 9 Suppose Assumption 1 holds,

a) When the indirect effect works in the same direction as the direct effect \( \frac{\partial \hat{x}_i^{\text{Standard,Ind}}}{\partial \theta_i} > 0 \), each decision of the narrow thinker under-reacts to shocks to its own fundamental,

\[
\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \theta_i} \leq \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \theta_i} \quad \forall i;
\]

b) When the indirect effect works in the opposite direction to the direct effect \( \frac{\partial \hat{x}_i^{\text{Standard,Ind}}}{\partial \theta_i} < 0 \), each decision of the narrow thinker over-reacts to shocks to its own fundamental,

\[
\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \theta_i} \geq \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \theta_i} \quad \forall i.
\]

The reason that additional conditions in Assumption 1 are required to establish Proposition 9 is because of the coexistence of opposing indirect effects. There could be some components of the indirect effect (e.g. through one decision \( x_{j1} \)) that positively influence the optimal decision \( i \) and there could be some components of the indirect effect (e.g. through another decision \( x_{j2} \)) that negatively influence the optimal decision \( i \). Dampening of each component (see Appendix D about the dampening of each component of the indirect effect) may not mean dampening of the net total. Each one of the additional conditions provided in Assumption 1 guarantees that one direction of the indirect effect dominates, and the net total of the indirect effect is dampened under narrow thinking. Depending on whether the net total of the indirect effect works in the same direction as the direct effect or not, narrow thinking then translates into over- or under- reaction.

In the proof of Proposition 9, I also show how all applications in Sections 3 and 4 satisfy Assumption 1. It is worth briefly explaining the third condition in Assumption 1, “Substitutes with a single factor structure.” This condition is satisfied when the interaction across decisions comes from a common source, such as the budget constraint in the consumer theory example with income effects in Section 4.1.

6 Additional Applications and Extensions

In this section, I study how narrow thinking can help explain two other types of narrow bracketing behavior: the neglect of “adding-up” effects and myopic loss aversion. They are deeply connected to the decision maker’s difficulty in coordinating her decisions, but are quite different from the mental accounting behavior studied in Section 2. At the end, I also provide a framework to endogeneize narrow thinking: in this problem, besides making the multiple-decisions, the decision
maker also chooses what information each decision is based upon, subject to a cognitive constraint. As different decisions are based on different decision rules, each self is “interested in” different parts of the fundamental. For example, in the simple consumer theory example above, each self wants to know more about the price of the good she buys. In this sense, narrow thinking arises endogenously.

**Neglect of “adding-up” effects.** One behavior often connected to narrow bracketing is the neglect of “adding-up” effects (Read, Loewenstein and Rabin, 1999). Consider the decision to smoke. The health consequence of a cigarette is small, but the cumulative health consequence of smoking can be large (i.e. the “adding-up” effects). Moreover, the cumulative benefit from smoking increases much more slowly than the cumulative costs. If the decision maker can perfectly coordinate all her smoking decisions, she will not smoke much. However, in practice, the decision maker decides on how much to smoke on different occasions separately, and may face difficulties in coordinating her smoking decisions.

Let me briefly summarize about how the narrow thinker may smoke excessively because she neglects the adding-up costs of smoking (see Appendix E for details). Specifically, I consider a decision maker who faces a convex cost based on total smoking. For the narrow thinker, each self is in charge of the smoking decision on one occasion, knows how attractive to smoke on that occasion, but does not know about the attractiveness on other occasions. When smoking becomes more attractive, the narrow thinker under-estimates how other selves will increase smoking, neglects the adding-up costs, and smokes excessively.

Theoretically, this application is different from those in Sections 3 - 5. Here I study the impact of a common shock to the attractiveness of smoking on all occasions. More generally, in response to common shocks to the fundamental, if different selves’ decisions are strategic substitutes, narrow thinking leads to overreaction relative to the frictionless benchmark. This case arises when the decision maker faces convex add-up costs (e.g. smoking) or concave add-up benefits (e.g. demand for variety in choices). On the other hand, if different selves’ decisions are strategic complements, narrow thinking leads to under reaction relative to the frictionless benchmark. This case arise when the decision maker faces convex add-up benefits (e.g. skill acquisition) and concave add-up costs (e.g. habituation).\(^{30}\)

**Myopic loss aversion.** Another behavior often connected to narrow bracketing is the decision maker’s aversion to combine small and favorable gamble (Samuelson, 1963). The existing explanation of this behavior, such as Benartzi and Thaler (1995), Barberis, Huang and Thaler (2006), and Rabin and Weizsacker (2009), contains two elements: first, the the decision maker suffers\(^{30}\)

\[^{30}\text{This relationship between strategic complementarity/substitutability and under-/over-reaction under narrow thinking only holds in response to a common shock. If the shock is idiosyncratic, as the case in Sections 3 - 5, we should reply on Proposition 9 to predict whether narrow thinking leads to over-reaction or under-reaction in a given environment.}\]
from loss aversion; second, she decides on each gamble in isolation.\footnote{Loss aversion alone is not enough, as the decision make can combine independent and favorable gambles to avoid the loss.} Narrow thinking provides a formalization of such myopic loss aversion, without directly requiring the decision maker to decide on each gamble in isolation.

Let me briefly summarize the analysis here (see Appendix E for details). I consider a decision maker with loss aversion. She faces two gambles. For each gamble, there is a 50\% chance that it turns out to be a loss of $1, and another 50\% chance that it turns out to be a gain of $\left(1 + \mu_i\right)$, where $\mu_i > 0$ is a random variable. Whether each gamble turns out to be a gain or a loss is independent from that of the other gamble. The benefits of combing two gambles is: if one gamble turns out to be a loss and another gamble turns out to be a gain, the decision maker who combines them will not suffer from the loss aversion. In fact, for a frictionless decision maker, she will coordinate her gambling decisions such that she either invests in two gambles together or does not invest in any of them.

For a narrow thinker, however, each self $i \in \{1, 2\}$ only knows the return of her gamble. She does not perfectly know the return of the other self’s gamble. Two selves cannot perfectly coordinate their gambling decisions. This difficulty in coordinating decisions makes it harder for the narrow thinker to enjoy the benefits of combing two gambles together. Together, this leads to a lower probability of investing in each gamble and provides a model of myopic loss aversion.\footnote{Theoretically, this application is different from those in Sections 3 - 5 because each self’s gambling decision is discrete.}

Endogenous narrow thinking. The previous analysis lets different decisions be made based on different, but exogenous, information. Here, I provide a framework to endogenize such information. In this problem, besides making the multiple-decisions, the decision maker also chooses what information each decision is based upon, subject to a cognitive constraint. The problem studies the optimal information choice problem at the decision-level, going beyond the standard rational inattention paradigm.

Same as Section 2, let $(S, F, P)$ be the underlying probability space. The decision maker’s utility is given by $u (\# x, \# \theta)$, where $u$ is twice continuously differentiable and strictly concave over $x \in X_1 \times \cdots \times X_N$ and $X_i$, a convex set on $\mathbb{R}$, denotes the set of possible decision $x_i$. The payoff relevant fundamental, $\bar{\theta}$, is the realization of an exogenously drawn random vector $\tilde{\theta} : S \rightarrow \Theta$, where $\Theta \subseteq \mathbb{R}^M$ denotes the set of possible fundamental. Here, for clarity, I use bold letters to denote random variables and normal letters to denote their realizations.

I then use $\omega_i$ to denote the signal (potentially multi-dimensional) under which each decision $i$ is made. $\omega_i$ is the realization of a random vector $\omega_i : S \rightarrow \Omega_i$, where $\Omega_i$ denotes the set of possible signal realizations for decision $i$. $\omega_i$, which summarizes how decision $i$’s signal is generated, is now
chosen endogenously from a set of random vectors $\Omega_i$.

Specifically, the decision maker chooses jointly the information upon which each decision is made upon $\{\omega_i \in \Omega_i\}_{i=1}^N$, and the decision rules $\{x_i(\cdot): \Omega_i \to X_i\}_{i=1}^N$. She maximizes her expected utility, subject to a cognitive constraint:

$$\max_{\{\omega_i \in \Omega_i, x_i(\cdot)\}_{i=1}^N} E \left[ u \left( x_1(\omega_1), \cdots, x_N(\omega_N), \vec{\theta} \right) \right] \tag{27}$$

s.t. $\sum_{i=1}^N I \left( \omega_i; \vec{\theta} \right) \leq \tau. \tag{28}$

In the cognitive constraint (28), $I \left( \omega_i; \vec{\theta} \right)$ denotes the mutual information between decision $i$'s signal $\omega_i$ and the fundamental $\vec{\theta}$, which equals to the entropy reduction $H \left( \vec{\theta} \right) - H \left( \vec{\theta} | \omega_i \right)$. It captures the cognitive cost for decision $i$. (28) then means the sum of cognitive costs used by all decisions $i$ cannot surpass the decision maker’s total cognitive capacity, $\tau$.

The above problem can be decomposed into two sub-problems. The first is about how decisions are made given the chosen information $\{\omega_i\}_{i=1}^N$. This sub-problem is the same as the one studied in Section 2, and the optimal decision rule can be characterized by (3). The second is about choosing the optimal information $\{\omega_i \in \Omega_i\}_{i=1}^N$ for each decision $i$, subject to the cognitive constraint in (28). One can henceforth interpret the problem in (27) as follows. The decision maker first chooses $\{\omega_i\}_{i=1}^N$, i.e. how each self $i$'s signal is generated, subject to the cognitive constraint in (28). Given the information structure $\{\omega_i\}_{i=1}^N$, different selves play the equivalent incomplete information Bayesian game defined in Proposition 1.

It is worth highlighting the difference of the problem in (27) under narrow thinking from the canonical rational inattention and sparsity paradigms. There, the decision maker decides what information about the fundamental to acquire subject to a cognitive constraint, but different decisions are based on the same information. The optimal information choice problem is at the decision-maker level. Here, the information is decision specific, and the optimal information choice problem is at the decision level. It captures the idea that, when the decision maker makes a particular decision, she cannot effortlessly use/recall the information used for other decisions.

The analysis of the problem in (27) is involved. For the problem to be analytically solvable, one can consider a quadratic or log-quadratic approximation of the utility function, and restrict signals to be normally or log-normally distributed. In Appendix F, I first revisit the simple consumer theory in Section 3. Now, I do not directly impose that each self $i$ has perfect knowledge of the price of the good she buys, $p_i$. I instead show that each self $i$ endogenously chooses to know more about $p_i$. In this sense, narrow thinking can arise endogenously. I then study a general problem in which I allow the signals to depend on the fundamental flexibly. The general insight is similar: as
different decisions are based on different decision rules, each self is “interested in” different parts of the fundamental; it is then optimal for different selves’ signals to take different forms, and narrow thinking arises endogenously.

7 Conclusion

Each decision maker faces multiple economic decisions, and makes these decisions separately. Nevertheless, in standard modeling practice, we implicitly assume perfect self-coordination among all these decisions. It is as if the decision maker determines all her decisions together. In this paper, I try to break such perfection. I develop an approach, narrow thinking, to capture the decision maker’s difficulty in coordinating her multiple decisions. The notion of narrow thinking is that different decisions are based on different, non-nested, information. This notion is motivated by the psychological observation that the decision maker may not incorporate all the relevant information when making each decision. Under narrow thinking, each decision of the decision maker is made with an imperfect understanding of other decisions. The friction then effectively attenuates the interaction across decisions and provides a model of narrow bracketing. Narrow thinking can then explain the related empirical phenomenon: mental accounting, the neglect of “adding-up” effects, and myopic loss aversion. It also generates new predictions about what drives the degree of frictional behavior, such as the degree of mental accounting.
Appendix A: Proofs

Proof of Lemma 1. The solution of (2) must satisfy the decision-by-decision optimality condition in (3). This proves the necessity part. Now we turn to the sufficiency. If the sufficiency is not true, consider \( \{x_1^* (\cdot), \ldots, x_N^* (\cdot)\} \) that satisfies the decision-by-decision optimality condition in (3) but is not the optimum in (2). We then have \( \{y_1^* (\cdot), \ldots, y_N^* (\cdot)\} \) such that

\[
E \left[ u \left( y_1^* (\omega_1), \ldots, y_N^* (\omega_N), \bar{\theta} \right) \right] > E \left[ u \left( x_1^* (\omega_1), \ldots, x_N^* (\omega_N), \bar{\theta} \right) \right].
\]

We then define

\[
f (t) = E \left[ u \left( x_1^* (\omega_1) + t (y_1^* (\omega_1) - x_1^* (\omega_1)), \ldots, x_N^* (\omega_1) + t (y_N^* (\omega_N) - x_N^* (\omega_N)), \bar{\theta} \right) \right].
\]

From the decision-by-decision optimality condition in (3) and the fact that \( u \) is twice continuously differentiable, we have, for all \( i \),

\[
E \left[ \frac{\partial u}{\partial x_i} \left( x_1^* (\omega_1), \ldots, x_N^* (\omega_N), \bar{\theta} \right) (y_i^* (\omega_i) - x_i^* (\omega_i)) \right] = 0.
\]

Moreover, we have

\[
f' (0) = \sum_{i=1}^{N} \left\{ E \left[ \frac{\partial u}{\partial x_i} \left( x_1^* (\omega_1), \ldots, x_N^* (\omega_N), \bar{\theta} \right) (y_i^* (\omega_i) - x_i^* (\omega_i)) \right] \right\}.
\]

Now, using law of iterated expectations, we have, for each \( i \),

\[
E \left[ \frac{\partial u}{\partial x_i} \left( x_1^* (\omega_1), \ldots, x_N^* (\omega_N), \bar{\theta} \right) (y_i^* (\omega_i) - x_i^* (\omega_i)) \right] = E \left[ E \left[ \frac{\partial u}{\partial x_i} \left( x_1^* (\omega_1), \ldots, x_N^* (\omega_N), \bar{\theta} \right) (y_i^* (\omega_i) - x_i^* (\omega_i)) \right] \right] = E \left[ E \left[ u \left( x_1^* (\omega_1), \ldots, x_N^* (\omega_N), \bar{\theta} \right) \right] (y_i^* (\omega_i) - x_i^* (\omega_i)) \right] = 0.
\]

As a result \( f' (0) = 0 \).

Because \( u \) is strictly concave over \( \bar{x} \), \( f (t) \) is also strictly concave. This means that \( t = 0 \) is the maximum of \( f (t) \). However, we have \( f (1) > f (0) = u \left( x_1^* (\omega_1), \ldots, x_N^* (\omega_N), \bar{\theta} \right) \). This is contradictory. In fact, this proposition is essentially Theorem 1 in Chapter 5 of Marschak and Radner (1972).

Comment: It is also useful to prove that the optimum of (2), if exists, is unique. If the optimum of (2) is not unique, consider two solutions of (2), \( \{x_1^* (\cdot), \ldots, x_N^* (\cdot)\} \) and \( \{y_1^* (\cdot), \ldots, y_N^* (\cdot)\} \), such that they differ with a non-zero probability. Now, consider \( \{z_1^* (\cdot), \ldots, z_N^* (\cdot)\} \), such that for all \( i \),

\[
z_i^* (\cdot) = \lambda x_i^* (\cdot) + (1 - \lambda) y_i^* (\cdot)
\]

with \( \lambda \in (0, 1) \). Because \( u \) is strictly concave over \( \bar{x} \), we have

\[
E \left[ u \left( z_1^* (\omega_1), \ldots, z_N^* (\omega_N), \bar{\theta} \right) \right] > E \left[ u \left( x_1^* (\omega_1), \ldots, x_N^* (\omega_N), \bar{\theta} \right) \right] \text{ and } E \left[ u \left( z_1^* (\omega_1), \ldots, z_N^* (\omega_N), \bar{\theta} \right) \right] > E \left[ u \left( z_1^* (\omega_1), \ldots, z_N^* (\omega_N), \bar{\theta} \right) \right].
\]

This contradicts with the optimality of \( \{x_1^* (\cdot), \ldots, x_N^* (\cdot)\} \) and \( \{y_1^* (\cdot), \ldots, y_N^* (\cdot)\} \).

\footnote{Uniqueness is in the sense that, in any two optima, decision rules are the same almost surely.}
Proof of Proposition 1. The optimality condition for each player \( i \) in the equivalent game is the same as the decision-specific optimality condition for decision \( i \) in (3). The equivalence between the Bayesian Nash Equilibrium in the Bayesian game played by multiple selves and the solution of (2) is then a direct corollary of Lemma 1.

Proof of Proposition 8. We first prove Proposition 8. This will help prove all results in Sections 3 and 4. For notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization.

Based on Lemma 1, I use guess and verify approach to find the unique optimum. I conjecture the optimal decision rule for each self \( i \), \( x_i^* (\omega_i) \), is linear in her signals,

\[
x_i^* (\omega_i) = \sum_{k=1}^{M} \alpha_{i,k} s_{i,k}.
\]

(29)

Given the information structure, we have, for all \( i \neq j \) and \( k \),

\[
E_i [s_{j,k}] = E_i [\theta_k] = \lambda_{i,k} s_{i,k},
\]

where \( \lambda_{i,k} = \frac{\sigma_{\theta_k}^2}{\sigma_{s_k}^2 + \sigma_{s_i}^2} \in (0, 1] \). We then have

\[
E_i [x_j^*] = \sum_{k=1}^{N} \lambda_{i,k} \alpha_{j,k} s_{i,k}.
\]

(30)

Together with the optimal decision rule in (24) and the guess in (29), we have, for all \( i \),

\[
x_i^* (\omega_i) = \sum_{1 \leq k \leq M} \psi_{i,k} \lambda_{i,k} s_{i,k} + \sum_{j \neq i} \gamma_{i,j} \sum_{k=1}^{M} \lambda_{i,k} \alpha_{j,k} s_{i,k}.
\]

For the guess in (29) to be valid, we then need to have, for all \( i, k \),

\[
\alpha_{i,k} = \psi_{i,k} \lambda_{i,k} + \sum_{j \neq i} \lambda_{i,k} \gamma_{i,j} \alpha_{j,k}.
\]

(31)

(31) is satisfied when

\[
\begin{pmatrix}
\alpha_{1,k} \\
\alpha_{2,k} \\
\vdots \\
\alpha_{N,k}
\end{pmatrix} = \left[ I_N - \left( \begin{array}{cccc}
1 & \lambda_{1,k} & \cdots & \lambda_{1,k} \\
\lambda_{2,k} & 1 & \cdots & \lambda_{2,k} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_{N,k} & \lambda_{N,k} & \cdots & 1
\end{array} \right) \circ \Gamma \right]^{-1} \begin{pmatrix}
\lambda_{1,k} \psi_{1,k} \\
\lambda_{2,k} \psi_{2,k} \\
\vdots \\
\lambda_{N,k} \psi_{N,k}
\end{pmatrix}.
\]
This verifies that the guess in (29) indeed characterizes the narrow thinker’s optimal decision rules. To prove Proposition 8, note that based on the definition in (25), we then have, for all $i$,

$$x_i^{\text{Narrow}}(\tilde{\theta}) = \sum_{k=1}^{N} \alpha_{i,k} \theta_k.$$ 

Taking partial derivative with respect to each $\theta_k$ then leads to Proposition 8.

