# A Theory of Narrow Thinking*

**Chen Lian**†

May 24, 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. The narrow thinker then makes each decision with an imperfect understanding of the others. Formally, it is as if the decision-maker is a collection of multiple selves playing an incomplete-information game. The friction effectively attenuates the degree of interaction across decisions and can translate into either over- or under-reaction depending on the environment. Narrow thinking leads to a violation of the fungibility principle and a smooth model of mental accounting. Narrow thinking also reconciles other seemingly disparate phenomena in a unified framework, such as excess smoothness to taste shocks, the small wage elasticity of daily labor supply, and the label effect. Finally, I study an endogenous narrow thinking problem: the decision maker chooses optimally what information each decision is based upon, subject to a cognitive constraint.

**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, Ludwig Straub, Juuso Toikka, Laura Veldkamp, Olivier Wang, Ivan Werning, and seminar participants at ASSA, Barcelona GSE, D-TEA and MIT for insightful comments.

†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 perfect self-coordination among 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 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), 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 in fact break the decision-maker's ability to perfectly coordinate her multiple decisions, and can explain seemingly disparate behavioral phenomena in a unified framework.

**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 top of her mind. When she purchases gasoline, she knows the gasoline price, but does not have the food price at the top of her mind. As her two consumption decisions are based on different and non-nested information, such a decision maker is a narrow thinker as defined above. As explained shortly, such within-person, cross-decision, frictions cause the decision-maker to effectively discount the influence of other decisions when making a particular decision.

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 $\vec{\theta}$: $u\left(x_1, \cdots, x_N, \vec{\theta}\right)$. 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 show the decision problem under narrow thinking is mathematically equivalent to multiple selves playing an *incomplete-information*, common interest, game. As each self does not perfectly know other selves' signals (states of mind), each self can no longer perfectly predict other selves' decisions. In the unique equilibrium of the game, each decision is then made with an imperfect understanding of other decisions. In this sense, narrow thinking formalizes within-person coordination frictions.

**Behavior under narrow thinking, under-reaction and over-reaction.** I then establish several general properties of the narrow thinker's behavior. As each self of the narrow thinker has an imperfect perception of the others' decisions, her beliefs about the others' decisions are anchored in response to shocks to fundamentals. Such belief anchoring leads to an effective attenuation of interaction across decisions: it is as if each of the narrow maker's decision is less influenced by other decisions, and she thinks "narrowly." Narrow thinking then leads to a dampening of indirect effects — the movement of one decision driven by the movement of other decisions.

Depending on the environment, such a dampening of indirect effects under narrow thinking can translate into either under- or over-reaction. I introduce a general principle which helps predict whether the narrow thinker over- or under-reacts in a specific economic environment. When the indirect effect works in the same direction as the direct effect (i.e. the influence of the fundamental $\vec{\theta}$ on each decision, while holding other decisions fixed), 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. The analysis then contrasts with the often-held belief that noises in the decision maker's mental representation of the world typically lead to under-reaction.

**Applications.** Using this principle, I then lay out the implications of narrow thinking for concrete economic applications. I start from the classical consumer problem of demanding multiple goods, and I analytically solve two polar cases to tease out the mechanism. In the first, the decision maker's utility is quasi-linear, there are no income effects and the interaction across different consumption decisions comes from the complementarity/substitutability embedded in the utility function. In the second, the decision maker's utility is separable but there are income effects. The interaction across different decisions now comes from the budget constraint. Guided by the general principle above, I will explain why narrow thinking leads to excess smoothness to own price shocks in the first case, but excess sensitivity in the second case. I then further explore how narrow thinking can explain other empirical examples of under-reaction (e.g. excess smoothness to taste shocks (Heath and Soll, 1996) and the small wage elasticity of daily labor supply (Camerer et al., 1997) and over-reaction (e.g. excess sensitivity to income shocks (Thaler, 1999) and the label effect (Abeler and Marklein, 2016).