**Comment.** In the proof, one may wonder why $I_N + \sum_{j \neq i}^{} u_{x_i,x_j} \alpha_{i,j,k} = \lambda_{i,k}^{-1} u_{x_i,\theta_k} \tilde{\theta}_k$, where $u_{i,j} = \frac{\partial^2 u}{\partial x_i \partial x_j}$. To prove $I_N - \left( \begin{array}{ccc} 1 & \lambda_{1,k} & \cdots & \lambda_{1,k} \\ \lambda_{2,k} & 1 & \cdots & \lambda_{2,k} \\ \vdots & \ddots & \ddots & \vdots \\ \lambda_{N,k} & \cdots & \lambda_{N,k} & 1 \end{array} \right) \circ \Gamma$ is invertible is then equivalent to prove $\left( \begin{array}{ccc} \lambda_{1,k}^{-1} & 1 & \cdots & 1 \\ 1 & \lambda_{2,k}^{-1} & \cdots & 1 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & \lambda_{N,k}^{-1} \end{array} \right) \circ U$ is invertible, where $U (i,j) = u_{i,j}$ is a negative definite matrix (as $u$ is strictly concave over $x$). Then note that,

$$U = U + \text{diag} \left\{ \left( \lambda_{1,k}^{-1} - 1 \right) u_{1,1}, \cdots, \left( \lambda_{N,k}^{-1} - 1 \right) u_{N,N} \right\}$$

This uses $\bar{x}_i \neq 0$ for all $i$. Otherwise log-linearization will be invalid.
is also negative definite. As a result,

\[
\begin{pmatrix}
\lambda_{1,k}^{-1} & 1 & \ldots & 1 & 1 \\
1 & \lambda_{2,k}^{-1} & \ldots & 1 & 1 \\
\vdots & & & \ddots & \vdots \\
1 & 1 & \ldots & 1 & \lambda_{N,k}^{-1}
\end{pmatrix}
\circ U
\]

and thus

\[
\begin{pmatrix}
1 & \lambda_{1,k} & \ldots & \lambda_{1,k} & \lambda_{1,k} \\
\lambda_{2,k} & 1 & \ldots & \lambda_{2,k} & \lambda_{2,k} \\
\vdots & & & \ddots & \vdots \\
\lambda_{N,k} & \lambda_{N,k} & \ldots & \lambda_{N,k} & 1
\end{pmatrix}
\circ \Gamma
\]

is invertible.

**Proof of Lemma 2, Proposition 2 and Corollary 1.** From the optimal consumption decisions (6) and similar to the proof of Proposition 8, we have: for \(i \in \{1,2\}\),

\[
\frac{\partial \hat{x}_{i}^\text{Narrow}}{\partial \hat{p}_i} = -\psi_i + \gamma_{i,-i} \frac{\partial \hat{x}_{-i}^\text{Narrow}}{\partial \hat{p}_i} \quad \text{and} \quad \frac{\partial \hat{x}_{-i}^\text{Narrow}}{\partial \hat{p}_i} = \lambda_{-i,i} \gamma_{-i,i} \frac{\partial \hat{x}_{i}^\text{Narrow}}{\partial \hat{p}_i}.
\]

Solving the above two equations, we have, for \(i \in \{1,2\}\),

\[
\frac{\partial \hat{x}_{i}^\text{Narrow}}{\partial \hat{p}_i} = \frac{-\psi_i}{1 - \lambda_{-i,i} \gamma_{-i,i}} \quad \text{and} \quad \frac{\partial \hat{x}_{-i}^\text{Narrow}}{\partial \hat{p}_i} = \frac{-\psi_i \lambda_{-i,i} \gamma_{-i,i}}{1 - \lambda_{-i,i} \gamma_{-i,i}}. \tag{32}
\]

As a result, for \(i \in \{1,2\}\),

\[
\begin{align*}
\frac{\partial \hat{x}_{i}^\text{Narrow}}{\partial \hat{p}_i} &= \omega_i \frac{\partial \hat{x}_{i}^\text{Neglect}}{\partial \hat{p}_i} + (1 - \omega_i) \frac{\partial \hat{x}_{i}^\text{Standard}}{\partial \hat{p}_i}, \tag{33} \\
\frac{\partial \hat{x}_{-i}^\text{Narrow}}{\partial \hat{p}_i} &= \lambda_{-i,i} \left[ \omega_i \frac{\partial \hat{x}_{i}^\text{Neglect}}{\partial \hat{p}_i} + (1 - \omega_i) \frac{\partial \hat{x}_{i}^\text{Standard}}{\partial \hat{p}_i} \right], \tag{34}
\end{align*}
\]

where

\[
\omega_i = \frac{1 - \lambda_{-i,i}}{1 - \lambda_{-i,i} \gamma_{-i,i}} \in [0,1], \tag{35}
\]

from the definition of \(x_{i}^\text{Neglect}\) in (8),

\[
\frac{\partial \hat{x}_{i}^\text{Neglect}}{\partial \hat{p}_i} = -\psi_i \quad \text{and} \quad \frac{\partial \hat{x}_{-i}^\text{Neglect}}{\partial \hat{p}_i} = 0.
\]
and

\[
\frac{\partial \hat{x}^\text{Standard}_{-i}}{\partial \hat{p}_i} = \frac{-\psi_i}{1 - \gamma_{i,-i}\gamma_{-i,i}} \quad \text{and} \quad \frac{\partial \hat{x}_i^\text{Standard}}{\partial \hat{p}_i} = \frac{-\psi_i \gamma_{-i,i}}{1 - \gamma_{i,-i}\gamma_{-i,i}}.
\]

The comparative statics in Proposition 2 follow directly from the formula of \( \omega_i \) in (9). Corollary 1 follows directly from the formula of \( \frac{\partial \hat{x}^\text{Narrow}_{-i}}{\partial \hat{p}_i} \) and \( \frac{\partial \hat{x}^\text{Narrow}_i}{\partial \hat{p}_i} \) in (32). Lemma 2 follows from (32) when \( \sigma^2_{1,2}, \sigma^2_{2,1} = +\infty \) \( (\lambda_{1,2}, \lambda_{2,1} = 0) \).

**Proof of Proposition 3.** As noted in main text, this Proposition holds within the context of the log-linearized decision rule in (6). Taking an unconditional expectation of (6) averaging over the realization of all fundamentals and signals, we have

\[
E[\hat{x}^\text{Narrow}_i] = E[-\psi_i \hat{p}_i] + \gamma_{i,-i}E[\hat{x}^\text{Narrow}_{-i}] \quad \forall i \in \{1, 2\},
\]

where the law of iterated expectation is used. The above condition also holds in standard consumer theory.

\[
E[\hat{x}^\text{Standard}_i] = E[-\psi_i \hat{p}_i] + \gamma_{i,-i}E[\hat{x}^\text{Standard}_{-i}] \quad \forall i \in \{1, 2\}.
\]

We then have

\[
E[\hat{x}^\text{Narrow}_i] = E[\hat{x}^\text{Standard}_i] \quad \forall i \in \{1, 2\}.
\]

**Proof of Lemma 3.** Under narrow thinking, if condition (16) holds, averaging over the realization of noises in signals, we then have

\[
-\kappa_i \hat{x}^\text{Narrow}_i (\hat{p}_1, \cdots, \hat{p}_N) - \hat{p}_i = -\kappa_i \hat{x}^\text{Narrow}_j (\hat{p}_1, \cdots, \hat{p}_N) - \hat{p}_j,
\]

and

\[
-\kappa_i \frac{\partial \hat{x}^\text{Narrow}_i}{\partial \hat{p}_i} - 1 = -\kappa_i \frac{\partial \hat{x}^\text{Narrow}_j}{\partial \hat{p}_i}.
\]

This is inconsistent with the formula about price elasticities in the proof of Proposition 4.

**Proof of Proposition 4 and Corollary 3.** From conditions (13) and (14), and similar to the proof of Proposition 8, we have, for all \( i \),

\[
-\kappa_i \frac{\partial \hat{x}^\text{Narrow}_i}{\partial \hat{p}_i} = 1 - \kappa_y \frac{\partial \hat{y}^\text{Narrow}}{\partial \hat{p}_i},
\]

\[
= 1 + \frac{\kappa_y}{\mu_y} \left( \sum_{j=1}^{N} \mu_j \frac{\partial \hat{z}^\text{Narrow}_j}{\partial \hat{p}_i} + \mu_i \right),
\]

and

\[
\frac{\partial \hat{y}^\text{Narrow}}{\partial \hat{p}_i} = \frac{\partial \hat{z}^\text{Narrow}_{-i}}{\partial \hat{p}_i},
\]

\[
= \frac{\partial \hat{z}^\text{Narrow}_i}{\partial \hat{p}_i},
\]

\[
= \frac{\partial \hat{z}^\text{Narrow}_j}{\partial \hat{p}_i} = \frac{\partial \hat{z}^\text{Narrow}_{-i}}{\partial \hat{p}_i}.
\]
and for all $i \neq k$,

$$-\kappa_k \frac{\partial \hat{x}^\text{Narrow}_k}{\partial \hat{p}_i} = -\kappa_y \lambda_{k,i} \frac{\partial \hat{y}^\text{Narrow}_k}{\partial \hat{p}_i} + \kappa_y (1 - \lambda_{k,i}) \mu_k \frac{\partial \hat{x}^\text{Narrow}_k}{\partial \hat{p}_i},$$

$$= \kappa_y \mu_k \frac{\partial \hat{x}^\text{Narrow}_k}{\partial \hat{p}_i} + \lambda_{k,i} \left( \sum_{j \neq k} \mu_j \frac{\partial \hat{x}^\text{Narrow}_j}{\partial \hat{p}_i} + \mu_i \right).$$

Together, we have, for all $i$,

$$\frac{\partial \hat{x}^\text{Narrow}_i}{\partial \hat{p}_i} = - 1 - \frac{1 - \kappa_i}{\kappa_i} \left( \sum_{j \neq i} \mu_j \frac{\mu_j}{\kappa_j} \frac{\lambda_{j,i}}{\gamma_j + \nu_j} + \mu_y \frac{\mu_y}{\kappa_y} \right),$$

$$\frac{\partial \hat{y}^\text{Narrow}_i}{\partial \hat{p}_i} = \kappa_y \left( \frac{\mu_i}{\kappa_i} + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\lambda_{j,i}}{\gamma_j + \nu_j} + \mu_y \frac{\mu_y}{\kappa_y} \right).$$

And for all $i \neq k$,

$$\frac{\partial \hat{x}^\text{Narrow}_k}{\partial \hat{p}_i} = \kappa_k + \frac{\mu_y}{\kappa_y} (1 - \lambda_{k,i}) \mu_k \left( \frac{\mu_i}{\kappa_i} + \sum_{j \neq k} \frac{\mu_j}{\kappa_j} \frac{\lambda_{j,i}}{\gamma_j + \nu_j} + \mu_y \frac{\mu_y}{\kappa_y} \right).$$

From (36), we have, for $i \in \{1, \ldots, N\}$,

$$\frac{\partial \hat{x}^\text{Narrow}_i}{\partial \hat{p}_i} = \omega_i \frac{\partial \hat{x}^\text{Explicit}_i}{\partial \hat{p}_i} + (1 - \omega_i) \frac{\partial \hat{x}^\text{Standard}_i}{\partial \hat{p}_i},$$

where

$$\omega_i = 1 - \left( \sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\mu_y}{\kappa_y} \frac{\lambda_{j,i}}{\gamma_j + \nu_j} + \mu_y \frac{\mu_y}{\kappa_y} \right) \left( \sum_{j \neq i} \frac{\mu_j}{\kappa_j} + \mu_y \frac{\mu_y}{\kappa_y} \right),$$

$$\frac{\partial \hat{x}^\text{Standard}_i}{\partial \hat{p}_i} = -1 - \frac{1 - \kappa_i}{\kappa_i} \left( \sum_{j \neq i} \frac{\mu_j}{\kappa_j} + \frac{\mu_y}{\kappa_y} \right) \frac{\lambda_{j,i}}{\gamma_j + \nu_j},$$

and $\frac{\partial \hat{x}^\text{Explicit}_i}{\partial \hat{p}_i} = -1$ based on (17). This proves 4.

To prove Corollary 3, note that $\frac{\partial \hat{x}^\text{Standard}_i}{\partial \hat{p}_i} > -1$ when $\kappa_i > 1$ and $\frac{\partial \hat{x}^\text{Standard}_i}{\partial \hat{p}_i} < -1$ when $\kappa_i < 1$. Corollary 3 then follows from Proposition 4.

**Proof of Proposition 5.** First define $f(a, b, c, x) = \frac{(a+x)(b+c+x)}{(a+c+x)(b+x)}$ where $a, b, c, x > 0$ and $b > a$. Based on the formula for $\omega_i$ in (20), we have $\omega_i = 1 - f(a, b, c, x)$ where $a = \sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\mu_y}{\kappa_y} \frac{\lambda_{j,i}}{\gamma_j + \nu_j}, b = \sum_{j \neq i} \frac{\mu_j}{\kappa_j}, c = \frac{\mu_y}{\kappa_y}$, and $x = \frac{\mu_y}{\kappa_y}$.

To prove (i), we need that $f(a, b, c, x)$ increases in $a$, which is true because $b > a$.
To prove (ii), we need that $f(a, b, c, x)$ increases in $x$. This is also true because

$$\frac{\partial f(a, b, c, x)}{\partial x} = \frac{(a + b + c) + 2x}{(b + x)(a + c + x)} - \left[\frac{(a + b + c) + 2x}{(b + x)(a + c + x)}\right]^2 > 0.$$ 

To prove (iii), we need that $f(a, b, c, x)$ decreases in $c$, which is true because $b > a$.

To prove (iv), we start from (20) and take the limit of $\lambda_{j,i} \to 0$ for all $j \neq i$ and $\frac{\mu_u}{\kappa_y} \to 0$.

**Proof of Corollary 2.** When, for all $i \neq k \in \{1, \cdots, N\}$, $\kappa_i = \kappa_y = \kappa$ and $\lambda_{i,k} = \lambda$, the formula for $\omega_i$ in (20) becomes

$$\omega_i = 1 - \frac{(\sum_{j \neq i} \frac{\mu_j}{\mu_y} \frac{\mu_i \lambda}{\mu_y (1-\lambda) \mu_1} + 1) \left( \frac{\mu_x}{\mu_y} + 1 \right)}{(\sum_{j \neq i} \frac{\mu_j}{\mu_y} \frac{\lambda \mu_1}{\mu_y (1-\lambda) \mu_1} + \frac{\mu_i}{\mu_y} + 1) \left( \frac{\mu_x}{\mu_y} (1 - \frac{\mu_i}{\mu_y}) + 1 \right)} \geq 1 - \frac{(\sum_{j \neq i} \frac{\mu_j}{\mu_i} \lambda + \frac{\mu_i}{\mu_y} + 1) \left( \frac{\mu_x}{\mu_y} (1 - \frac{\mu_i}{\mu_y}) + 1 \right)}{(1 - \lambda) \frac{\mu_x}{\mu_y} \frac{\mu_i}{\mu_y} + \lambda \frac{\mu_i}{\mu_y} + 1) \left( \frac{\mu_x}{\mu_y} (1 - \frac{\mu_i}{\mu_y}) + 1 \right)} = \omega_i$$

where $\mu_x = \sum_{i=1}^N \mu_i$.

**Proof of Proposition 6.** The optimality condition for each decision $i$ and the budget constraint become

$$\hat{\varphi}_i - \kappa_i \hat{x}_i^* (\omega_i) = -\kappa_h \bar{E}_i [\hat{y}^*] \quad \forall i,$$

$$\sum_{i=1}^N \mu_i \hat{x}_i^* (\omega_i) + \mu_y \hat{y}^* = 0,$$

where, as in Section 4.1, $\kappa_h = -\frac{h''(\bar{y})}{h'(\bar{y})}$, $\mu_i = \frac{\bar{t}_i}{\bar{w}}$, and $\mu_y = \frac{\bar{y}}{\bar{w}}$. Similar to the proof of Proposition 8, we have, for all $i$ and $k \neq i$,

$$1 - \kappa_i \frac{\partial \hat{x}_i^{Narrow}}{\partial \hat{\varphi}_i} = \frac{\kappa_y}{\mu_y} \left( \sum_{j=1}^N \mu_j \frac{\partial \hat{x}_j^{Narrow}}{\partial \hat{\varphi}_i} \right),$$

$$-\kappa_k \frac{\partial \hat{x}_k^{Narrow}}{\partial \hat{\varphi}_i} = \frac{\kappa_y}{\mu_y} \left[ \mu_k \frac{\partial \hat{x}_k^{Narrow}}{\partial \hat{\varphi}_i} + \lambda_{k,i} \left( \sum_{j \neq k} \mu_j \frac{\partial \hat{x}_j^{Narrow}}{\partial \hat{\varphi}_i} + \mu_i \right) \right].$$
where \( \lambda_{k,i} = \frac{\sigma_{w}^2}{\sigma_{w}^2 + \sigma_{k,i}^2} \). Solving the above two equations, we have for all \( i \) and \( k \neq i \),

\[
\frac{\partial \hat{x}_{i}^{\text{Narrow}}}{\partial \hat{\varphi}_i} = \frac{1}{\kappa_i} \frac{\sum_{j \neq i} \mu_j \frac{\mu_j}{\kappa_j} \frac{\lambda_{j,i}}{\kappa_j} + \mu_y}{\kappa_i + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\lambda_{j,i}}{\kappa_j} + \frac{\mu_y}{\kappa_y}},
\]

\[
\frac{\partial \hat{x}_{k}^{\text{Narrow}}}{\partial \hat{\varphi}_i} = \frac{\lambda_{k,i}}{\kappa_k + \sum_{j \neq i} \mu_j \frac{\lambda_{j,i}}{\kappa_j} + \mu_y}.
\]

We then have for \( i \in \{1, \ldots, N\} \),

\[
\frac{\partial \hat{x}_{i}^{\text{Narrow}}}{\partial \hat{\varphi}_i} = \omega_i \frac{\partial \hat{x}_{i}^{\text{Explicit}}}{\partial \hat{\varphi}_i} + (1 - \omega_i) \frac{\partial \hat{x}_{i}^{\text{Standard}}}{\partial \hat{\varphi}_i},
\]

where the weight \( \omega_i \in [0, 1] \) is still given by (20), \( \frac{\partial \hat{x}_{i}^{\text{Standard}}}{\partial \hat{\varphi}_i} = \frac{1}{\kappa_i + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} + \frac{\mu_y}{\kappa_y}} > 0 \), and \( \frac{\partial \hat{x}_{i}^{\text{Explicit}}}{\partial \hat{\varphi}_i} = 0 \) based on (17).

We then have \( \frac{\partial \hat{x}_{i}^{\text{Standard}}}{\partial \hat{\varphi}_i} > \frac{\partial \hat{x}_{i}^{\text{Narrow}}}{\partial \hat{\varphi}_i} \). This proves Proposition 6.

**Proof of Proposition 7.** The log-linearized optimal decision rule for each consumption \( x_i^* (\omega_i) \) and the budget constraint are given by:

\[
-k_i \hat{x}_i^*(\omega_i) = -k_y \bar{E}_i [\bar{y}^*],
\]

\[
N \sum_{i=1}^{N} \mu_i \hat{x}_i^*(\omega_i) + \mu_y \bar{y}^* = \sum_{i=1}^{N} \mu_i^w \bar{w}_i,
\]

where \( \kappa_h = -\frac{h''(\bar{y})}{h'(\bar{y})} \), \( \mu_i = \frac{\mu_i}{\mu_y} \), \( \mu_y = \frac{\bar{y}}{\bar{w}} \), and \( \mu_i^w = \frac{\bar{w}}{\bar{w}} \). Similar to the proof of Proposition 8, we have, for all \( i \) and \( k \neq i \),

\[
-k_i \frac{\partial \hat{x}_{i}^{\text{Narrow}}}{\partial \hat{w}_i} = \frac{\kappa_y}{\mu_y} \left( \sum_{j=1}^{N} \mu_j \frac{\partial \hat{x}_{j}^{\text{Narrow}}}{\partial \hat{w}_i} - \mu_i^w \right),
\]

\[
-k_k \frac{\partial \hat{x}_{k}^{\text{Narrow}}}{\partial \hat{w}_i} = \frac{\kappa_y}{\mu_y} \left( \mu_k \frac{\partial \hat{x}_{k}^{\text{Narrow}}}{\partial \hat{w}_i} + \lambda_{k,i} \left( \sum_{j \neq k} \mu_j \frac{\partial \hat{x}_{j}^{\text{Narrow}}}{\partial p_i} - \mu_i^w \right) \right),
\]

where \( \lambda_{k,i} = \frac{\sigma_{w}^2}{\sigma_{w}^2 + \sigma_{k,i}^2} \). Solving the above two equations, for all \( i \),

\[
\frac{\partial \hat{x}_{i}^{\text{Narrow}}}{\partial \hat{w}_i} = \frac{\kappa_y}{\kappa_i} \left( \frac{\mu_i}{\kappa_i} + \sum_{j \neq i} \frac{\mu_j}{\kappa_j} \frac{\lambda_{j,i}}{\kappa_j} + \frac{\mu_y}{\kappa_y} \right) \frac{\mu_i^w}{\mu_i^w}.
\]


We then have for $i \in \{1, \cdots, N\}$,

$$
\frac{\partial x^\text{Narrow}_i}{\partial \hat{w}_i} = \omega_i \frac{\partial x^\text{Explicit}_i}{\partial \hat{w}_i} + (1 - \omega_i) \frac{\partial x^\text{Standard}_i}{\partial \hat{w}_i},
$$

where the weight $\omega_i \in [0, 1]$ is still given by (20), $\frac{\partial x^\text{Standard}_i}{\partial \hat{w}_i} = \frac{\kappa_i}{\kappa_i} \left( \frac{\mu^w_i}{\mu^w_i + \sum_{j \neq i} \mu^w_j + \gamma^w_i} \right) > 0$, and $\frac{\partial x^\text{Explicit}_i}{\partial \hat{w}_i} = \frac{\mu^w_i}{\mu^w_i}.
$

Here, by explicit budgeting, I mean the consumer uses the entirety of $\hat{w}_i$ on the consumption of good $i$.

The last expression then comes from the log-linearization.

Because $\frac{\partial x^\text{Explicit}_i}{\partial \hat{w}_i} > \frac{\partial x^\text{Standard}_i}{\partial \hat{w}_i}$, we have $\frac{\partial x^\text{Narrow}_i}{\partial \hat{w}_i} > \frac{\partial x^\text{Standard}_i}{\partial \hat{w}_i}$. This proves Proposition 7.

**Proof of Lemma 4.** I start from the decision-specific optimality in (3), and take a first order condition:

$$
E \left[ \frac{\partial u}{\partial x_i} \left( x^*_i(\omega_i), \bar{x}^*, \vec{\theta} \right) \right] = 0 \quad \forall i, \omega_i \in \Omega_i.
$$

Log-linearize this condition, I arrive at (24).

**Proof of Proposition 9.** For notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. I prove the Proposition 9 case by case. In the case (1) “Symmetry,” there exists $\psi, \Psi > 0$, $\lambda \in (0, 1)$, and $\gamma \in (-1, \frac{1}{N-1})$, such that $\psi_{i,i} = \psi$, $\psi_{i,k} = \Psi$, $\gamma_{i,j} = \gamma$, and $\lambda_{i,j} = \lambda$ for all $j, k \neq i$.

From (26), we can express the own-sensitivity as

$$
\frac{\partial x^\text{Narrow}_i}{\partial \theta_i} = \psi + \gamma (N - 1) \frac{\lambda \Psi + \lambda \gamma \psi (N - 1)}{1 - \lambda \gamma^2 (N - 1) - \lambda \gamma (N - 2)} \quad \forall i,
$$

with $\frac{\partial x^\text{Standard, Ind}_i}{\partial \theta_i} = \gamma (N - 1) \frac{\Psi + \psi \gamma \psi (N - 1)}{1 - \gamma^2 (N - 1) - \gamma (N - 2)}$. Using the fact that $\lambda \in [0, 1)$, $\psi, \Psi > 0$ and $\gamma \in (-1, \frac{1}{N-1})$, Proposition 9 follows directly.

In the case (2) “Complements,” we have $\gamma_{i,j} \geq 0$ for all $i \neq j$ and $\sum_{j \neq i} \gamma_{i,j} < 1$ for all $i$. In this case, the game among multiple selves are solvable by iterating best response. From (26), we have

$$
\frac{\partial x^\text{Narrow}_i}{\partial \theta_i} = \psi_{i,i} + \sum_{j \neq i} \gamma_{i,j} \psi_{j,i} + \sum_{j \neq i} \gamma_{i,j} \sum_{l \neq j} \lambda_{j,l} \gamma_{j,l} \psi_{l,i} + \cdots .
$$

As each term in (42) is non-negative, the indirect effect always works in the same direction as the direct effect, and the result follows directly.

In the case (3) “Substitutes with a single factor structure,” there exists non-negative scalars $\{\rho_i, \Gamma_i, \Delta_i\}_{i=1}^N$ such that $\gamma_{i,j} = -\rho_i \Gamma_j$ and $\psi_{i,k} = \rho_i \Delta_k$ for all $j, k \neq i$ and $\rho_i \Gamma_i < 1$ for all $i$. Define $y_{-i} = \sum_{j \neq i} \Gamma_j x_j$ for

\[\text{This means } 1 - \lambda \gamma^2 (N - 1) - \lambda \gamma (N - 2) > 0.\]
all $i$. Based on the proof of Proposition 8, we have, for all $i \neq k$,

\[
\frac{\partial x_i}{\partial \theta_i} = \psi_{i,i} - \rho_i \frac{\partial y_{-i}}{\partial \theta_i},
\]

\[
\frac{\partial x_j}{\partial \theta_i} = \lambda_{j,i} \psi_{j,i} - \rho_j \lambda_{j,i} \left( \Gamma_i \frac{\partial x_i}{\partial \theta_i} + \frac{\partial y_{-i}}{\partial \theta_i} - \Gamma_j \frac{\partial x_j}{\partial \theta_i} \right)
\]

\[
\frac{\partial y_{-i}}{\partial \theta_i} = \sum_{j \neq i} \Gamma_j \frac{\partial x_j}{\partial \theta_i}.
\]

Together, we have

\[
\frac{\partial x_i}{\partial \theta_i} = \psi_{i,i} - \rho_i \left( \Delta_i - \Gamma_i \psi_i \right) \frac{\sum_{j \neq i} \lambda_{j,i} \rho_j \Gamma_j}{1 + \sum_{j \neq i} \lambda_{j,i} \rho_j \Gamma_j},
\]

with $\frac{\partial u_i}{\partial \theta_i} = -\rho_i \left( \Delta_i - \Gamma_i \psi_i \right) \frac{\sum_{j \neq i} \lambda_{j,i} \rho_j \Gamma_j}{1 + \sum_{j \neq i} \lambda_{j,i} \rho_j \Gamma_j}$. Using the fact that $1 - \lambda_{j,i} \rho_j \Gamma_j > 0$ and $\frac{\sum_{j \neq i} \lambda_{j,i} \rho_j \Gamma_j}{1 + \sum_{j \neq i} \lambda_{j,i} \rho_j \Gamma_j}$ increases with each $\lambda_{j,i}$, the result follows.

Now, let me show that all applications in Sections 3 and 4 satisfy Assumption 1.

i) When $u_{x_1,x_2} > 0$, two goods are complements, from the optimal decision rule (6), the environment falls into case 2) in Assumption 1.

When $u_{x_1,x_2} < 0$, two goods are substitutes, from the optimal decision rule (6), the environment falls into case 3) in Assumption 1, with $\rho_i = -\gamma_{i,-i}$, $\Gamma_i = 1$ and $\Delta_i = 0$.

ii) For the environment in Section 4.1, based on each self’s optimal consumption rule in (18), it falls into case 3) of Assumption 1, with $\rho_i = \frac{\kappa_{x_i}}{1 + \frac{\kappa_{x_i}}{\mu_x}}$ and $\Gamma_i = \Delta_i = \frac{\mu_i}{\mu_y}$.

iii) For the environment in Proposition 6, from (37) and (38), we know each self’s optimal consumption rule can be written as

\[
x_i^*(\omega_i) = \frac{1}{1 + \frac{\mu_i}{\mu_y}} \hat{x}_i^* - E_i \left[ \sum_{j \neq i} \frac{1}{1 + \frac{\mu_i}{\mu_y}} \hat{x}_j^*(\omega_j) \right].
\]

The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\mu_i}{1 + \frac{\mu_i}{\mu_y}}$, $\Gamma_i = \frac{\mu_i}{\mu_y}$ and $\Delta_i = 0$.

iv) For the environment in Proposition 7, from (37) and (38), we know each self’s optimal consumption rule can be written as

\[
\hat{x}_i^* (\omega_i) = E_i \left[ \sum_{j=1}^{N} \frac{\mu_i}{\mu_y} \hat{x}_j^* - \sum_{j \neq i} \frac{\mu_i}{\mu_y} \hat{x}_j^* (\omega_j) \right],
\]

The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\mu_i}{1 + \frac{\mu_i}{\mu_y}}$, $\Gamma_i = \frac{\mu_i}{\mu_y}$ and $\Delta_i = \frac{\mu_i}{\mu_y}$.

It is worth noting that for case 3) of Assumption 1, the “single factor structure” only restricts $\psi_{i,k}$ for $i \neq k$. On the other hand, $\psi_{i,i}$, can be any non-negative scalar.
Appendix B: A Simple Consumer Theory Example: Narrow Thinking Implies Narrow Bracketing

Identification of Demand Elasticities.

Consider an environment with \( K \) consumers. All consumers have the same utility as in Section 3. There are \( T \) periods. In each period, each consumer solves the same consumer problem with a newly drawn price vector. Specifically, the price vector faced by consumer \( k \in \{1, \cdots, K\} \) at period \( t \in \{1, \cdots, T\} \), \( p_{k,t} \), is drawn i.i.d (across time, consumers and goods) from consumer \( k \)'s price distribution, \( \log N(\log p_{k,1}, \sigma_{p_k}^2) \).

Same as Section 3, each self \( i \in \{1, 2\} \) of the consumer \( k \) at period \( t \) perfectly knows the price of the good she buys \( p_{k,t} \), but only receives a noisy signal about the other price faced by her other self, \( p_{k-2,t} \).

All consumers share the same signal-to-noise ratio of their signals (thus same \( \{\lambda_{i,k}\} \)) in each period. As different consumers have the same utility and same \( \lambda \),s, they all share the same demand elasticities in response to price shocks. However, the mean demand for each good \( i \) differs across different consumers, as the price distribution for each consumer is different.

Specifically, first, as the price distribution is drawn i.i.d across time, one can study how each consumer responds to the temporary price shocks she faces. For each consumer \( k \), we can look at how each of her consumption \( x_{i,k,t} \) moves with respect to \( \hat{p}_k = (p_{1,t}, \cdots, p_{N,t}) \), for \( t \in \{1, \cdots, T\} \). This will identify the narrow thinker’s demand elasticity with respect to price shocks studied in main text.

Second, one can first calculate the average demand and the average price for each consumer (across all \( T \) periods), and then study how such each consumer’s average demand varies with her average price. Such method will identify a different demand elasticity under narrow thinking. The one identified will coincide with the frictionless demand elasticity under standard consumer theory.

Observational Equivalence with Other Forms of Bounded Recall.

In Section 3, we show that, in a sequential setting, the narrow thinker’s information structure introduced above can be interpreted as a particular form of bounded recall (or selective retrieval from memory). A natural question is how different the narrow thinker’s behavior analyzed in Section 3 is the different from the behavior of a decision maker, whose bounded recall is captured by a noisy signal about the past endogenous decision. Here, I show that, in terms of cross-price and own-price elasticities, these two approaches of modeling bounded rationality are observationally equivalent.

To illustrate the result in a clean fashion, consider a symmetric version of the environment in Section 3. That is, the optimal consumption for each self \( i \in \{1, 2\} \) in (6) is given by

\[
\hat{x}_i^* (\omega_i) = E_i \left[ -\psi \hat{p}_i + \gamma \hat{x}_{-i}^* (\omega_{-i}) \right], \tag{43}
\]
Each self $i \in \{1, 2\}$ of the narrow thinker, who is in charge of purchasing good $i$, perfectly knows $p_i \sim \log \mathcal{N}(\log \bar{p}, \sigma_i^2)$, but receives a noisy signal about each of the other $p_{-i}$: $s_{i,-i} = p_i \epsilon_{i,-i}$, with $\epsilon_{i,-i} \sim \log \mathcal{N}(0, \sigma^2)$ and $\sigma^2 > 0$. All $e$s and $p$s are independent from each other. In other words, for $i \in \{1, 2\}$, $\omega_i = \{p_i, s_{i,-i}\}$. 

Now, consider a decision maker, whose bounded recall is captured by a noisy signal about the past endogenous decision. For such a decision maker (BR), in the case that decision 1 is made before decision 2, her self 1’s information is given by $\omega_1 = \{p_1, s_{1,2}^{BR} = p_2^{BR}\}$, where $\epsilon_{1,2}^{BR} \sim \log \mathcal{N}(0, \sigma_{BR}^2)$. That is, she perfectly knows $p_1$, and receives a noisy signal about the future $p_2$, and $\sigma_{BR}^2$ captures the size of the noise. On the other hand, her self 2’s information is given by $\omega_2 = \{p_2, s_{2,1}^{BR} = x_1 \epsilon_{2,1}^{BR}\}$, where $\epsilon_{2,1}^{BR} \sim \log \mathcal{N}(0, \Sigma_{BR}^2)$. That is, she perfectly knows $p_2$, but cannot perfectly recall her past decision $x_1$. Instead, she receives a noisy signal directly about $x_1$ and $\Sigma_{BR}^2$ captures the size of the noise. Similarly, for the case that the decision 2 is made before the decision 1, we have $\omega_2 = \{p_2, s_{2,1}^{BR} = p_1 \epsilon_{2,1}^{BR}\}$ and $\omega_1 = \{p_1, s_{1,2}^{BR} = x_2 \epsilon_{1,2}^{BR}\}$, where $\epsilon_{2,1}^{BR} \sim \log \mathcal{N}(0, \Sigma_{BR}^2)$ and $\epsilon_{1,2}^{BR} \sim \log \mathcal{N}(0, \Sigma_{BR}^2)$.

Note that the information and behavior of such a decision maker (BR), whose bounded recall is captured by a noisy signal about the past endogenous decision, will depend on the order of the decisions. To compare her behavior with that of the above narrow thinker, we define $\hat{x}_i^{BR}(\hat{p}_1, \hat{p}_2) \equiv E[\hat{x}_i^*(\omega_i) | \hat{p}_1, \hat{p}_2]$, where $E[\cdot | \hat{p}_1, \hat{p}_2]$ averages over not only the realization of noises in signals but also the potential order of decisions.\textsuperscript{36}

In fact, without additional knowledge about the order of decisions, the own-price and cross-price elasticities of the narrow thinker and those of the decision maker (BR) are observationally equivalent.\textsuperscript{37}

**Lemma 5** For a narrow thinker with any size of noise $\sigma^2 > 0$, one can find a decision maker (BR) with sizes of noise $\sigma_{BR}^2 > 0$ and $\Sigma_{BR}^2 > 0$, such that the two decision makers share the same own-price and cross-price elasticities:

$$\frac{\partial \hat{x}_i^{Narrow}}{\partial \hat{p}_i} = \frac{\partial \hat{x}_i^{BR}}{\partial \hat{p}_i} \quad \text{and} \quad \frac{\partial \hat{x}_i^{Narrow}}{\partial \hat{p}_{-i}} = \frac{\partial \hat{x}_i^{BR}}{\partial \hat{p}_{-i}} \quad \forall i \in \{1, 2\}. \quad (44)$$

This equivalence result should not be surprising: in the end, the noisy signals in both cases capture bounded recall and restrict the later decision to be made based on an imperfect perception of the earlier decision. One may then wonder why I focus on the case with noisy signals about fundamentals throughout the paper, instead of the case with noisy signals directly about decisions. As mentioned above, one advantage of the former case is that the analysis does not require knowledge about the exact order of decisions. Moreover, for the case with noisy signals directly about decisions, it is hard to analytically characterize the decision maker’s behavior when we have $N \geq 3$ decisions. To illustrate the difficulty, consider self $N$, who

\textsuperscript{36}I assume with 50% of probability that decision 1 is made before decision 2, and with another 50% of probability that decision 2 is made before decision 1.