**Violation of the fungibility principle and a "smooth" mental accounting model.** One application of independent interest is the above case of the consumer problem with income effects. Under narrow thinking, different selves have different beliefs about the marginal value of money. The narrow thinker hence violates the fungibility principle: she behaves as if the money allocated to one good cannot perfectly substitute for the money allocated to another. Such a violation then generates mental accounting-type behavior which has "no agreed-upon model" (Farhi and Gabaix, 2015), e.g. excess sensitivity to own-price changes (Hastings and Shapiro, 2013).

To illustrate how narrow thinking can generate such excess sensitivity, consider an increase in the food price. The price increase has a negative direct effect on food consumption. Under standard consumer theory, the decision maker can decrease other consumption to smooth out the drop in food consumption, exerting a positive indirect effect on food consumption. Under narrow thinking, however, the coordinated response of other consumption is limited, and the indirect effect from smoothing out other consumption is dampened. As the indirect effect works in the opposite direction to the direct effect, its dampening then leads to excess sensitivity to own-price changes: the narrow thinker decreases food consumption more in response to the price increase.

The narrow thinking approach provides a model of the violation of the fungibility principle without relying on the explicit mental budgeting in Heath and Soll (1996). Nevertheless, narrow thinking brings the consumer's demand elasticity closer to that under the explicit mental budgeting model. In this sense, narrow thinking leads to a "smooth" model of mental accounting. The approach also generates new testable predictions. For example, excess sensitivity to own price changes under narrow thinking only happens for goods with respect to which the decision maker's demand is not very elastic.

**Relationship between narrow thinking and rational inattention.** The narrow thinking approach builds upon the rational inattention literature (Sims, 2003; Matejka and McKay, 2015; Koszegi and Matejka, 2018) by using imprecise information (noisy signals) to capture bounded rationality, but with a few key differences. The key friction of interest for narrow thinking is the decision maker's difficulty in coordinating her multiple decisions. The narrow thinker's different decisions are based on *different* information, and each of her decision is then made with an imperfect understanding of other decisions. By contrast, the key friction of interest for rational inattention is the decision maker's imperfect perception of the fundamental. When applying it to static multiple-decision problems such as those studied in this paper (e.g. demand of multiple goods), different decisions are based on the *same*, imperfect, information (e.g. Koszegi and Matejka, 2018). The decision maker then perfectly knows her other decisions when making a particular decision.[1]

[1]When applying the rational inattention approach to dynamic problems (Steiner, Stewart and Matejka, 2017), similar to the standard sequential decision problem, the typical assumption is that the information of the earlier

In a complementary approach to rational inattention, Gabaix (2014) develops a novel "sparsity" method to model the decision maker's sparse representation of the state of the world. There, multiple decisions are made based on the *same,* imprecise, perception of the fundamental.[2]

**A bounded recall interpretation in sequential settings.** The decision problem under narrow thinking is analyzed through the lens of a static, incomplete-information, game among multiple selves, as the multiple-decision problems studied in the paper are traditionally treated as static (e.g. demand of multiple goods). One may naturally wonder how to interpret the above analysis if the sequential nature of the decision-making process is modeled explicitly. In such a sequential setting, I explain how the form of narrow thinking studied in the paper can be interpreted as a particular form of bounded recall, and how it is observationally similar to other forms of bounded recall. A particular feature of my approach is that, because of the equivalence to a static incomplete information game, the narrow thinker's behavior is insensitive to the exact order of the decisions. As a result, one can view my approach as a reduced form method to study the impact of bounded recall and selective retrieval of memory, when the analyst does not want to take a stand on the exact order of decisions.

**Costly contemplation.** The main analysis lets different decisions be based on different, but exogenous, information. In the last part of the paper, I study a "costly contemplation" problem in which such information is endogenized. 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 costly contemplation problem studies the optimal information choice problem at the *decision*-level, going beyond the standard rational inattention paradigm. It captures the idea that, when the decision maker makes a particular decision, she cannot effortlessly use/recall the information used for other decisions.

As different decisions are based on different decision rules, each self is "interested in" different parts of the fundamental. As a result, it is optimal for different selves' signals to take different forms. For example, in consumer theory, it is optimal for each self to receive a more precise signal about the price of the good she buys. In this sense, narrow thinking arises endogenously.

**Layout.** Section 2 defines the notion of narrow thinking, and provides a game theoretic representation. Section 3 establishes a few key properties of optimal behavior under narrow thinking. Section 4 discusses the interpretation in sequential settings. Section 5 focuses on narrow thinking's predictions in concrete economic applications. Section 6 studies the costly contemplation problem. Section 7 concludes. The Appendix contains proofs and additional results.

---

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. This will be discussed in detail in Section 2.

[2] Gabaix (2014)'s sparsity approach does not use noisy signals and the perception of the fundamental there is imperfect but deterministic.

## 1.1 Related Literature

This paper builds on, and adds to, the growing literature on rational inattention (Sims, 2003; Mackowiak and Wiederholt, 2009; Matejka, 2015, 2016; Matejka and McKay, 2015; Koszegi and Matejka, 2018)[3] and sparsity (Gabaix, 2014, 2017). There, the key friction is the decision maker's imperfect perception of the fundamental, while the decision maker perfectly knows her other decisions when making a particular decision. On the other hand, the narrow thinking approach lets different decisions be based on different, non-nested, information and captures the friction that each decision can be made with an imperfect understanding of other decisions.

As the decision problem under narrow thinking is equivalent to multiple selves playing an incomplete information game, the paper also builds upon the literature on incomplete information "beauty contests" (Morris and Shin, 2002; Angeletos and Pavan, 2007; Bergemann and Morris, 2013). This literature studies linear best-response games under incomplete information. A key insight from the literature is that incomplete information can attenuate the equilibrium interaction (Angeletos and Lian, 2016, 2018; Bergemann, Heumann and Morris, 2017). In these works, the behavior of each individual is frictionless and the focus is on *inter*-personal coordination friction and macroeconomic applications. The current paper, on the other hand, focuses on *intra*-personal friction in coordinating a decision maker's multiple decisions and behavioral applications. This change of focus permits me to build a bridge between the incomplete information literature and the bounded rationality literature. The paper then shows how such intra-personal frictions can deliver a novel theory to reconcile seemingly disparate behavioral phenomena.

By viewing the decision maker as a team of multiple selves, the paper also connects to the literature on multiple-selves and team theory. The multiple-selves literature (Piccione and Rubinstein, 1997; Benabou and Tirole, 2002, 2003, 2004; Gottlieb, 2014, 2017) uses environments in which multiple selves have conflicted interests to model motivated beliefs and reasoning, and explores reasons why the decision maker's beliefs and behavior can be systematically biased. The multiple selves of the narrow thinker, on the other hand, have common interests, and the focus of the current paper is about frictional behavior in response to shocks. Narrow thinking does not necessarily lead to systematical bias on average. The team theory literature (Marschak and Radner, 1972; Dessein, Galeotti and Santos, 2016), on the other hand, mostly focuses on optimal information design in an organization. Angeletos and Pavan (2007) develop a method to use team theory to find the constrained efficient allocation in an economy with dispersed information. An-

[3]Other notable contributions in the rational inattention and endogenous information acquisition literature include, but are not limited to: Hellwig and Veldkamp (2009), Myatt and Wallace (2012, 2017), Colombo, Femminis and Pavan (2014), Stevens (2015), Morris and Yang (2016), Hébert and Woodford (2017), Denti (2017*b*) and Steiner, Stewart and Matejka (2017);. See Sims (2010), Veldkamp (2011), Mackowiak, Matejka and Wiederholt (2018) for further references.

geletos and Pavan (2009), Lorenzoni (2010) and Angeletos and La'O (2018) then use the method to characterize optimal policy with informational frictions.

Multiple cognitive frictions can let different decisions be made based on different, non-nested, information. For example, Gennaioli and Shleifer (2010), Kahana (2012), Wilson (2014), Bordalo, Gennaioli and Shleifer (2017) and Jehiel and Steiner (2018) on bounded recall, and Tversky and Kahneman (1973) and Kahneman (2011) on heuristics, biases, and selective retrieval from memory. These works therefore provide complementarity justification for the kind of frictions that define narrow thinking. Gennaioli and Shleifer (2010) use the term "local thinker" to describe a decision maker whose bounded recall follows the representativeness heuristic, and study the implication for single-decision problems. The current paper, on the other hand, focuses on the decision maker's difficulty in coordinating her multiple decisions.