\textsuperscript{37}Such a decision maker (BR) is still a narrow thinker based on the general definition in Definition 1, but has different information from the narrow thinker studied in Section 5 and beyond.
receives noisy signals about the first $N-1$ decisions: $s_{N,i}^{BR} = x_i e_{N,i}^{BR}, i \in \{1, \ldots, N-1\}$. As each decision is a function of all fundamentals ($p_1, \ldots, p_N$), the signal about each decision $i$ will also be informative about all other decisions and vice versa. In other words, the “rational confusion” among signals is prevalent, and the analysis becomes intractable. On the other hand, with noisy signals about (independent) fundamentals, one can still characterize analytically the narrow thinker’s behavior in many interesting economic environments, as studied in Section 4.

In fact, this technical choice echoes that in the literature on interpersonal coordination friction, i.e. incomplete information “beauty contests” (Morris and Shin, 2002; Angeletos and Pavan, 2007). That literature also mainly focuses on information structures with noisy signals about fundamentals (instead of endogenous actions), and uses these signals to model each agent’s imperfect perception of other agents’ decisions and to introduce coordination frictions among agents.

**Proof of Lemma 5.** For notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. Under narrow thinking, from the proof of Proposition 2, we have, for $i \in \{1, 2\}$,

$$
\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} = -\psi \left( 1 + \frac{\lambda \gamma^2}{1 - \lambda \gamma^2} \right) \quad \text{and} \quad \frac{\partial x_i^{\text{Narrow}}}{\partial p_{-i}} = -\psi \frac{\lambda \gamma^2}{1 - \lambda \gamma^2},
$$

where $\lambda = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$. Now, we consider the decision maker BR, whose bounded recall is captured by a noisy signal about the past endogenous decision. First consider the case that the decision 1 is made before the decision 2. We use guess and verify approach, and suppose that two decisions can be characterized by

$$
\begin{align*}
x_1^* (\omega_1) &= \alpha_1 p_1 + \alpha_2 s_{1,2}^{BR} = \alpha_1 p_1 + \alpha_2 (p_2 + e_{1,2}^{BR}), \\
x_2^* (\omega_2) &= \beta_2 p_2 + \beta_1 s_{2,1}^{BR} = \beta_2 p_2 + \beta_1 (x_1 + e_{2,1}^{BR}).
\end{align*}
$$

From self 1’s optimality, we then have

$$
x_1^* (\omega_1) = -\psi p_1 + E_1 [\gamma x_2^* (\omega_2)] = -\psi p_1 + \gamma E_1 [\beta_2 p_2 + \beta_1 x_1].
$$

As a result,

$$
\alpha_1 = -\frac{\psi}{1 - \beta_1 \gamma} \quad \text{and} \quad \alpha_2 = \frac{\lambda_1^{BR} \beta_2}{1 - \beta_1 \gamma},
$$

49
where \( \lambda^1_{BR} = \frac{\sigma^2}{\sigma^2 + \sigma^2_{BR}} \). From self 2’s optimality, we then have

\[
x^*_2(\omega_2) = -\psi p_2 + \gamma E_2 \left[ \alpha_1 p_1 + \alpha_2 \left( p_2 + \epsilon^1_{2,2} \right) \right],
\]

\[
= (-\psi + \gamma \alpha_2) p_2 + \gamma E_2 \left[ \alpha_1 p_1 + \alpha_2 \epsilon^1_{2,2} \right],
\]

\[
= (-\psi + \gamma \alpha_2) p_2 + \lambda^2_{BR} \gamma \left( x_1 + \epsilon^2_{2,1} - \alpha_2 p_2 \right),
\]

where \( \lambda^2_{BR} = \frac{\alpha_1^2 \text{Var}(p_1) + \alpha_2^2 \sigma^2_{BR}}{\alpha_1^2 \text{Var}(p_1) + \alpha_2^2 \sigma^2_{BR} + \Sigma_{BR}} \). As a result,

\[
\beta_1 = \lambda^2_{BR} \gamma \quad \text{and} \quad \beta_2 = -\psi + \gamma \left( 1 - \lambda^2_{BR} \right) \alpha_2.
\]

Together, we have

\[
\alpha_1 = -\frac{\psi}{1 - \lambda^2_{BR} \gamma} \quad \text{and} \quad \alpha_2 = -\frac{\psi \lambda^1_{BR} \gamma}{1 - \gamma^2 \left( \lambda^2_{BR} \left( 1 - \lambda^1_{BR} \right) + \lambda^1_{BR} \right)}
\]

\[
\beta_1 = -\psi \lambda^2_{BR} \gamma \quad \text{and} \quad \beta_2 = -\frac{\psi \left( 1 - \gamma^2 \lambda^2_{BR} \right)}{1 - \gamma^2 \left( \lambda^2_{BR} \left( 1 - \lambda^1_{BR} \right) + \lambda^1_{BR} \right)}.
\]

Similarly, in the case that the decision 2 is made before the decision 1, we have

\[
x^*_2(\omega_2) = \alpha_1 p_2 + \alpha_2 \left( p_1 + \epsilon^2_{2,1} \right),
\]

\[
x^*_1(\omega_1) = \beta_2 p_1 + \beta_1 \left( x_2 + \epsilon^1_{1,2} \right).
\]

Averaging over not only the realization of noises in signals but also the potential order of decisions, we have, for \( i \in \{1, 2\} \),

\[
\frac{\partial x^i_{BR}}{\partial p_i} = \frac{1}{2} \left( \alpha_1 + \beta_2 + \beta_1 \alpha_2 \right) = -\psi \left[ \frac{1 + \gamma}{2} \left( \frac{\lambda^2_{BR} \gamma}{1 - \lambda^2_{BR} \gamma^2} + \frac{\lambda^1_{BR} \gamma}{1 - \gamma^2 \left( \lambda^2_{BR} \left( 1 - \lambda^1_{BR} \right) + \lambda^1_{BR} \right)} \right) \right] \tag{46}
\]

\[
\frac{\partial x^i_{PR}}{\partial p_{-i}} = \frac{1}{2} \left( \alpha_2 + \beta_1 \alpha_1 \right) = -\frac{\psi}{2} \left( \frac{\lambda^2_{BR} \gamma}{1 - \lambda^2_{BR} \gamma^2} + \frac{\lambda^1_{BR} \gamma}{1 - \gamma^2 \left( \lambda^2_{BR} \left( 1 - \lambda^1_{BR} \right) + \lambda^1_{BR} \right)} \right). \tag{47}
\]

Compared to the formula about own- and cross- sensitivities under narrow thinking in (45), to prove Lemma 5, we only need to prove that, for any \( \lambda \in (0, 1) \), we can find \( \sigma^2_{BR}, \Sigma^2_{BR} > 0 \) such that

\[
\frac{\lambda \gamma}{1 - \lambda \gamma^2} = \frac{1}{2} \left( \frac{\lambda^2_{BR} \gamma}{1 - \lambda^2_{BR} \gamma^2} + \frac{\lambda^1_{BR} \gamma}{1 - \gamma^2 \left( \lambda^2_{BR} \left( 1 - \lambda^1_{BR} \right) + \lambda^1_{BR} \right)} \right).
\]

This is true as

\[
\lim_{\sigma^2_{BR} \to 0, \Sigma^2_{BR} \to 0} = \frac{1}{2} \left( \frac{\lambda^2_{BR} \gamma}{1 - \lambda^2_{BR} \gamma^2} + \frac{\lambda^1_{BR} \gamma}{1 - \gamma^2 \left( \lambda^2_{BR} \left( 1 - \lambda^1_{BR} \right) + \lambda^1_{BR} \right)} \right) = \frac{\gamma}{1 - \gamma^2}
\]
and
\[
\lim_{\sigma_{\tilde{n}} \to +\infty, \Sigma_{\tilde{n}} \to +\infty} \frac{1}{2} \left( \frac{\lambda_2^{\text{BR}} \gamma}{1 - \lambda_2^{\text{BR}} \gamma^2} + \frac{\lambda_1^{\text{BR}} \gamma}{1 - \gamma^2 \left( \lambda_2^{\text{BR}} (1 - \lambda_1^{\text{BR}}) + \lambda_1^{\text{BR}} \right)} \right) = 0.
\]

**Comparison with Standard Sequential Decisions.**

Let me now compare the narrow thinker’s behavior with the behavior of a decision maker under perfect recall, who can perfectly recall her earlier fundamentals, signals, and decisions when making later decisions. Consider the same sequential environment as the one studied in Lemma 5. Here, I compare the narrow thinker’s behavior with the behavior of a decision maker with perfect recall. Specifically, for a decision maker with perfect recall (PR), in the case that the decision 1 is made before the decision 2, as above, her self 1’s information is given by \( \omega_1 = \{ p_1, s_{1,2}^{\text{PR}} = p_2 \epsilon_{1,2}^{\text{PR}} \} \), where \( \epsilon_{1,2}^{\text{PR}} \sim \log N(0, \sigma_{\text{PR}}^2) \). That is, she perfectly knows \( p_1 \), and receives a noisy signal about the future \( p_2 \), and \( \sigma_{\text{PR}}^2 \) captures the size of the noise in this signal. On the other hand, her self 2’s information is given by \( \omega_2 = \{ p_1, p_2, s_{1,2}^{\text{PR}}, x_1 \} \). That is, when her self 2 decides on \( x_2 \), she not only perfectly knows her local fundamental \( \theta_2 \), but also perfectly recalls her past fundamentals, signals and decisions. Similarly, for the case that the decision 2 is made before the decision 1, we have \( \omega_2 = \{ p_2, s_{2,1}^{\text{PR}} = p_1 \epsilon_{2,1}^{\text{PR}} \} \) and \( \omega_1 = \{ p_1, p_2, s_{2,1}^{\text{PR}}, x_2 \} \).

We then compare the narrow thinker’s behavior \( \hat{x}_i^{\text{Narrow}}(\hat{p}_1, \hat{p}_2) \) with the behavior of this decision maker with perfect recall (PR), \( \hat{x}_i^{\text{PR}}(\hat{p}_1, \hat{p}_2) \equiv E[\hat{x}_i^*(\omega_i)|\hat{p}_1, \hat{p}_2] \), where \( E[\cdot|\hat{p}_1, \hat{p}_2] \) averages over not only the realization of noises in signals but also the potential order of decisions.

**Lemma 6** There exists a \( \tilde{\sigma} > 0 \) such that for any narrow thinker with the size of noise \( \sigma > \tilde{\sigma} \), her own- and cross- sensitivities deviate more from the frictionless counterparts than the counterparts of any decision maker with perfect recall (PR, with any \( \sigma_{\text{PR}}^2 \)):

\[
\left| \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} \right| \leq \frac{\partial \hat{x}_i^{\text{PR}}}{\partial \hat{p}_i} \leq \left| \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} \right| \quad \text{and} \quad \left| \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} \right| \leq \left| \frac{\partial \hat{x}_i^{\text{PR}}}{\partial \hat{p}_i} \right| \leq \left| \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} \right|.
\]

To understand why the extent of frictional behavior with perfect recall is limited, note that, for the decision that comes first (supposed to be the decision 1), the decision maker PR’s decision is given by \( \hat{x}_i^*(\omega_1) = x_1^{\text{Standard}}(\hat{p}_1, E_1[\hat{p}_2]) \). That is, a form of certainty equivalence emerges: one can use the standard, frictionless, decision function to characterize her behavior. In other words, there is no coordination friction among the two selves, and the only friction comes from the earlier self’s uncertainty about the future fundamental. As a result, the total degree of frictional behavior under perfect recall is limited.

**Proof of Lemma 6.** For notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. First consider the case that the decision 1 is made before the decision 2. As self 2 has perfect recall, we have

\[
x_2^*(\omega_2) = -\psi p_2 + \gamma x_1^*(\omega_1).
\]
From self 1’s perceptive, we then have

\[ x_1^* (\omega_1) = -\psi p_1 + E_1 [\gamma x_2^* (\omega_2)] = -\psi p_1 + E_1 [\psi \gamma p_2 + \gamma^2 x_1^* (\omega_1)] . \]

As a result,

\[ x_1^* (\omega_1) = -\frac{\psi}{1 - \gamma^2} p_1 - \frac{\psi \lambda_{PR}^\gamma}{1 - \gamma^2} (p_2 + \epsilon_{1,2}^{PR}) , \]

\[ x_2^* (\omega_2) = -\psi p_2 \left(1 + \frac{\lambda_{PR}^\gamma}{1 - \gamma^2}\right) - \frac{\psi \gamma}{1 - \gamma^2} p_1 - \frac{\psi \lambda_{PR}^\gamma}{1 - \gamma^2} \epsilon_{1,2}^{PR} , \]

where \( \lambda_{PR} = \frac{\sigma_p^2}{\sigma_p^2 + \epsilon_{PR}^2} \).

Similarly, in the case that the decision 2 is made before the decision 1, we have

\[ x_2^* (\omega_2) = -\frac{\psi}{1 - \gamma^2} p_2 - \frac{\psi \lambda_{PR}^\gamma}{1 - \gamma^2} (p_1 + \epsilon_{2,1}^{PR}) , \]

\[ x_1^* (\omega_1) = -\psi p_1 \left(1 + \frac{\lambda_{PR}^\gamma}{1 - \gamma^2}\right) - \frac{\psi \gamma}{1 - \gamma^2} p_2 - \frac{\psi \lambda_{PR}^\gamma}{1 - \gamma^2} \epsilon_{2,1}^{PR} . \]

Averaging over not only the realization of noises in signals but also the potential order of decisions, we have, for \( i \in \{1, 2\} \),

\[ \frac{\partial x_i^{PR}}{\partial p_i} = -\psi \left(1 + \frac{1}{2} \frac{\gamma^2 (1 + \lambda_{PR})}{1 - \gamma^2}\right) \quad \text{and} \quad \frac{\partial x_i^{PR}}{\partial p_{-i}} = -\psi \left(\frac{1}{2} \frac{\gamma (1 + \lambda_{PR})}{1 - \gamma^2}\right) . \] (48)

Compared to the formula about own- and cross- sensitivities under narrow thinking in (45), Lemma 6 follows with \( \sigma = \sqrt{1 - \gamma^2} \sigma_p \).

**Appendix C: A Smooth Model of Mental Accounting**

**Interpreting Each \( i \) as a Spending Category.**

In the main analysis, I interpret \( x_i \) as the consumption of a good \( i \), which is decided by self \( i \). An alternative interpretation is that each \( x_i \) captures the composite consumption for a spending category, and \( p_i \) the price index of that category:

\[ x_i = \mathcal{V}_i (x_{i,1}, \cdots, x_{i,M_i}) \quad \text{and} \quad p_i = \mathcal{P}_i (p_{i,1}, \cdots, p_{i,M_i}) \equiv \min_{\mathcal{V}_i (x_{i,1}, \cdots, x_{i,M_i}) = \mathcal{V}_i; \sum_{m=1}^{M_i} p_{i,m} x_{i,m},} \]

where \( \mathcal{V}_i \) is homogenous of degree one.

I now let each self \( i \) be in charge of deciding consumption of all goods \((x_{i,1}, \cdots, x_{i,M_i})\) in spending
category \(i\). She perfectly knows the price of all goods \((p_{i,1}, \ldots, p_{i,M})\), and receives the a noisy signal about the price index in other spending category: \(s_{i,k} = p_k \epsilon_{i,k}\), for \(k \neq i\).

In this case, as each self \(i\) can perfectly coordinate consumption of all goods within the spending category \(i\), one can effectively collapse all consumption decisions within spending category \(i\) into the composite consumption decision \(x_i\) based on the price index \(p_i\). The main analysis can then be interpreted as consumption decisions of different spending categories and Proposition 4 can then be re-interpreted as a smooth model of mental accounting at the spending category level.

By the same token, one can interpret Corollary 3 as how the composite consumption of a spending category decreases excessively in response to a common price shock to all goods within that spending category. This is what Hastings and Shapiro (2013) find: after a common price shock to all different types of gasoline, consumers significantly decrease their total gasoline expenditure, to an extent that cannot be explained by neoclassical effects. In the next part of the appendix, I further extend my setting to accommodate the discrete choices among different grades of gasoline as in Hastings and Shapiro (2013). I show how their estimates can help inform about the degree of narrow thinking \(\lambda\) and imply a sizable deviation of gasoline consumption from the frictionless benchmark.

**Mapping to Hastings and Shapiro (2013).**

Here, I consider the environment in Hastings and Shapiro (2013) with discrete choices among different grades of gasoline. I show how their estimates about the violation of the fungibility principle can help inform about the degree of narrow thinking \(\lambda\) and imply a sizable deviation of gasoline consumption from the frictionless benchmark.

**Environment.** The consumer chooses one gasoline grade among different grades of gasoline. Let me use \(o_j \in \{0,1\}\) to denote whether gasoline grade \(j \in \{0,\ldots,J\}\) has been chosen. As only one grade of gasoline is chosen, \(\sum_{j=0}^{J} o_j = 1\). Following Hastings and Shapiro (2013), the quantity of her gasoline consumption is fixed at a constant \(q_g\) (gallons).

The rest of the consumer’s utility is similar to the environment in Section 4.1. Specifically, her utility is given by

\[
\sum_{j=0}^{J} u_j o_j q_g + \sum_{i=1}^{N} v_i (x_i) + h(y),
\]

where \(u_j\) captures the per-gallon utility from consuming grade \(j\) of gasoline, \(v_i (x_i) = \frac{x_i^{1-\xi_i}}{1-\xi_i}\) captures the consumer’s utility from consuming good \(i\), and \(h(y)\), a strictly concave function on \(\mathbb{R}\), captures the consumer’s utility from the last decision, which can be interpreted as utility from saving (borrowing).