On the applied side, narrow thinking provides a unified framework to explain different behavioral phenomena. Depending on the environment, narrow thinking can translate into either over- or under-reaction. Applications studied in the paper connect to the literature on narrow bracketing (Rabin and Weizsacker, 2009), mental accounting (Thaler, 1985, 1999; Heath and Soll, 1996; Hastings and Shapiro, 2013; Abeler and Marklein, 2016), excessive sensitivity to temporary income shock (Parker et al., 2013; Kueng, 2018), the small cross-price demand elasticity (Gabaix and Laibson, 2006; Abaluck and Gruber, 2011, 2016; Allcott and Taubinsky, 2015) and the small wage elasticity of daily labor supply (Camerer et al., 1997; Farber, 2015; Thakral and To, 2017). For each application, narrow thinking's distinct economic implications and testable predictions will be discussed. Koszegi and Matejka (2018) is a recent, complementary, paper that shares the focus on an information-based theory of mental accounting. That paper stays within the rational inattention paradigm, and different decisions are based on the same, imperfect, information.

# 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. I then explain why such a narrow thinker makes each decision with an imperfect understanding of the others, and the sense in which she faces frictions in coordinating her multiple decisions. I later discuss the psychological justifications for narrow thinking, that is, why different decisions are based on different, non-nested, information. I finally discuss how to map constrained problems to the unconstrained problem introduced here.

## 2.1 Environment and the Definition of Narrow Thinking

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

$$u\left(\vec{x}, \vec{\theta}\right), \tag{1}$$

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.[4]

To accommodate narrow thinking, I introduce *decision*-specific information. For each decision $i \in \{1, \cdots, N\}$, 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. As further discussed in Section 2.3, one should 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 costly contemplation problem in which the decision maker chooses endogenously the information upon which each decision is based. I finally let $\mathcal{F}_i$ be the $\sigma-$algebra (on the probability space) generated by decision $i$'s signal $\omega_i$.[5]

**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

$$\max_{\{x_i(\cdot)\}_{i=1}^N} E\left[u\left(x_1(\omega_1), \cdots, x_N(\omega_N), \vec{\theta}\right)\right]. \tag{2}$$

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 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

[4] 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.

[5] $\mathcal{F}_i \equiv \left\{\omega_i^{-1}(B) : B \in \mathbb{B}_l\right\}$, where $\mathbb{B}_l$ is the collection of Borel set on $\mathbb{R}^l$ and $l$ is the dimensionality of $\omega_i$.

decision maker, and the constraint is a self-specific information constraint.

It is worth noting that, as $u$ is strictly concave over $\vec{x}$, the optimum of (2), if exists, is unique.[6,7]

**Lemma 1** *If the optimum of (2) exists, it is unique.*

**Narrow thinking.** Now, I introduce the notion of narrow thinking used throughout the paper: different decisions are made based on different, non-nested, information.

**Definition 1** *A decision maker is a narrow thinker if there exists a pair of* $(i, j) \in \{1, \cdots, N\}$ *such that* $\mathcal{F}_i \nsubseteq \mathcal{F}_j$ *and* $\mathcal{F}_j \nsubseteq \mathcal{F}_i$.

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.[8]

To understand what the definition of narrow thinking captures, 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 top 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 2.3, I further 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. Formally, it means, for all* $i \neq j$, $\mathcal{F}_i = \mathcal{F}_j$.

In the context of multiple-decision problems that are traditionally treated as static (the main focus of the paper), the notion of broad thinking nests both classical consumer theory and a few standard bounded rationality approaches (e.g. rational inattention and sparsity). To illustrate, consider a standard consumer problem of demanding multiple goods. In the case of standard

[6]Uniqueness is in the sense that, in any two optima, decision rules are the same almost surely.