The consumer is subject to the budget constraint:

\[
\left( p_g + \sum_{j=1}^{J} \Delta_j o_j \right) q_g + \sum_{i=1}^{N} p_i x_i + y \leq w,
\]
where $p_g$ captures the per-gallon price of grade 0 gasoline, and $\Delta_j$ captures the premium of grade $j \in \{1, \ldots, J\}$ gasoline.

**Information.** Following Hastings and Shapiro (2013)'s identification strategies, I study shocks to gasoline price $p_g$ and income/wealth $w$, and treat other parameters as constants. As in the main text, the analysis is essentially the same if other parameters are independently distributed random variables.

Specifically, for the narrow thinker, her self $g$, who in charge of deciding which gasoline to use, perfectly knows $p_g \sim \mathcal{N}(\bar{p}_g, \sigma_{p_g}^2)$. Her self $i \in \{1, \ldots, N\}$, who is charge of purchasing good $i$, receives a noisy signal about $p_g$: $s_{i,g} = p_g + \epsilon_{i,g}$, with $\epsilon_{i,g} \sim \mathcal{N}(0, \sigma_{i,g}^2)$ and $\sigma_{i,g}^2 > 0$. Similar to the main text, I use $\lambda_{i,g} \equiv \frac{\sigma_{i,g}^2}{\sigma_{p_g}^2 + \sigma_{i,g}^2} \in [0, 1]$ to capture the precision of self $i$’s signal about $p_g$. Moreover, each self $i \in \{g, 1, \ldots, N\}$ receives a noisy signal about income/wealth $w$: $s_{i,w} = w + \epsilon_{i,w}$, with $w \sim \mathcal{N}(\bar{w}, \sigma_w^2)$ and $\epsilon_{i,w} \sim \mathcal{N}(0, (\sigma_{i,w}^w)^2)$. I use $\lambda_{i,w} \equiv \frac{\sigma_{i,w}^w}{\sigma_{i,w}^2 + (\sigma_{i,w}^w)^2} \in [0, 1]$ to capture the precision of self $i$’s signal about $w$. All fundamentals and noises are independent from each other. Note that here fundamentals and noises are distributed Normally instead of log-normally. This is because I will follow Hastings and Shapiro (2013) to use linearization instead of log-linearization. This allows me to directly map my model to their estimates.

**Gasoline choice.** Following Hastings and Shapiro (2013), the choice of which grade of gasoline to use can be summarized by the comparison among the per-gallon utility from consuming each gasoline grade $j \in \{0, \ldots, J\}$:

$$u_j^g = u_j^0 - \mu \Delta_j,$$

where $\mu$ captures the marginal utility of non-gasoline expenditures.

In the narrow thinking model here, as the last decision $y$ will adjust based on the gasoline expenditure, the relevant $\mu$ for gasoline decision is the gasoline self’s belief about $h'(y)$:

$$\mu = E_g [h'(y)] . \tag{49}$$

**Connecting to Hastings and Shapiro (2013)’s estimates.** Hastings and Shapiro (2013) find that, from the perspective of gasoline decisions, the marginal utility of money increases 15 times more in response to an increase in the gasoline price than in response to an equivalent decrease in income. Now, I show how, in the context of the narrow thinking model here, one can interpret their estimates as the degree of narrow thinking.

Here, to connect with their estimates directly, I follow them to use linearization and use a hat over a variable to denote its deviation from the point where each price and the income is fixed at its average. The following Lemma connects parameters in my models to Hastings and Shapiro (2013)’s estimates about how marginal value of money $\mu$ responds to gasoline price and income changes. In the Lemma, similar to main text, for each endogenous variable $x$, I define $x^{Narrow} = E [\hat{x}/\hat{p}_g, \hat{p}_1, \cdots, \hat{p}_N, \hat{w}]$, averaging over the realization of noises in signals.
Lemma 7 Assume all selves’ noisy signals have the same signal to noisy ratio. That is, \( \lambda_{i,g} = \lambda_i^w = \lambda \) for all \( i \).\(^{38}\) There exists scalars \( \eta^{g,Narrow} \) and \( \eta^{w,Narrow} \) such that the marginal value of money \( \mu \) in (49) can be expressed as:

\[
\hat{\mu} = \eta^{g,Narrow} \left( \hat{p}_g + \sum_j \Delta_j \hat{\omega}_j^{Narrow} \right) q^g - \eta^{w,Narrow} \hat{w} ,
\]

where \( \left( \hat{p}_g + \sum_j \Delta_j \hat{\omega}_j^{Narrow} \right) q^g \) is the gasoline expenditure. Moreover,

\[
\eta^{w,Narrow} / \eta^{g,Narrow} = \lambda .
\]

The two objects in (50) are the direct counterparts of the objects that Hastings and Shapiro (2013) estimate in their equation (6). Though with a nontrivial assumption (common \( \lambda \) and the fact that narrow thinking is the only friction), (51) directly connects the degree of narrow thinking \( \lambda \) with Hastings and Shapiro (2013)’s estimates about the differential impact of gasoline expenditures and incomes on the marginal utility of wealth. Based on their estimates, we have \( \lambda \approx 1/15 \).

Narrow thinking’s impact of gasoline choice. Similar to the spirit of the calibration exercise in Corollary 2, let me provide a first look at how narrow thinking can influence gasoline choice.

Corollary 4 Assume common \( \kappa \) and \( \lambda \). That is, for all \( i \), \( \kappa_i = \kappa_y = \kappa_i \)\(^{39}\) and, for all \( i \neq k \in \{ g, 1, \cdots , N \} \), \( \lambda_{i,k} = \lambda_i^w = \lambda \). The following expression provides a lower bound about how narrow thinking increases the sensitivity of marginal value of money to gasoline expenditure compared to the frictionless benchmark:

\[
\frac{\eta^{g,Narrow}}{\eta^{g,Standard}} \geq \frac{\mu_x}{\lambda \mu_y} + 1 ,
\]

where \( \mu_x = \sum_{i=1}^N \mu_i \) captures the total spending share (excluding gasoline).

Similar to the main text, we calibrate \( \frac{\mu_x}{\mu_y} = 5.8 \cdot 0.96 = 5.568 \), where the first term comes from the ratio between the median household’s consumption and net wealth in Baker (2018) and the second term comes from the share of non-gasoline expenditure in total expenditure from NIPA Table 2.4.5. Together with the calibration of \( \lambda \approx 1/15 \) above, we have \( \frac{\eta^{g,Narrow}}{\eta^{g,Standard}} \geq 4.8 \). Note that in Hastings and Shapiro (2013), the probability to switch to regular gasoline after a gasoline price increase is approximately linear to \( \eta^g \).\(^{40}\) This means that a typical household’s propensity to switch to regular gasoline after a gasoline price increase is at least around 4.8 times larger under narrow thinking.

\( ^{38} \)This means, for all \( i \in \{ 1, \cdots , N \} \), \( \lambda_{i,g} = \lambda \). And, for all \( i \in \{ 1, \cdots , N \} \), \( \lambda_{i}^w = \lambda \).

\( ^{39} \)Similar to the main text, \( \kappa_y = \frac{h'(y) \hat{y}}{h(y)} \) captures the rate at which the marginal value of money \( h'(y) \) moves with respect to \( y \).

\( ^{40} \)For example, see their Table 3.
Proof of Lemma 7. Similar to (13), the optimal consumption for each self $i \in \{1, \cdots, N\}$ is given by:

$$\kappa_i \frac{\hat{x}_i^*(\omega_i)}{x_i} = -\tilde{p}_i + \kappa_y E_i \left[ \frac{\hat{y}^*}{y} \right], \quad (52)$$

where $\kappa_y = -\frac{h''(y)}{E(y)}$ and a bar over a variable to denote its value when each price and the income is fixed at its average. The budget constraint can be written as

$$\left( \hat{p}_g + \sum_{j=1}^{J} \Delta_j \hat{o}_j^*(\omega_g) \right) q^g + \sum_{i=1}^{N} (\hat{p}_i \hat{x}_i + \tilde{p}_i \hat{x}_i^*(\omega_i)) + \hat{y}^* = 0, \quad (53)$$

where $\omega_g$ captures the gasoline self’s signal.

First, consider response to shocks to gasoline price $\hat{p}_g$. From each self’s optimality (52) and similar to the proof of Proposition 4, for self $i \in \{1, \cdots, N\}$,

$$\frac{\kappa_i \partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_g} = \frac{\kappa_y \lambda \partial \hat{y}^{\text{Narrow}}}{\hat{y}} - \frac{(1 - \lambda) \kappa_y \hat{p}_i}{\hat{y}} \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_g}. \quad (54)$$

From the budget constraint (53),

$$\left(1 + \sum_{j=1}^{J} \Delta_j \frac{\partial \hat{o}_j^{\text{Narrow}}}{\partial \hat{p}_g} \right) q^g + \sum_{i=1}^{N} \tilde{p}_i \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_g} + \frac{\partial \hat{y}^{\text{Narrow}}}{\partial \hat{p}_g} = 0. \quad (55)$$

Together, we have

$$\frac{\partial \hat{y}^{\text{Narrow}}}{\partial \hat{p}_g} = -\left(1 + \sum_{j=1}^{J} \Delta_j \frac{\partial \hat{o}_j^{\text{Narrow}}}{\partial \hat{p}_g} \right) q^g \left(\sum_{i=1}^{N} \frac{\lambda \kappa_i}{\mu_i + (1 - \lambda) \frac{\kappa_i}{\hat{y}_i}} + 1\right), \quad (56)$$

where $\mu_i = \frac{\tilde{p}_i \hat{x}_i}{\hat{w}}$ is the spending share of good $i$ and $\mu_y = \frac{\hat{y}}{\hat{w}}$.

Now consider response to shocks to $\hat{w}$. From each self’s optimality (52) and similar to the proof of Proposition 4, for self $i \in \{1, \cdots, N\}$,

$$\frac{\kappa_i \partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{w}} = \frac{\kappa_y \lambda \partial \hat{y}^{\text{Narrow}}}{\hat{y}} \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{w}} - \frac{(1 - \lambda) \kappa_y \hat{p}_i}{\hat{y}} \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{w}}. \quad (57)$$

From the budget constraint (53),

$$\left(\sum_{j=1}^{J} \Delta_j \frac{\partial \hat{o}_j^{\text{Narrow}}}{\partial \hat{w}} \right) q^g + \sum_{i=1}^{N} \tilde{p}_i \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{w}} + \frac{\partial \hat{y}^{\text{Narrow}}}{\partial \hat{w}} = 1. \quad (58)$$

56
Together, we have

$$\frac{\partial \hat{y}^{\text{Narrow}}}{\partial \hat{w}} = 1 - \left( \sum_{j=1}^{J} \Delta_j \frac{\partial \hat{y}^{\text{Narrow}}}{\partial \hat{w}} \right) q^g \left( \sum_{i=1}^{N} \frac{\lambda \mu_i}{\kappa_y + (1-\lambda) \mu_i} + 1 \right), \tag{55}$$

From (54) and (55), we have

$$\hat{y}^{\text{Narrow}} = \frac{\left( \hat{p}_g + \sum_{j=1}^{J} \Delta_j \hat{\phi}_j \right) q^g - \hat{w}}{\sum_{i=1}^{N} \frac{\lambda \mu_i}{\kappa_y + (1-\lambda) \mu_i} + 1} \tag{56}$$

From (49), we know $\mu^* = -\kappa_y E_g \left[ \frac{\hat{y}^*}{\hat{y}} \right]$. Moreover, we know that

$$E_g \left[ \hat{p}_g + \sum_{j=1}^{J} \Delta_j \hat{\phi}_j (\omega_g) \right] = \left[ \hat{p}_g + \sum_{j=1}^{J} \Delta_j \hat{\phi}_j (\omega_g) \right],$$

and $E_g [\hat{w}] = \lambda (\hat{w} + \varepsilon^w)$. Using (56) and averaging over the realizations of noises, we have

$$\hat{\mu} = \eta^{g, \text{Narrow}} \left( \hat{p}_g + \sum_{j=1}^{J} \Delta_j \hat{\phi}_j^{\text{Narrow}} \right) q^g - \eta^{w, \text{Narrow}} \hat{w},$$

where

$$\eta^{g, \text{Narrow}} = \frac{1}{\sum_{i=1}^{N} \frac{\lambda \mu_i}{\kappa_y + (1-\lambda) \mu_i} + 1} \quad \text{and} \quad \eta^{w, \text{Narrow}} = \frac{\lambda}{\sum_{i=1}^{N} \frac{\lambda \mu_i}{\kappa_y + (1-\lambda) \mu_i} + 1}.$$ 

So $\eta^{w, \text{Narrow}} / \eta^{g, \text{Narrow}} = \lambda$.

**Proof of Corollary 4.** Using the formula of $\eta^g$ in the proof of Lemma 7, we have

$$\frac{\eta^{g, \text{Narrow}}}{\eta^{g, \text{Standard}}} = \frac{\sum_{i=1}^{N} \frac{\mu_i}{\mu_y} + 1}{\sum_{i=1}^{N} \frac{\lambda \mu_i}{\kappa_y + (1-\lambda) \mu_i} + 1} = \frac{\frac{\mu_i}{\mu_y} + 1}{\frac{\lambda \mu_i}{\mu_y} + 1}.$$  

**The Small Wage Elasticity of Daily Labor Supply.**

In the standard labor supply theory, when the wage on a particular day increases, the decision maker will coordinate her behavior by increasing her labor supply on the day of wage increase and decreasing her labor supply on other days. Such a coordinated response generates a large elasticity of daily labor supply. Under narrow thinking, however, labor supply on other days may not be as responsive, and such friction will prevent a large increase in labor supply on the day of wage increase.
Environment. To formalize, consider a decision maker whose utility is

$$\sum_{i=1}^{N} -v (l_i) + h (y)$$

where $l_i$ is the labor supply on day $i$, $v (l_i) = \frac{1}{1+n} \log (l_i)$ captures the disutility of labor on day $i$, and $\kappa > 0$. $h (y)$ captures her consumption utility, and is a strictly concave function on $\mathbb{R}$. The decision maker is subject to the budget constraint: $\sum w_i l_i + w \leq y$, where $w$ is her initial wealth level (constant) and $w_i$ is her wage on day $i$. As the “residual decision” $y$ is allowed to be negative, the budget constrain will always be satisfied.

Information. In this environment, each self $i \in \{1, \ldots, N\}$ should be interpreted as in charge of labor supply decisions for a day. Each self $i$ of the narrow thinker perfectly knows the wage she faces $w_i$, and receives a noisy signal about each of the other selves’ $w_j$. Specifically, for $i \in \{1, \ldots, N\}$, self $i$’s information (signals) is given by $\omega_i = \{s_{i,j} \}_{j \in \{1, \ldots, N\}}$, where $s_{i,i} = w_i \sim \log N (\log \bar{w}_i, \sigma^2_{w_i})$ and, for $i \neq j$, $s_{i,j} = w_j \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \log N \left(0, \sigma^2_{i,j}\right)$ and $\sigma^2_{i,j} > 0$. $\epsilon$s are independent from each other and all $w$s.

Narrow thinker’s behavior. Similar to the main text, I use a hat over a variable to denote its log-deviation from the point of log-linearization.\(^{41}\) The optimal labor supply condition for each $i$ and the budget constraint are given by

$$\kappa \hat{l}_i^* (\omega_i) = \bar{w}_i - \kappa_y E_i [\bar{y}^*],$$

$$\sum_{i=1}^{N} \mu_i \left(\hat{l}_i^* (\omega_i) + \bar{w}_i\right) = \bar{y}^*, \tag{58}$$

where $\kappa_y = -\frac{h'' (\bar{y}) \bar{y}}{h' (\bar{y})}$ and $\mu_i = \bar{w}_i \bar{l}_i / \bar{y}$ is the share of day $i$ income in total wealth at the point of log-linearization.

Small wage elasticity of daily labor supply. I then study how the narrow thinker’s labor supply on each day $i$ responds to shocks to the wage on that day. Similar to condition (7), for each $i$, I define the narrow thinker’s (log) labor supply function as $\hat{l}_i^{\text{Narrow}} (\bar{w}_1, \ldots, \bar{w}_N) \equiv E \left[\hat{l}_i^* (\omega_i) \mid \bar{w}_1, \ldots, \bar{w}_N\right]$. Compared to the standard frictionless case when each decision is made with perfect knowledge of all fundamentals (indexed by the superscript Standard, as above), one can then establish excess sensitivity to temporary income shocks under narrow thinking.

**Proposition 10** For each $i$, the narrow thinker’s labor supply $l_i$ is smaller (larger) in response to positive (negative) shocks to $w_i$:

$$\frac{\partial \hat{l}_i^{\text{Narrow}}}{\partial \bar{w}_i} = \omega_i \frac{\partial \hat{l}_i^{\text{Target}}}{\partial \bar{w}_i} + (1 - \omega_i) \frac{\partial \hat{l}_i^{\text{Standard}}}{\partial \bar{w}_i} \leq \frac{\partial \hat{l}_i^{\text{Standard}}}{\partial \bar{w}_i}, \tag{59}$$

where $\omega_i \in [0, 1]$ and $\hat{l}_i^{\text{Target}}$ captures a daily income target model. That is, there exits $m_i$ such that $w_i l_i = m_i$.

\(^{41}\) I log-linearize around the point where each wage is fixed at $\bar{w}_i$ and each decision is made with perfect knowledge of all wages.
Proposition 10 expresses the narrow thinker’s wage elasticity of daily labor supply as a weighted average between that in the standard labor supply theory ($l_{\text{Standard}}^i$) and that with a daily income target ($x_{\text{Target}}^i$). This leads to a smaller wage elasticity of daily labor supply under narrow thinking. To see the mechanism behind the small wage elasticity of daily labor supply, note that an increase in $w_i$ increases $l_i$ (a positive direct effect) and decreases $l_j$ for $j \neq i$ (both in standard consumer theory and under narrow thinking). This is because the income effect of $w_i$ on $l_j$ (negative) and the substitution effect of $w_i$ on $l_j$ (negative) work in the same direction. The decrease of other $l_j$ then further increases $l_i$ (a positive indirect effect). Under narrow thinking, in response to an increase in $w_i$, the decision maker decreases labor supply on other days less. The indirect effect from this coordinated response is dampened, and the narrow thinker’s $l_i$ is smaller in response to the increase in $w_i$.

**Economic implications.** First, as Farber (2015) points out, daily income target model predicts negative wage elasticity of daily labor supply, which is inconsistent with the empirical evidence. By providing a smooth version of the daily income target model in (59), my approach can then explain the empirically documented positive, but small, wage elasticity of daily labor supply.

Second, in line with Proposition 3, the smaller wage elasticity of labor supply under narrow thinking is about response to temporary daily wage shocks. The narrow thinker’s labor supply decision, average across days, as a function of the average wage can coincide with that in the standard benchmark. Such prediction is consistent with the larger wage elasticity of labor supply found in Fehr and Goette (2007) and Angrist, Caldwell and Hall (2017) based on wage variations at longer frequency.

**Proof of Proposition 10.** Similar to the proof of Proposition 8, from (57) and (58), we have, for all $i$ and $k \neq i$,

$$\kappa \frac{\partial \hat{l}_{\text{Narrow}}^i}{\partial \hat{w}_i} = 1 - \kappa y \left( \sum_{j=1}^{N} \mu_j \frac{\partial \hat{l}_{\text{Narrow}}^j}{\partial \hat{w}_i} + \mu_i \right),$$

$$\kappa \frac{\partial \hat{l}_{\text{Narrow}}^k}{\partial \hat{w}_i} = -\kappa y \left[ \lambda_{k,i} \left( \sum_{j \neq k} \mu_j \frac{\partial \hat{l}_{\text{Narrow}}^j}{\partial \hat{w}_i} + \mu_k \right) - \mu_k \frac{\partial \hat{l}_{\text{Narrow}}^k}{\partial \hat{w}_i} \right],$$

where $\lambda_{k,i} = \frac{\sigma_{u_i}^2}{\sigma_{u_i}^2 + \sigma_{k,i}^2}$. Solving the above two equations, we have for all $i$,

$$\frac{\partial \hat{l}_{\text{Narrow}}^i}{\partial \hat{w}_i} = \frac{1}{\kappa} - \frac{\mu_i \frac{\kappa+1}{\kappa}}{\kappa \left( \frac{1}{\kappa y} + \sum_{j \neq i} \frac{\mu_j \frac{\kappa}{\kappa y} \frac{1}{1-\lambda_{j,i}}}{1-\lambda_{j,i}} \right)}.$$

(60)

We then have for $i \in \{1, \cdots, N\}$,

$$\frac{\partial \hat{l}_{\text{Narrow}}^i}{\partial \hat{w}_i} = \omega_i \frac{\partial \hat{l}_{\text{Target}}^i}{\partial \hat{w}_i} + (1 - \omega_i) \frac{\partial \hat{l}_{\text{Standard}}^i}{\partial \hat{w}_i},$$
where
\[
\omega_i = 1 - \frac{1}{\kappa_y} + \sum_{j \neq i} \frac{\mu_{ij} \lambda_{ij}}{\kappa_y + (1 - \lambda_{ij}) \frac{\mu_{ij}}{\kappa_y}} - \frac{1}{\kappa_y} + \sum_{j \neq i} \frac{\mu_{ij} \lambda_{ij}}{\kappa_y + (1 - \lambda_{ij}) \frac{\mu_{ij}}{\kappa_y}} \in [0, 1],
\]
\[
\frac{\partial \text{Standard}_{\omega_i}}{\partial \omega_i} = \frac{1}{\kappa} - \left( \frac{\mu_i + \mu_j}{\kappa + \frac{\mu_i + \mu_j}{\kappa + \sum_{j \neq i} \mu_{ij}}} \right) > 0, \text{ and } \frac{\partial \text{Target}_{\omega_i}}{\partial \omega_i} = -1 \text{ from the daily income target.}
\]
As \( \frac{\partial \text{Standard}_{\omega_i}}{\partial \omega_i} > \frac{\partial \text{Target}_{\omega_i}}{\partial \omega_i} = -1 \), we have \( \frac{\partial \text{Standard}_{\omega_i}}{\partial \omega_i} > \frac{\partial \text{Narrow}_{\omega_i}}{\partial \omega_i} \).

### Appendix D: General Properties

#### Dampening of Indirect Effects.

Consider the optimal decision rules in (24). As defined there, each decision \( i \)'s response to shocks to fundamentals can be decomposed into a direct and an indirect effect. Here, I formalize the dampening of indirect effects under narrow thinking.

One needs to deal with an additional complication. There could be some components of the indirect effect that positively influence the optimal decision \( i \) and there could be some components of the indirect effect that negatively influence the optimal decision \( i \). Dampening of each component may not mean dampening of the net total. Nevertheless, as the above logic suggests, I can further decompose the indirect effect into positive and negative components. I can then show each component is dampened under narrow thinking.

I first impose conditions such that the game among multiple selves are solvable by iterating best response.

**Assumption 2** The absolute value of all eigenvalues of \( \Gamma \) are less than one.

Iterating the optimal decisions rule in condition (24), we have

\[
\hat{x}_i^*(\omega_i) = \sum_{k=1}^{M} \psi_{i,k} E_i \left[ \hat{\theta}_k \right] + \sum_{j \neq i} \gamma_{i,j} E_i \left[ \hat{x}_j^* \right] \\
= \sum_{k=1}^{M} \psi_{i,k} E_i \left[ \hat{\theta}_k \right] + \sum_{j \neq i} \gamma_{i,j} \left( \sum_{k=1}^{M} \psi_{j,k} E_i \left[ E_j \left[ \hat{\theta}_k \right] \right] \right) + \sum_{j \neq i} \gamma_{i,j} \left( \sum_{l \neq j} \gamma_{j,l} E_i \left[ E_j \left[ \hat{\theta}_l \right] \right] \right) \\
= \sum_{k=1}^{M} \psi_{i,k} E_i \left[ \hat{\theta}_k \right] + \sum_{j \neq i} \gamma_{i,j} \left( \sum_{k=1}^{M} \psi_{j,k} E_i \left[ \hat{\theta}_k \right] \right) + \sum_{j \neq i} \gamma_{i,j} \left( \sum_{l \neq j} \gamma_{j,l} \left( \sum_{k=1}^{M} \psi_{l,k} E_i \left[ E_j \left[ \hat{\theta}_l \right] \right] \right) \right) + \cdots 
\]

(61)

The above representation shows that, as the indirect effect for each decision \( i \) comes from self \( i \)'s belief about other decisions, it in turn depends on self \( i \)'s belief about other selves’ belief about \( \theta_k \), self \( i \)'s belief
about other selves’ beliefs about other selves’ belief about \( \theta_s \), ad infinitum. I can then define, \( x_{i}^{\text{Ind},+} (\omega_i) \), the indirect effect that positively influences \( x_i \), by collecting all belief terms with positive coefficients. I can also define, \( x_{i}^{\text{Ind},-} (\omega_i) \), the indirect effect that negatively influences \( x_i \), as the collection of all belief terms with negative coefficients. Similar to condition (25), averaging over the realization of noises in signals, one can then define each part of the indirect effects as a function of fundamentals: 

\[
\frac{\partial \hat{x}_{i}^{\text{Ind},+,\text{Narrow}}}{\partial \hat{\theta}_k} \quad \text{and} \quad \frac{\partial \hat{x}_{i}^{\text{Ind},-,\text{Narrow}}}{\partial \hat{\theta}_k} \quad \forall i, k
\]

where, as above, a superscript Standard denotes the case when each self perfectly knows all the fundamentals.

**Proposition 11** Under Assumption 2, for each decision \( x_i \), each part of its indirect effect is dampened under narrow thinking in response to shocks to each \( \theta_k \),

\[
\frac{\partial \hat{x}_{i}^{\text{Ind},+,\text{Narrow}}}{\partial \hat{\theta}_k} \leq \frac{\partial \hat{x}_{i}^{\text{Ind},+,\text{Standard}}}{\partial \hat{\theta}_k} \quad \text{and} \quad \frac{\partial \hat{x}_{i}^{\text{Ind},-,\text{Narrow}}}{\partial \hat{\theta}_k} \leq \frac{\partial \hat{x}_{i}^{\text{Ind},-,\text{Standard}}}{\partial \hat{\theta}_k} \quad \forall i, k
\]

Proof of Proposition 11. For all \( i_1, \cdots, i_m \in \{1, \cdots, N\} \), where \( i_l \neq i_{l+1} \) for \( 1 \leq l \leq m - 1 \), we have

\[
E_{i_1} \left[ E_{i_2} \left[ \cdots E_{i_m} \left[ \hat{\theta}_k \right] \right] \right] = \lambda_{i_1, k} \cdots \lambda_{i_m, k} \hat{\theta}_{i_m, k}, \quad (62)
\]

and

\[
E \left[ E_{i_1} \left[ E_{i_2} \left[ \cdots E_{i_m} \left[ \hat{\theta}_k \right] \right] \right] \right] = \lambda_{i_1, k} \cdots \lambda_{i_m, k} \hat{\theta}_k. \quad (63)
\]

From conditions (61) and (63), we have

\[
\frac{\partial \hat{x}_{i}^{\text{Narrow}}}{\partial \theta_k} = \lambda_{i, k} \psi_{i, k} + \sum_{j \neq i} \lambda_{i, k} \gamma_{i, j} \psi_{j, k} + \sum_{j \neq i} \lambda_{i, k} \gamma_{i, j} \sum_{l \neq j} \lambda_{j, k} \gamma_{j, l} \psi_{l, k} + \cdots
\]

Using the fact that each \( \lambda \) is a factor between 0 and 1 and collecting terms with positive and negative coefficients prove Proposition 11.

**Cross-Elasticities**

Consider the environment in Section 5.2. I now turn to how \( x_i \) responds to shocks to other fundamentals \( \theta_k \), for \( i \neq k \).

**Proposition 12** If either 1) or 2) in Assumption 1 holds, the narrow thinker’s cross-elasticities are attenuated:

\[
\left| \frac{\partial \hat{x}_{i}^{\text{Narrow}}}{\partial \theta_k} \right| \leq \left| \frac{\partial \hat{x}_{i}^{\text{Standard}}}{\partial \theta_k} \right| \quad \forall i \neq k.
\]

61
Similar to Proposition 9, the attenuation of cross-elasticity is not always true because of the coexistence of opposing effects. Dampening of each component may not mean dampening of the net total. The additional conditions guarantee that the net total of cross-elasticity is dampened.

Proof of Proposition 12. For notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. We prove the Proposition 12 case by case. In the case (1) “Symmetry,” there exists \( \psi, \Psi > 0, \lambda \in (0, 1) \), and \( \gamma \in \left(-1, \frac{1}{N - 1}\right) \), such that \( \psi_{i,i} = \psi, \psi_{i,k} = \Psi, \gamma_{i,j} = \gamma \), and \( \lambda_{i,j} = \lambda \) for all \( j, k \neq i \).

From (26), we can express the cross-sensitivity as

\[
\frac{\partial \tilde{F}_{i}^{\text{Narrow}}}{\partial \theta_k} = \frac{\lambda \Psi + \lambda \gamma \psi}{1 - \lambda \gamma^2 (N - 1) - \lambda \gamma (N - 2)} \quad \forall i.
\]

Using the fact that \( \lambda \in [0, 1], \psi, \Psi > 0 \) and \( \gamma \in \left(-1, \frac{1}{N - 1}\right) \), Proposition 12 follows directly.

In the case (2) “Complements,” we have \( \gamma_{i,j} \geq 0 \) for all \( i \neq j \) and \( \sum_{j \neq i} \gamma_{i,j} < 1 \) for all \( i \). In this case, the game among multiple selves are solvable by iterating best response. From (26), we have

\[
\frac{\partial \tilde{F}_{i}^{\text{Narrow}}}{\partial \theta_k} = \lambda_{i,k} \psi_{i,k} + \sum_{j \neq i} \lambda_{i,k} \gamma_{i,j} \psi_{j,k} + \sum_{j \neq i} \lambda_{i,k} \gamma_{i,j} \sum_{l \neq j} \lambda_{j,k} \gamma_{j,l} \psi_{l,k} + \cdots.
\]

As each term in the above expression is non-negative, the result follows directly.

A General Consumer Theory Problem

Here, I show how a general consumer theory problem can be nested into the general environment studied in Sections 2 and 5.

The decision maker’s utility depends on her consumption of \( N \) goods, \((x_1, \cdots, x_N)\), and the numeraire \( y \in \mathbb{R} \) (which can be interpreted as saving/borrowing or money). Her utility is given by

\[
\tilde{u}(x_1, \cdots, x_N, y),
\]

where \( \tilde{u} \) is strictly increasing in each of her arguments, strictly concave and twice continuously differentiable. She is subject to the budget constraint

\[
\sum_{i=1}^{N} p_i x_i + y \leq w,
\]

where \( p_i \) is good \( i \)'s price and \( w \) is the decision maker’s total wealth (treated as a constant, as I am interested in response to price shocks here).\(^{43}\)

\(^{42}\)This means \( 1 - \lambda \gamma^2 (N - 1) - \lambda \gamma (N - 2) > 0 \).

\(^{43}\)For notation simplicity, I normalize the price of the last good \( y \) is normalized to 1. In fact, as long as its price
I let \( \tilde{u} \) be well defined for all \( y \in \mathbb{R} \). Based on the discussion about constraint problems in Section 2, this allows the “residual decision” \( y \) to be negative and guarantees that the budget constraint, \( \sum_{i=1}^{N} p_i x_i + y \leq w \), will always be satisfied. As the budget constraint always binds in the optimum, one can use it to substitute \( y \):  

\[
  u(x_1, \cdots, x_N, \tilde{p}) = \tilde{u} \left( x_1, \cdots, x_N, w - \sum_{i=1}^{N} p_i x_i \right). 
\]

This is then nested in the unconstrained problem in (1), with \( \tilde{\theta} = \tilde{p} \).

**Appendix E: Additional Applications**

**Neglect of “Adding-up” Effects.**

One behavior often connected to narrow bracketing is the neglect of “adding-up” effects (Read, Loewenstein and Rabin, 1999). Consider the decision to smoke. The health consequence of a cigarette is small, but the cumulative health consequence of smoking can be large (the adding-up effects). Moreover, the cumulative benefit from smoking seems to increase much more slowly than the cumulative costs. If the decision maker can perfectly coordinate all her smoking decisions, she will not smoke much. However, in practice, the decision maker decides on how much to smoke on different occasions separately, and may face difficulties in coordinating her smoking decisions.

Specifically, the decision maker’s utility is given by

\[
  \sum_{i=1}^{N} \varphi_i v(x_i) - c \left( \sum_{i=1}^{N} x_i \right),
\]

where \( x_i \) captures how much she smokes on occasion \( i \), \( \varphi_i v(x_i) = \varphi_i x_i^{1-\kappa} \) with \( \kappa > 0 \) captures her utility from doing so, and \( \varphi_i \) parametrizes the attractiveness to smoke on occasion \( i \). On the other hand, \( c \left( \sum_{i=1}^{N} x_i \right) = \frac{\left( \sum_{i=1}^{N} x_i \right)^{1+\kappa \varphi}}{1+\kappa \varphi} \) captures the convex cost based on total smoking.

Here I study the impact of a common shock to the attractiveness of smoking on all occasions. For tractability, I let the stochastic property of shocks and the information structure be symmetric across each \( i \). Specifically, the attractiveness to smoke on occasion \( i \) \( \varphi_i \) has an idiosyncratic and a common component: \( \varphi_i = \varphi \delta_i \), where \( \varphi \sim \log \mathcal{N} (\log \varphi, \sigma^2_{\varphi}) \), \( \delta_i \sim \log \mathcal{N} (0, \sigma^2_{\delta}) \) and they are independent from each other. Similar to the information structure considered throughout, each self \( i \) perfectly knows her own \( \varphi_i \), and receives a noisy signal about each of the other \( \varphi_j : s_{i,j} = \varphi_j \epsilon_{i,j} \ \forall j \neq i \). Noise \( \epsilon_{i,j} \sim \log \mathcal{N} (0, \sigma^2_{\epsilon}) \) is independent from the fundamentals and each other.

Similar to the main analysis, I define the narrow thinker’s (log) decision as a function of the fundamentals as \( \hat{x}^{\text{Narrow}}_i (\hat{\varphi}_1, \cdots, \hat{\varphi}_N) \equiv E [ \hat{x}_i^*(\omega_i) | \hat{\varphi}_1, \cdots, \hat{\varphi}_N ] \). I then study \( \frac{d \hat{x}^{\text{Narrow}}_i}{d \varphi} = \lim_{\varphi \to 0} \frac{\hat{x}^{\text{Narrow}}_i (\hat{\varphi}_1, \cdots, \hat{\varphi}_N) - \hat{x}^{\text{Narrow}}_i (0, \cdots, 0)}{\hat{\varphi}} \), is common knowledge across different selves, this normalization is without loss of generality.
which summarizes each decision \( i \)'s response to the common shock.

**Proposition 13** For each \( i \), the narrow thinker increases (decreases) her smoking on occasion \( i \) more in response to positive (negative) common taste shocks \( \varphi \):

\[
\frac{d x^\mathrm{Narrow}_i}{d \varphi} > \frac{d x^\mathrm{Standard}_i}{d \varphi} > 0 \quad \forall i.
\]

To understand the intuition behind the result, note that the common increase in \( \varphi \) will have positive direct effects on all \( x_i \) through the increase in each \( \varphi_i \). As each self \( i \) perfectly knows her own \( \varphi_i \), such direct effects are maintained under narrow thinking. Nevertheless, as self \( i \) does not perfectly know other \( \varphi_j \)'s, she has imperfect perception about how other \( x_j \)'s will respond to the common shock. The indirect effect of the common shock on \( x_i \) through other \( x_j \)'s will then be dampened under narrow thinking. In this context, the indirect effect of the common shock through the increase in other \( x_j \)'s negatively influences \( x_i \), as the cost function is convex. As a result, the direct effect and the indirect effect of the common shock work in opposite directions, and narrow thinking leads to over-reaction. Intuitively, when smoking becomes more attractive, the narrow thinker under-estimates how other selves will increase smoking, neglects the adding-up costs, and smokes excessively.