[7]For generality, I do not restrict the potential set for each $x_i$, $\mathcal{X}_i$, 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.

[8]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 $i$.

consumer theory (Mas-Colell, Whinston and Green, 1995), all decisions are based on the same, perfect, knowledge of the fundamental $\vec{\theta}$, i.e. the price vector. In the case of rational inattention (Koszegi and Matejka, 2018) 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.

## 2.2 Narrow Thinking as an Incomplete Information Game

**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.[9] To further understand the mathematical nature of the problem under narrow thinking in (2) and provide an alternative interpretation, it is useful to provide an equivalent, game-theoretic, representation of (2).[10] 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 2** $\{x_1^*(\cdot), \cdots, x_N^*(\cdot)\}$ *solves (2) if and only if*

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

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 2 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 $\vec{x} = (x_1, \cdots, x_N) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ 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).*

[9] In the context of broad thinking, such a decision-maker level planning problem maps to the standard practice: given her information, the decision maker chooses all her decisions jointly to maximize her expected utility.

[10] The main goal of the game-theoretic representation is to help clarify the mathematical nature of the decision problem under narrow thinking. It does not mean the decision maker literally solves the Bayesian game in her mind. One should view the analysis in the paper as a disciplined method to capture narrow bracketing: the only departure from the standard paradigm is that $\omega_i$, the decision maker's state of mind when deciding on $x_i$, may not summarize all the relevant information.

**Narrow thinking as a formalization of within-person coordination frictions.** The game-theoretic representation then permits me to explain in what sense the narrow thinker faces frictions in coordinating her multiple 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 uncertainty about other selves' information then translates into her uncertainty 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. In other words, by letting each self $i$ have a different signal (state of mind) $\omega_i$, I allow frictions in each self's higher order reasoning: each self can no longer perfectly predict other selves' decisions. As the later analysis shows, such friction effectively attenuates the interaction across decisions in response to shocks to the fundamental and it is as if each self "thinks narrowly."

Under broad thinking, however, different decisions are made based on the same information. The game among multiple selves becomes a complete information game. In the unique Bayesian Nash equilibrium, 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.

## 2.3 Additional Discussion

**Sources of narrow thinking.** 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. In the above consumer theory context, 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, another notion in cognitive psychology (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. Such a notion is also closely related to the study of heuristics and biases in decision making (Tversky and Kahneman, 1973; Gennaioli and Shleifer, 2010). One relevant heuristics here is the "What You See Is All There Is" principle emphasized by Kahneman (2011) and Enke (2018): in the above consumer theory context, when the decision maker purchases food (gasoline), she only sees the food (gasoline) price, and does not have the gasoline (food) price at the top of her mind.

Finally, in section 6, I also study a costly contemplation problem: the decision maker chooses optimally what information each decision is based upon, subject to a cognitive constraint. As different decisions are based on different decision rules, it is optimal for different decisions' signals to take different forms.

In sum, due to these cognitive frictions, $\omega_i$, which captures the *state of mind* when the decision maker decides on $x_i$, may not summarize all the relevant information. This why different decisions can be made based on different, non-nested, information.

**Sequential settings.** As the multiple-decision problems studied in the paper are traditionally treated as static (e.g. demand of multiple goods), the multiple-decision problem in (2) is analyzed through the lens of a static incomplete-information, game among multiple selves. However, it is worth noting that individual level "team" problem in (2) can also be interpreted when the sequential nature of the decision-making process is modeled explicitly. For example, for any $i < j$, one could think that the decision $i$ is made before decision $j$. In this case, the standard practice imposes the information of the earlier decision $i$ is perfectly nested by the information of the later decision $j$ (perfect recall). Formally, it means $\mathcal{F}_i \subseteq \mathcal{F}_j$ for all $i < j$. On the other hand, under narrow thinking, different decisions are made based on different, non-nested, information, because of the cognitive frictions discussed above. That is, $\mathcal{F}_i \nsubseteq \mathcal{F}_j$ for some $i < j$. Section 4 further studies how the narrow thinker's information structure studied in the paper can be mapped to a particular form of bounded recall (or selective retrieval from memory), and how it is related to other forms of bounded recall.