More generally, in response to common shocks to the fundamental, if different selves’ decisions are strategic substitutes, narrow thinking leads to overreaction relative to the frictionless benchmark. This case arises when the decision maker faces convex add-up costs (e.g. smoking) or concave add-up benefits (e.g. demand for variety in choices). On the other hand, if different selves’ decisions are strategic complements, narrow thinking leads to under reaction relative to the frictionless benchmark. This case arise when the decision maker faces convex add-up benefits (e.g. skill acquisition) and concave add-up costs (e.g. habituation). See the proof of Proposition 13 for details.\(^{44}\)

**Proof of Proposition 13.** In the proof, for notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. Given the environment, the optimal decision rule for each \( i \) is

\[
x^*_i (\omega_i) = E_i \left[ \psi \varphi_i - \gamma \sum_{j \neq i} \frac{x^*_j (\omega_j)}{N} \right], \tag{66}
\]

where \( \psi = \frac{1}{\kappa + \frac{1}{N}} > 0 \) and \( \gamma = \frac{a}{\kappa + \frac{1}{N}} \in (0, 1) \).

Given the information structure, we have

\[
E_i [\varphi_i] = \varphi_i \quad \forall i, \tag{67}
\]

\(^{44}\)This relationship between strategic complementarity/substitutability and under-/over-reaction under narrow thinking only holds in response to a common shock. If the shock is idiosyncratic, as the case in Sections 3 - 5, we should reply on 9 to predict whether narrow thinking leads to over-reaction or under-reaction in a given environment.
Collecting terms, we have

\[ E_i [\varphi] = \frac{\sigma_\delta^{-2}}{\sigma_\varphi^{-2} + \sigma_\varphi^{-2} + (N - 1) (\sigma_\delta^2 + \sigma_\varepsilon^2)^{-1}} \varphi_i + \sum_{l \neq i} \frac{(\sigma_\delta^2 + \sigma_\varepsilon^2)^{-1}}{\sigma_\varphi^{-2} + \sigma_\varphi^{-2} + (N - 1) (\sigma_\delta^2 + \sigma_\varepsilon^2)^{-1}} s_{l,i} \quad \forall i, \]

\[ E_i [\varphi_j] = E_i [\varphi] + \frac{\sigma_\delta^{-2}}{\sigma_\varphi^{-2} + \sigma_\varphi^{-2} + (N - 1) (\sigma_\delta^2 + \sigma_\varepsilon^2)^{-1}} (s_{i,j} - E_i [\varphi]) \]

(68)

\[ \equiv \lambda s_{i,j} + \mu \varphi_i + \omega \sum_{l \neq i, j} s_{i,l} \quad \forall i \neq j, \]

where \( \lambda, \mu, \omega \in (0, 1) \) and \( \lambda + \mu + \omega (N - 2) < 1 \).

Similar to the proof of Proposition 8, as the optimal decision rule (66) is linear and all variables are distributed Normally, for all \( i \), \( x_i^*(\omega_i) \) is linear in its signal and \( x_i^L(\varphi_1, \ldots, \varphi_N) \) is linear in all \( \varphi \). From condition (66) and the fact that the noise in each self’s private signal is not predictable, we have

\[ x_i^*(\omega_i) = \psi \varphi_i - \gamma \sum_{j \neq i} x_j^{\text{Narrow}} (E_i [\varphi_1], \ldots, E_i [\varphi_N]), \]

where \( \psi = \frac{1}{\kappa + \frac{\omega}{\kappa}} > 0 \) and \( \gamma = \frac{\kappa}{\kappa + \frac{\omega}{\kappa}} \in (0, 1) \).

Using (67) and (68), averaging across noises in the realizations of signals, and taking partial derivatives with respect to each \( \theta_j \), we have

\[ \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} = \psi - \gamma \sum_{j \neq i} \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_i} - \mu \gamma \sum_{j \neq i} \sum_{l \neq i} \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_l} \quad \forall i, \]

\[ \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} = -\lambda \gamma \sum_{j \neq i} \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_k} - \omega \gamma \sum_{j \neq i} \sum_{l \neq k, i} \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_l} \quad \forall i, k. \]

Using symmetry, we know \( \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} \) are equal for each \( i \) and \( \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} \) are equal for each \( i \neq k \), we then have,

\[ \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} = \psi - \gamma (N - 1) \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} - \mu \gamma \left( (N - 1) \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} + (N - 1) (N - 2) \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} \right), \]

\[ \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} = -\lambda \gamma \left( (N - 2) \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} + \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} \right) - \omega \gamma \left( (N - 2) \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} + (N - 2)^2 \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} \right). \]

Collecting terms, we have

\[ \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} = \frac{\psi}{1 + \mu \gamma (N - 1) - \frac{\gamma^2 (N - 1) (1 + \mu (N - 2)) (\lambda + \omega (N - 2))}{(1 + \lambda \gamma (N - 2) + \omega \gamma (N - 2)^2)}}, \]

\[ \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} = -\frac{\gamma (\lambda + \omega (N - 2))}{(1 + \gamma (N - 2) (\lambda + \omega (N - 2)))} \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i}. \]
Based on the definition of \( \frac{dx_i^{\text{Narrow}}}{d\varphi} \), we then have

\[
\frac{dx_i^{\text{Narrow}}}{d\varphi} = \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} + (N-1) \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} \frac{\psi}{1 + \mu \gamma (N-1)} - \frac{\gamma (\lambda + \omega (N-2)) (N-1)}{(1 + \gamma (N-2) (\lambda + \omega (N-2)))} (1 + \gamma (N-2) (\lambda + \omega (N-2)))
\]

Using \( \lambda + \mu + \omega (N-2) < 1 \) and letting \( t = \lambda + \omega (N-2) \in (0,1) \), we then have

\[
\frac{dx_i^{\text{Narrow}}}{d\varphi} \geq \frac{\psi (1 - \gamma t)}{1 + \gamma (N-1) (1-t) + \gamma (N-2 - \gamma (N-1)) t} = \frac{\psi}{1 + \gamma (N-1)} = \frac{dx_i^{S}}{d\varphi}.
\]

**Comment.** In the proof, what drives the over-reaction is that different decisions are strategic substitutes (from 66, this means \( \gamma > 0 \)). If different decisions are strategic complements (\( \gamma < 0 \)), we have under-reaction.

\[
\frac{dx_i^{\text{Narrow}}}{d\varphi} \leq \frac{\psi (1 - \gamma t)}{1 + \gamma (N-1) (1-t) + \gamma (N-2 - \gamma (N-1)) t} = \frac{\psi}{1 + \gamma (N-1)} = \frac{dx_i^{S}}{d\varphi}
\]

**Myopic Loss Aversion.**

Another behavior often connected to narrow bracketing is the decision maker’s aversion to combine small and favorable gamble (Samuelson, 1963). The existing explanation of this behavior, such as Benartzi and Thaler (1995), Barberis, Huang and Thaler (2006), and Rabin and Weizsacker (2009), contains two elements: first, the the decision maker suffers from loss aversion; second, she decides on each gamble in isolation. Narrow thinking provides a formalization of such myopic loss aversion, without directly requiring the decision maker to decide on each gamble in isolation.

Specifically, I consider a decision maker who faces two gambles \( i \in \{1,2\} \) and has loss aversion. Her utility is given by

\[
v(c) = \begin{cases} 
  c & \text{if } c > 0 \\
  \delta c & \text{if } c < 0 
\end{cases},
\]

where \( \delta > 1 \) captures the degree of loss aversion, \( c = x_1 r_1 + x_2 r_2 \), \( r_i \) is the return on gamble \( i \in \{1,2\} \), and \( x_i = \{0,1\} \) captures whether the decision maker takes gamble \( i \in \{1,2\} \) or not.

For each gamble \( i \in \{1,2\} \), there is a 50% chance that it turns out to be a loss of \( r_i = -1 \) dollars, and another 50% chance that it turns out to be a gain of \( r_i = 1 + \mu_i \) dollars, where \( \mu_i = \mu e_i > 0 \) is a random
variable, with \( \mu \sim \log \mathcal{N} \left( \log (\mu^{avg}) - \frac{1}{2}\sigma^2, \sigma^2 \right) \) and \( \epsilon_i \sim \log \mathcal{N} \left( -\frac{1}{2}\sigma_e^2, \sigma_e^2 \right) \) for \( i \in \{1, 2\} \). \( \mu, \epsilon_1, \) and \( \epsilon_2 \) are independent from each other. We then have \( E[e^{\mu_i}] = \mu > 0 \) for each \( i \in \{1, 2\} \). Moreover, whether each gamble turns out to be a gain or a loss is independent from that of the other gamble and the returns of gambles.

First consider a frictionless decision maker (indexed by “Standard”) who makes each gambling decisions with knowledge of returns on both gambles. She can coordinate her gambling decisions by combing two gambles: if one gamble turns out to be a loss and another gamble turns out to be a gain, she will not suffer from the loss aversion. In fact, she either invests in two gambles together or does not invest in any of them:

\[
\begin{align*}
x_{1}^{\text{Standard}} &= x_{2}^{\text{Standard}} = 1 & \text{if } \mu (\epsilon_1 + \epsilon_2) > \delta - 1 \\
\text{and} \\
x_{1}^{\text{Standard}} &= x_{2}^{\text{Standard}} = 0 & \text{if } \mu (\epsilon_1 + \epsilon_2) < \delta - 1.
\end{align*}
\]

Now let us turn to the narrow thinker. Each self \( i \in \{1, 2\} \) is in charge of deciding whether to bet on gamble \( i \) or not, i.e. the decision on \( x_i \in \{0, 1\} \). Each self \( i \) perfectly knows the return of her own gamble \( \mu_i \), but does not know the return on the other gamble \( \mu_{-i} \). She instead needs to use \( \mu_i \) to infer \( \mu_{-i} \).\(^{45}\)

As a result, the narrow thinker cannot perfectly coordinate their gambling decisions. This difficulty in coordinating decisions makes it harder for the narrow thinker to enjoy the benefits of combing two gambles together. This leads to a lower probability of investing in each gamble and provides a model of myopic loss aversion.

In particular, Figure 2 plots the probability of investing in each gamble, i.e. \( \text{Prob}(x_i = 1) \), for the standard decision maker and the narrow thinker. It plots the probability of gambling as a function of the noise in individual return \( \sigma_e \), while setting \( \delta = 2.25 \) (Tversky and Kahneman, 1992) and normalizing \( \sigma = 1 \). We can see that narrow thinker always has a lower probability of investing in each gamble.

**Characterizing the narrow thinker’s behavior.** I use guess and verify approach. I guess the threshold equilibrium for the narrow thinker takes the following form: for \( i \in \{1, 2\} \),

\[
\begin{align*}
x_{i}^{\text{Narrow}} &= 1 & \text{if } \mu_i > \bar{\mu} \\
x_{i}^{\text{Narrow}} &= 0 & \text{if } \mu_i < \bar{\mu}.
\end{align*}
\]

Now consider self \( i \in \{1, 2\} \)’s optimal decision, who assumes that the other self \( -i \)'s decision is given by the above threshold equilibrium. If she did not invest, her utility is given by

\[
V^0 (\mu_i) \equiv 0.5 \text{Prob} (\mu_{-i} > \bar{\mu} | \mu_i) \left( -\delta + 1 + \mathbb{E} [\mu_{-i} | \mu_{-i} > \bar{\mu}, \mu_i] \right).
\]

\(^{45}\)In the current set up, each self's only information about \( \mu_{-i} \) comes from her own gamble \( \mu_i \). She does not receive an additional signal about \( \mu_{-i} \). This helps facilitate the characterization of the narrow thinker’s behavior below: each self’s gambling strategy can be characterized by a single-dimension cutoff strategy.
If she invests, her utility is given by

\[ V^1 (\mu_i) \equiv 0.5 \left[ \text{Prob} (\mu_{-i} < \bar{\mu} | \mu_i) (-\delta + 1 + \mu_i) + \text{Prob} (\mu_{-i} > \bar{\mu} | \mu_i) (-\delta + \mu_i + \mathbb{E} \{ \mu_{-i} | \mu_{-i} > \bar{\mu}, \mu_i \}) + 1) \right] \\
= V^0 (\mu_i) + 0.5 \left[ \text{Prob} (\mu_{-i} < \bar{\mu} | \mu_i) (-\delta + 1) + \mu_i \right]. \]

At the threshold \( \bar{\mu} \), we should have \( V^1 (\bar{\mu}) = V^0 (\bar{\mu}) \). That is,

\[ \text{Prob} (\mu_{-i} < \bar{\mu} | \mu_i) (\delta - 1) = \bar{\mu}. \]

Based on the distribution of \( \mu_i \) and \( \mu_{-i} \), I can find the unique solution for the threshold \( \bar{\mu} \) in the above expression. I can then calculate the probability of investing in each gamble for the narrow thinker in Figure 2.

**Appendix F: Endogenous Narrow Thinking**

**Revisiting the Simple Consumer Theory Example**

I revisit the simple consumer theory example in Section 3. That is, the consumer’s utility is given by

\[ u (x_1, x_2, p_1, p_2) = v (x_1, x_2) + w - p_1 x_1 - p_2 x_2. \]  

(69)

Here, for analytical solution, I assume \( v \) is strictly concave and quadratic. Alternatively, one can consider a quadratic or log-quadratic approximation of a more general utility function.
At the information side, I do not directly impose that each self has perfect knowledge of \( p_i \). Instead, I let the decision maker choose endogenously the precision of each self’s signal about \( p_1 \) and \( p_2 \). Specifically, each potential signal \( \omega_i = \{s_{i,1}, s_{i,2}\} \in \Omega_i \) for decision \( i \) consists of a noisy signal about \( p_1 \), \( s_{i,1} = p_1 + \epsilon_{i,1} \), and a noisy signal about \( p_2 \), \( s_{i,2} = p_2 + \epsilon_{i,2} \). All \( \epsilon \)s and \( p \)s are Normally distributed and independent from fundamentals and each other. The variances of the noises in these signals are free to choose, subject to the cognitive constraint in (28).

**Proposition 14** In the optimum in (27):

\[
(\sigma^*_{1,1})^2 < (\sigma^*_{2,1})^2 \quad \text{and} \quad (\sigma^*_{2,2})^2 < (\sigma^*_{1,2})^2,
\]

where \( \sigma^*_{i,k} \) is the variance of the noise of self \( i \)’s signal about \( p_k \) in the optimum.

Proposition 14 means that, in the optimum, self 1’s signal about \( p_1 \) is more precise than self 2’s signal about \( p_1 \). Similarly, self 2’s signal about \( p_2 \) is more precise than self 1’s signal about \( p_2 \). As \( p_i \) directly influences self \( i \)’s optimal consumption rule, it is optimal for self \( i \) to have a more precise signal about \( p_i \) than the other self. In other words, even though I do not impose that each self \( i \) knows more about the price of the good she buys, she endogenously chooses to know more about it. In this sense, narrow thinking can arise endogenously.

**Proof of Proposition 14.** As discussed in the main text, the problem in (27) can be divided into two subproblems, the optimal information choice subject to the cognitive constraint in (28), and the optimal decisions given the chosen information. From the utility (69), given any chosen information \( \{\omega_i\}_{i=1}^2 \), the optimal decision rule \( \{x^*_i(\omega_i)\}_{i=1}^2 \) for \( i \in \{1, 2\} \) can be characterized by

\[
E_i \left[ \frac{\partial v}{\partial x_i} \left( x^*_i(\omega_i), x^*_{-i}(\omega_{-i}) \right) - p_i \right] = 0. \tag{70}
\]

Using law of iterated expectations, we henceforth have

\[
E \left[ x^*_i(\omega_i) \frac{\partial v}{\partial x_i} \left( x^*_i(\omega_i), x^*_{-i}(\omega_{-i}) \right) - p_i x^*_i(\omega_i) \right] = 0.
\]

Substituting into the decision maker’s utility function, and using the fact that \( v \) is quadratic, the optimal
information choice in (28) is then equivalent to \(^{46}\)

\[
\max_{\{\omega_i \in \Omega_i\}_{i=1}^2} -\frac{1}{2} E \left[ p_1 x_1^* (\omega_1) + p_2 x_2^* (\omega_2) \right] \\
\text{s.t. } x_i^* (\omega_i) \text{ satisfy (70)} \\
\sum_{i=1}^2 I \left( \omega_i; \hat{\theta} \right) \leq \tau.
\]

Now, given the \(\Omega_i\) specified above, any \(\omega_i = \{s_{i,1}, s_{i,2}\}\) takes the form of \(s_{i,1} = p_1 + \varepsilon_{i,1}\) and \(s_{i,2} = p_2 + \varepsilon_{i,2}\), with \(\varepsilon_{i,1} \sim N \left( 0, \sigma_{i,1}^2 \right)\), \(\varepsilon_{i,2} \sim N \left( 0, \sigma_{i,2}^2 \right)\) and all \(\varepsilon\)s and \(p\)s are independent from each other. Similar to the proof of Proposition 8, we have

\[
\begin{align*}
\left( \frac{\partial E[x_1^* (\omega_1)|p_1, p_2]}{\partial p_1}, \frac{\partial E[x_2^* (\omega_2)|p_1, p_2]}{\partial p_2} \right) &= \left( I_N - \begin{pmatrix} 1 & \lambda_{1,1} \\ \lambda_{2,1} & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} -\lambda_{1,1} \psi_1 \\ 0 \end{pmatrix}, \\
\left( \frac{\partial E[x_1^* (\omega_1)|p_1, p_2]}{\partial p_2}, \frac{\partial E[x_2^* (\omega_2)|p_1, p_2]}{\partial p_1} \right) &= \left( I_N - \begin{pmatrix} 1 & \lambda_{1,2} \\ \lambda_{2,2} & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} 0 \\ -\lambda_{2,2} \psi_2 \end{pmatrix},
\end{align*}
\]

where \(\lambda_{i,j} = \frac{\sigma_i^2}{\sigma_j^2} \in (0, 1)\) and \(\psi_i = -\left( \frac{\partial \mu_i}{\partial \sigma_i^2} \right)^{-1} > 0\). The problem in (71) then becomes

\[
\max_{0 \leq \lambda_{i,j} \leq 1} \quad g \left( \{\lambda_{i,j}\} \right) \equiv -\frac{1}{2} \frac{\lambda_{1,1} u_{1,2}}{u_{1,1} u_{2,2} - \lambda_{1,1} \lambda_{2,1} (u_{1,2})^2} \sigma_{1,2}^2 - \frac{1}{2} \frac{\lambda_{2,2} u_{1,1}}{u_{1,1} u_{2,2} - \lambda_{2,2} \lambda_{1,2} (u_{1,2})^2} \sigma_{1,2}^2 \\
\text{s.t. } h \left( \{\lambda_{i,j}\} \right) \equiv \sum_{1 \leq i,j \leq 2} \frac{1}{2} \log_2 \left( \frac{1}{1 - \lambda_{i,j}} \right) \leq \tau,
\]

where I use the fact that all \(\varepsilon\)s and \(p\)s are independent from each other, and \(I (\omega_i; \hat{\theta}) = \frac{1}{2} \log_2 \left( \frac{1}{1 - \lambda_{i,1}} \right) + \frac{1}{2} \log_2 \left( \frac{1}{1 - \lambda_{i,2}} \right)\).