**Planning problem as an incomplete information game.** The decision problem in (2) is a single-agent planning problem with a decision-specific information constraint. I then map its optimum to the Bayesian Nash equilibrium of an incomplete information, common interest, game among multiple selves. The method is also reminiscent of Angeletos and Pavan (2007): they use team theory to find the social planer's constrained efficient allocation in a multiple-agent economy with dispersed information and potentially conflicting interests. They then relate the constrained efficient allocation to the equilibrium of a fictitious game.

**Constrained problems.** 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

$$B\left(\vec{x}, \vec{\theta}\right) \leq 0, \tag{4}$$

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

How to guarantee the constraint is satisfied under bounded rationality is a hotly debated issue in the literature. Here I choose a simple and standard approach, i.e. let the last decision adjust automatically given the constraint and other boundedly rational decisions.[11] Specifically, I let the last decision, $x_{N+1}$, be made with perfect knowledge about the fundamental and other decisions, and this guarantees that (4) holds. 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. For example, in the consumer theory setting (Section 5.1 and 5.2), I allow the last decision, which can be interpreted as saving or borrowing, to be negative. This guarantees that one can find a value for the last decision such that the budget constraint is satisfied. The key friction of interest is then the difficulty of coordinating the first $N$ decisions. In fact, when the constraint in (4) always binds in the optimum, one can use it to substitute $x_{N+1}$ in the utility and the problem becomes an unconstrained problem in (2) for the first $N$ decisions.

# 3 Optimal Behavior under Narrow Thinking

I now turn to optimal behavior under narrow thinking. In this section, I focus on a few theoretical results which illustrate several key abstract insights. As each self of the narrow thinker has an imperfect perception of the others' decisions, narrow thinking leads to an effective attenuation of interaction across decisions and a dampening of indirect effects — the movement of one decision driven by the movement of other decisions. I then establish results about when the underlying friction leads to under-reaction, and when it leads to over-reaction. These insights will then be applied to more concrete applications in Section 5.

## 3.1 Environment

**Optimal decision rules.** In this section, I directly work with the narrow thinker's optimal decision rules, which can be derived from the decision specific optimality conditions in (3). Specifically, the optimal decision rule for decision $i \in \{1, \cdots, N\}$ is given by

$$x_i^*(\omega_i) = E_i\left[BR_i\left(\vec{x}^*_{-i}; \vec{\theta}\right)\right] \equiv \underbrace{E_i\left[\sum_{1\leq k\leq M} \psi_{i,k}\theta_k\right]}_{\text{direct effect}} + \underbrace{E_i\left[\sum_{j\neq i, 1\leq j\leq N} \gamma_{i,j} x_j^*(\omega_j)\right]}_{\text{indirect effect}} \quad \forall i, \omega_i \in \Omega_i, \tag{5}$$

where $E_i[\cdot] = E[\cdot|\omega_i]$ denotes self $i$'s belief.

The first term in (5), which I call the direct effect, summarizes the fundamental $\vec{\theta}$'s direct influence on the optimal decision $i$, holding other decisions fixed. For each $i \in \{1, \cdots, N\}$ and $k \in \{1, \cdots, M\}$, $\psi_{i,k}$ captures how the fundamental $\theta_k$ directly influences the optimal decision $i$.

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

In the equivalent game among multiple selves, the first term in (5) then captures the fundamental's influence on self $i$'s best response function, i.e. the "intercept" of $BR_i\left(\vec{x}^*_{-i};\vec{\theta}\right)$.