Now we prove Proposition 14. If, in the optimum, we have \((\sigma_{1,1}^*)^2 \geq (\sigma_{2,1}^*)^2\). This means \(\lambda_{1,1}^* \leq \lambda_{2,1}^*\), where \(\lambda_{i,j}^* = \frac{\sigma_i^2}{\sigma_j^2} \in (0, 1)\). As a result, \(\frac{\partial h \left( \{\lambda_{i,j}^*\}_{1 \leq i,j \leq 2} \right)}{\partial \lambda_{1,1}^*} > \frac{\partial h \left( \{\lambda_{i,j}^*\}_{1 \leq i,j \leq 2} \right)}{\partial \lambda_{2,1}^*}\) and \(\frac{\partial h \left( \{\lambda_{i,j}^*\}_{1 \leq i,j \leq 2} \right)}{\partial \lambda_{1,1}^*} \leq \frac{\partial h \left( \{\lambda_{i,j}^*\}_{1 \leq i,j \leq 2} \right)}{\partial \lambda_{2,1}^*}\). This is inconsistent with the first order condition of (72):

\[
\frac{\partial g \left( \{\lambda_{i,j}^*\}_{1 \leq i,j \leq 2} \right)}{\partial \lambda_{1,1}^*} \bigg/ \frac{\partial h \left( \{\lambda_{i,j}^*\}_{1 \leq i,j \leq 2} \right)}{\partial \lambda_{1,1}^*} = \frac{\partial g \left( \{\lambda_{i,j}^*\}_{1 \leq i,j \leq 2} \right)}{\partial \lambda_{2,1}^*} \bigg/ \frac{\partial h \left( \{\lambda_{i,j}^*\}_{1 \leq i,j \leq 2} \right)}{\partial \lambda_{2,1}^*}.
\]

Therefore, \((\sigma_{1,1}^*)^2 < (\sigma_{2,1}^*)^2\). Similarly, we can prove \((\sigma_{2,2}^*)^2 < (\sigma_{1,2}^*)^2\).

\(^{46}\)In (71), I omit constants that are independent from the information choice.
The General Case: Flexible Information Acquisition

Environment. In the above example, I restrict each potential $\omega_i \in \Omega_i$ to have a particular form: each $\omega_i$ consists of $N = 2$ noisy signals, one for each price. This is consistent with the information structure studied in the rest of the paper. An alternative is to let the potential signals depend on the fundamental flexibly. The only restriction on potential signals is that $\omega_1, \omega_2, \cdots, \omega_N$ are always conditionally independent given $\tilde{\theta}$. That is, the noise in each decision $i$'s signal about $\tilde{\theta}$ is idiosyncratic.\(^{47}\) With such flexible form of information acquisition, I can achieve a sharp characterization about the optimum of (27) in the general multiple-decision setting.

Specifically, in this subsection, I allow arbitrary concave and quadratic utility functions $u$, and arbitrarily Normally correlated fundamentals $\tilde{\theta}$. For notation simplicity, I normalize the mean of $\tilde{\theta}$ to be $\bar{0}$. Without loss of generality, I also restrict that $u$ does not have terms which are linear functions of $\bar{x}$. Such terms will only add a constant to each optimal decision rule and will not influence the optimal information choice for each self.

To facilitate the exposition, I use $\{\omega_i^*\}_{i=1}^N$ to denote the optimally chosen signals and $\{x_i^* (\cdot)\}_{i=1}^N$ to denote the optimal chosen decision rules. I then use $\tau_i^* = I (\omega_i^*; \tilde{\theta})$ to denote the cognitive capacity allocated for decision $i$ in the optimum.

Optimal information choice. I first study the form of optimal information $\omega_i^*$ for each decision $i$, given the cognitive capacity allocated for decision $i$, $\tau_i^*$.

Lemma 8 With unrestricted $\Omega_i$, in the optimum of (27), each decision $i$ is based on an one-dimensional signal $s_i^*$:\(^{48}\)

$$\omega_i^* = \{s_i^*\} \quad \text{and} \quad s_i^* = \vartheta_i + \mathbb{E} \left[ \sum_{j \neq i} \gamma_{i,j} x_j^* (\omega_j^*) \right] + \epsilon_i \equiv t_i + \epsilon_i,$$

(73)

where $\vartheta_i \equiv \sum_{1 \leq m \leq M} \psi_{i,m} \theta_m$ is a linear function of $\tilde{\theta}$ that summarizes how the fundamental directly influences optimal decision $i$, holding other decisions fixed, $\psi_{i,m} = -\frac{\partial^2 u}{\partial x_i \partial \theta_m} \left( \frac{\partial^2 u}{\partial x_i^2} \right)^{-1}$ and $\gamma_{i,j} = -\frac{\partial^2 u}{\partial x_i \partial x_j} \left( \frac{\partial^2 u}{\partial x_i^2} \right)^{-1}$.\(^{48}\)

In (73), $\epsilon_i \sim N (0, \sigma_i^2)$ is the idiosyncratic noise in the signal, $\sigma_i^2$ is pinned down by $\frac{1}{2} \log_2 \left( \frac{\sigma_i^2 + \sigma_j^2}{\sigma_j^2} \right) = \tau_i^*$, $\sigma_j^2$ is the variance of $t_i$ defined in (73). Without the cognitive constraint, the optimal decision $i$ is given by $\vartheta_i + \sum_{j \neq i} \gamma_{i,j} x_j^* (\omega_j^*)$. Now, with limited cognitive capacity, Lemma 8 shows that the optimal information for decision $i$ will be given by a signal about the fundamental $\tilde{\theta}$ that is closest to $\vartheta_i + \sum_{j \neq i} \gamma_{i,j} x_j^* (\omega_j^*)$. The variance of the noise in this signal is pinned down by decision $i$’s allocated cognitive capacity $\tau_i^*$.

\(^{47}\) This follows the literature on information acquisition in games (e.g. Yang, 2015, Morris and Yang, 2019). Such assumption can be justifiable as the noise in each self’s signal comes from cognitive costs to perfectly track the fundamental. Based on this assumption, different decisions’ signals will always be different because of the idiosyncratic noise. Nevertheless, this section focuses on how different decisions’ signals take different forms.

\(^{48}\) For each $i$, the optimal signal $s_i^*$ is unique up to a linear transformation. That is, from an informational perspective, $s_i^*$ is equivalent to $\alpha s_i^* + \beta$, where $\alpha \neq 0$ and $\beta$ are scalars.
As different decisions are based on different decision rules, each self is “interested in” different parts of the fundamentals. As a result, the optimal signals for different decisions take different forms. In this sense, narrow thinking arises endogenously.

Given the optimal signal in (73), one can then solve optimal decision rules \( \{x_i^*(\cdot)\}_{i=1}^N \). From (73), we know each self’s optimal signal in turn depends on other selves’ optimal decision rules. Solving this fixed-point problem, one can then characterize how each optimal signal \( s_i^* \) depends on the fundamental \( \vec{\theta} \).

**Proposition 15** The optimal signals depend on the fundamental \( \vec{\theta} \) as follows:

\[
\begin{pmatrix}
E[s_1^*|\vec{\theta}]

\vdots

E[s_N^*|\vec{\theta}]
\end{pmatrix} = \begin{pmatrix}
t_1

\vdots

t_N
\end{pmatrix} = \left( I_N - \begin{pmatrix}
1 & \lambda_1 & \cdots & \lambda_1 & \lambda_1 \\
\lambda_2 & 1 & \cdots & \lambda_2 & \lambda_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\lambda_{N-1} & \lambda_{N-1} & \cdots & 1 & \lambda_{N-1} \\
\lambda_N & \lambda_N & \cdots & \lambda_N & 1
\end{pmatrix} \right) \circ \Gamma^{-1} \begin{pmatrix}
\vartheta_1 \\
\vdots \\
\vartheta_k \\
\vdots \\
\vartheta_N
\end{pmatrix},
\]

where \( \lambda_i = \frac{\sigma_i}{\sigma_i + \sigma_{i^*}} = 1 - 2^{2\tau_i^*} \in (0, 1) \) is pinned down by decision i’s allocated cognitive capacity \( \tau_i^* \).

Similar to Proposition 8, as self \( i \) does not perfectly know self \( j \)’s decision, the effective degree of interaction from decision \( j \) to decision \( i \) is attenuated by a factor \( \lambda_i \) between 0 and 1. As the effective interaction across decisions is attenuated, optimal decision \( i \) will be influenced more by \( \vartheta_i \), summarizing the fundamental’s direct influence. This in turn lets the optimal signal for self \( i \) depend more on her own \( \vartheta_i \). To further illustrate the last point, consider a symmetric optimum for the problem in (27) with two decisions. From (74), we have

\[
\begin{pmatrix}
t_1 \\
t_2
\end{pmatrix} = \left( I_2 - \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix}
\vartheta_1 \\
\vartheta_2
\end{pmatrix} = \begin{pmatrix}
\frac{\vartheta_1}{1 - \gamma \lambda^2} + \lambda \gamma \frac{\vartheta_2}{1 - \gamma \lambda^2} \\
\frac{\vartheta_2}{1 - \gamma \lambda^2} + \lambda \gamma \frac{\vartheta_1}{1 - \gamma \lambda^2}
\end{pmatrix},
\]

where \( \gamma = -\frac{\partial^2 u}{\partial x_1 \partial x_2} \left( \frac{\partial^2 u}{\partial x_1^2} \right)^{-1} \). One can see that, for the optimal signal for each self \( i \), the weight on the other self’s \( \vartheta_{-i} \) compared to her own \( \vartheta_i \) is attenuated by the factor \( \lambda \) between 0 and 1. In this sense, the within-person coordination friction induces the optimal signal for each self \( i \) to depend more on her own \( \vartheta_i \). In fact, when the cognitive constraint is severe (\( \tau \) is small so \( \lambda \) is close to zero), the optimal signal for each self \( i \) will effectively only depend on \( \vartheta_i \). The decision maker becomes a “completely” narrow thinker: each decision \( i \) is only based on her own \( \vartheta_i \), i.e. the fundamental’s direct influence.

**Allocation of cognitive capacities across different decisions.** We finally turn to the optimal allocation of cognitive capacities, \( \tau_i^* \), across different decisions.
Proposition 16. In the optimum of (27),
\[ \tau_i^* = \tau_j^* \iff \frac{\partial^2 u}{\partial x_i^2} \text{Var}(t_i) > \frac{\partial^2 u}{\partial x_j^2} \text{Var}(t_j). \]

Proposition 16 shows that more volatile decisions (with high \( \text{Var}(t_i) \)) and decisions with respect to which the marginal utility is more sensitive (with high \( \frac{\partial^2 u}{\partial x^2} \)) will be based on more precise information. For example, the decision maker may allocate more cognitive capacity to the self who is in charge of purchasing computers (high \( \frac{\partial^2 u}{\partial x_i^2} \)) than to the self who is in charge of purchasing apples (low \( \frac{\partial^2 u}{\partial x_i^2} \)). Similarly, the decision maker may allocate more cognitive capacities to the self who invests bitcoins (volatile \( t_i \)) than to the self who invests ETFs (stable \( t_i \)).

Proof of Lemma 8. A necessary condition for \( \{\omega_i^*, x_i^*(\cdot)\}_{i=1}^N \) to be an optimum of (27) is that, for each \( i \), \( (\omega_i^*, x_i^*(\cdot)) \) is optimally chosen, taking other \( \{\omega_j^*, x_j^*(\cdot)\}_{j \neq i} \) as given. That is, \( (\omega_i^*, x_i^*(\cdot)) \) solves
\[
\max_{\omega_i \in \Omega_i, x_i(\cdot)} E \left[ u \left( x_1(\omega_1), \ldots, x_N^*(\omega_N^*), \tilde{\theta} \right) \right] \\
\text{s.t.} \quad I(\omega_i; \tilde{\theta}) = \tau - \sum_{j \neq i} I(\omega_j^*; \tilde{\theta}),
\]
(75)
(76)
As \( u \) is quadratic, maximizes the objective in (75) is equivalent to maximizing
\[
E \left[ \frac{u_{i,i}}{2} \left( x_i(\omega_i) - \text{Var}(t_i) \right)^2 + \frac{u_{i,i}}{2} \sum_{j \neq i} \gamma_{i,j} \left( x_j^*(\omega_j^*) - \text{Var}(t_j) \right)^2 \right] \\
+ f \left( \{x_j^*(\omega_j^*)\}_{j \neq i}, \tilde{\theta} \right),
\]
(77)
where \( u_{i,i} = \frac{\partial^2 u}{\partial x_i^2} \) and I use the fact that \( \omega_1, \omega_2, \ldots, \omega_N \) is conditionally independent given \( \tilde{\theta} \). The problem based on the objective in (77) and the constraint in (76) is then the standard tracking problem with quadratic loss function and Normally distributed target. From Sims (2003), we know the optimal signal \( \omega_i \) takes the form in Lemma 8.

Proof of Proposition 15. Given the chosen information \( \{\omega_i^*\}_{i=1}^N \), the optimal decision rule \( \{x_i^*(\cdot)\}_{i=1}^N \) can be characterized by
\[
x_i^*(\omega_i^*) = E \left[ \tilde{\theta}_i + \sum_{j \neq i} \gamma_{i,j} x_j^* (\omega_j^*) | s_i^* \right] = \lambda_i s_i^*,
\]
\(49\)
One may wonder whether it is the case that \( \tau_i^* > \tau_j^* \iff \frac{\partial^2 u}{\partial x_i^2} \text{Var}(v_i) > \frac{\partial^2 u}{\partial x_j^2} \text{Var}(v_j) \), where \( \text{Var}(v_i) \) is defined above, summarizing how the fundamental directly influences optimal decision \( i \). This is not necessarily the case. Even if \( \tilde{\theta}_i \) is volatile, if the \( \tilde{\theta}_j \)s of all other decisions who influence decision \( i \) (i.e. decisions \( j \)s such that \( \gamma_{i,j} > 0 \)) are not volatile, the decision maker may not want to allocate a lot of cognitive capacities to self \( i \).
where \( \lambda_i = \frac{\sigma_i^2}{\sigma^2_i + \sigma_{i'}^2} \). Together with (73), we have

\[
t_i = \vartheta_i + \sum_{j \neq i} \lambda_j \gamma_{i,j} t_j.
\]

This leads to (74).

**Proof of Proposition 16.** For \( \{0 \leq \Lambda_i \leq 1, \xi_i^2\}_{i=1}^N \), define \( g \left( \{ \Lambda_i, \xi_i^2 \}_{i=1}^N \right) \equiv E \left[ u \left( x_1, \ldots, x_N, \hat{\theta} \right) \right] \), where for all \( i, x_i = \Lambda_i \left( t_i + \epsilon_i \right), \epsilon_i \sim N \left( 0, \xi_i^2 \right) \), and \( \{ t_i \}_{i=1}^N \) are given by

\[
\begin{pmatrix}
  t_1 \\
  \vdots \\
  t_N
\end{pmatrix} = \left( I_N - \begin{pmatrix}
  1 & \Lambda_1 & \ldots & \Lambda_1 & \Lambda_1 \\
  \Lambda_2 & 1 & \ldots & \Lambda_2 & \Lambda_2 \\
  \Lambda_{N-1} & \Lambda_{N-1} & \ldots & 1 & \Lambda_{N-1} \\
  \Lambda_N & \Lambda_N & \ldots & \Lambda_N & 1
\end{pmatrix} \otimes \Gamma \right)^{-1} \begin{pmatrix}
  \vartheta_1 \\
  \vdots \\
  \vartheta_k \\
  \vdots \\
  \vartheta_N
\end{pmatrix}.
\] (78)

Based on Lemma 8 and Proposition 15, in the optimum of (27), \( \{ \lambda_i, \sigma_i^2 \}_{i=1}^N \) defined in Lemma 8 and Proposition 15 must solve

\[
\max_{\{ \Lambda_i, \xi_i^2 \}_{i=1}^N} g \left( \{ \Lambda_i, \xi_i^2 \}_{i=1}^N \right) = \frac{1}{2} \sum_i \log_2 \frac{\xi_i^2 + \sigma_i^2}{\xi_i^2} \leq \tau,
\] (79)

where \( \sigma_i^2 \) is the variance of \( t_i \) defined based on (78). Rewrite \( g \) in a way similar to (77), we have

\[
\frac{\partial g \left( \{ \Lambda_i, \xi_i^2 \}_{i=1}^N \right)}{\partial \left( \xi_i^2 \right)} = \frac{u_i \Lambda_i^2}{2 \xi_i^2} \text{ and } \frac{\partial h \left( \{ \xi_i^2 \}_{i=1}^N \right)}{\partial \left( \xi_i^2 \right)} = -\frac{1}{2 \log_2} \frac{\sigma_i^2}{\xi_i^2 \left( \sigma^2_i + \xi_i^2 \right)}
\]

where \( u_i = \frac{\partial^2 u}{\partial x_i^2} \). As \( \{ \lambda_i = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_{i'}^2}, \sigma_i^2 \}_{i=1}^N \) must solve (72), in the optimum of (27), we must have

\[
u_{i,i} \Lambda_i^2 / \left( \frac{\sigma_i^2}{\sigma_i^2 \left( \sigma_i^2 + \sigma_{i'}^2 \right)} \right) = u_{j,j} \Lambda_j^2 / \left( \frac{\sigma_j^2}{\sigma_j^2 \left( \sigma_j^2 + \sigma_{j'}^2 \right)} \right) \forall i,j,
\]

and

\[
\nu_{i,i} \sigma_i^2 \left( 1 - \lambda_i \right) = u_{j,j} \sigma_j^2 \left( 1 - \lambda_j \right).
\]

As \( \lambda_i = 1 - 2^{-2\tau_i} \), Proposition 16 follows.
References


Marschak, Jacob, and Roy Radner. (1972) *Economic Theory of Teams*. 77