The second term in (5), which I call the indirect effect, summarizes how other decisions influence decision $i$. For each $i \neq j \in \{1,\cdots,N\}$, $\gamma_{i,j}$ 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 the equivalent game among multiple selves, $\gamma_{i,j}$ captures the slope of the best response function $BR_i(\cdot)$ with respect to $x_j$. In fact, one can think of (5) as the best response function of a linear network game (Bergemann, Heumann and Morris, 2017; Golub and Morris, 2017*a*,*b*; Denti, 2017*a*). The matrix $\Gamma = \{\gamma_{i,j}\}_{1\leq i,j\leq N}$ can then be interpreted as the interaction matrix, where $\gamma_{i,j}$ captures the "weight" that self $i$ assigns to self $j$'s decision in her best response function.[12]

Compared to the general framework in Section 2, the main restrictions embedded in (5) is that the optimal decision rule (5) is linear. This is naturally true when the utility function is quadratic (see the abstract example below). In the case of general utility, such a linear decision rule can follow from a linear or a log-linear approximation (see the consumer theory example below).

I also let the interaction matrix $\Gamma$ be such that $I-\Gamma$ be invertible. This guarantees a unique solution for (5), and follows naturally from the strict concavity of the utility function. Finally, though not strictly needed and for expositional purposes, I assume $\psi_{i,k} \geq 0$ for all $i$ and $k$.[13]

I now introduce a few concrete examples of the decision maker's optimal decision rules (5).

**An abstract example.** I start from a simple, abstract, example. There are $N$ decisions $\vec{x} = (x_1,\cdots,x_N) \in \mathbb{R}^N$ and $N$ fundamentals $\vec{\theta} = (\theta_1,\cdots,\theta_N) \in \mathbb{R}^N$. The decision maker's utility is quadratic and given by:

$$u\left(\vec{x},\vec{\theta}\right) = -\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta_i)^2 + \sum_{1\leq i<j\leq N}\frac{\gamma_{i,j}+\gamma_{j,i}}{2}x_i x_j, \tag{6}$$

where, for all $i \neq j$, $\gamma_{i,j} = \gamma_{j,i}$ and, for all $i$, $\sum_{j\neq i}|\gamma_{i,j}| < 1$, which guarantees that $u$ is strictly concave over $\vec{x}$. The decision maker's utility therefore has two components. The first captures how closely each of her decisions $x_i$ can track its "local fundamental," $\theta_i$. The second captures how her different decisions interact with each other. The scalar $\gamma_{i,j}$ then parametrizes the strength of the interaction from decision $j$ to decision $i$.

The optimal decision rule for each decision $i$ can be characterized by the decision specific

[12] For notation simplicity, I also set $\gamma_{i,i} = 0$ for all $i$.

[13] In the case that each fundamental has a negative direct effect on each decision, such as prices on consumption in the context of consumer theory studied below, one can define each $\theta_k$ as a negative multiple of the price $p_k$.

optimality condition in (3). Taking the first order condition of (3) and collecting terms, we have:

$$x_i^*(\omega_i) = E_i[\theta_i] + E_i\left[\sum_{j \neq i} \gamma_{i,j} x_j^*(\omega_j)\right] \quad \forall i, \omega_i \in \Omega_i, \tag{7}$$

(7) is then nested by the general decision rule in (5), with $\psi_{i,i} = 1$ and $\psi_{i,k} = 0$, for all $k \neq i$. Here, the direct effect of the fundamental $\vec{\theta}$ on each $x_i$ is summarized solely by its local fundamental $\theta_i$. If each self $i$, who is in charge of decision $i$, perfectly knows her own $\theta_i$, the direct effect in (7) is the same as in the frictionless case, and the friction under narrow thinking comes solely from each self's imperfect perception of other decisions, i.e. the frictional indirect effects. As further explained below, such an isolation will help me focus on the friction of interest.

**Consumer theory without income effects.** I now turn to the classical consumer problem of demanding multiple goods, part of the focus of Section 5. There, I study two cases. In the first, there are no income effects. The decision maker's utility is quasi-linear in the last good, and takes a CES form over the first $N$ goods. The interaction across decisions comes from the complementarity/substitutability embedded in the utility function, i.e. the second-order cross-derivatives of the utility function. As further explained in Section 5.1, the optimal consumption for each good $i \in \{1, \cdots, N\}$ is given by:

$$\hat{x}_i^*(\omega_i) = E_i[-\psi \hat{p}_i] + E_i\left[\gamma \sum_{j \neq i} \hat{x}_j^*(\omega_j)\right], \tag{8}$$

where a hat over a variable to denote its log-deviation. (8) is then nested by the general optimal decision rule in (17), with $\theta_i = -\hat{p}_i$, $\psi_{i,i} = \psi$, $\psi_{i,k} = 0$, and $\gamma_{i,j} = \gamma$, for all $j, k \neq i$. A positive (negative) $\gamma$ means that each pair of goods are complements (substitutes). Similar to the abstract example above, the direct effect of the fundamental on each $x_i$ is summarized by its price $p_i$.

**Consumer theory with income effects.** In the second case, the decision maker's utility is separable but there are income effects. The interaction across decisions then comes from the budget constraint. As further explained in Section 5.2, the optimal consumption for each good $i \in \{1, \cdots, N\}$ is given by:

$$\hat{x}_i^*(\omega_i) = E_i\left[-\psi_{i,i}\hat{p}_i - \sum_{k \neq i} \psi_{i,k}\hat{p}_k\right] + E_i\left[\sum_{j \neq i} \gamma_{i,j} \hat{x}_j^*(\omega_j)\right], \tag{9}$$

with $\psi_{i,i} > 0$, $\psi_{i,k} > 0$ and $\gamma_{i,j} < 0$, for all $j, k \neq i$. (9) is then nested by the general optimal decision rule in (5), with $\theta_i = -\hat{p}_i$. In this case, the existence of income effects introduces a direct effect of each price $p_k$ on each consumption $x_i$. Here, different selves' consumption decisions are strategic substitutes ($\gamma_{i,j} < 0$), due to the budget constraint.

**Information.** Motivated by the above examples and for sharper results, I will use the following information structure for the narrow thinker throughout the rest of the paper. Each self $i \in \{1, \cdots N\}$, who is in charge of decision $i$, receives a noisy signal about each $\theta_k \sim \mathcal{N}\left(\bar{\theta}_k, \sigma^2_{\theta_k}\right)$: $s_{i,k} = \theta_k + \epsilon_{i,k},\ \epsilon_{i,k} \sim \mathcal{N}\left(0, \sigma^2_{i,k}\right)$. That is, $\omega_i = \left\{\{s_{i,k}\}_{k\in\{1,\cdots,M\}}\right\}$. All fundamentals and noises are independent from each other.

Though not needed for the theoretical results in Subsection 3.2, in later applications, it is often natural to further let each self have perfect knowledge of one particular fundamental. For example, in the consumer theory examples in (8) and (9), it is natural that each self $i$, who buys good $i$, perfectly knows the price of the good she buys. That is, $\sigma^2_{i,i} = 0$. As explained soon in Subsection 3.3, by doing so, I can further isolate the friction of interest: when I study the impact of $\theta_i$ on $x_i$, the direct effect of $\theta_i$ on $x_i$ is maintained, and the friction fully comes from the indirect effect.

Finally, it is worth noting that the general definition of narrow thinking in Definition 1 allows for other potential information structures for the narrow thinker. However, for consistency, I will use this information structure throughout the rest of the paper. In section 4, I will further discuss the interpretation of this information structure, and compare it with other choices.

## 3.2 Effective Attenuation of Interaction

**Question of interest.** I now turn to the question of interest: how each narrow thinker's decision $x_i$ responds to shocks to each fundamental $\theta_k$. From (5), we know these responses depend on both the direct effects and the indirect effects. Naturally, the size of the direct effect of $\theta_k$ on $x_i$ is determined by the precision of self $i$'s signal about $\theta_k$, $s_{i,k}$. To understand the indirect effect of $\theta_k$ on $x_i$, we need to understand how each self $i$'s beliefs about how other decisions, i.e. $E_i\left[x^*_j\right]$, move with respect to $\theta_k$. Under narrow thinking, as each self has an imperfect perception of other selves' decisions, this type of beliefs will also be anchored.

**Belief anchoring.** Specifically, I use $E\left[\cdot|\theta_k\right]$ to denote the conditional expectation with respect to $\theta_k$. For $i \neq j$, $E\left[E_i\left[x^*_j\right]|\theta_k\right]$ captures how much self $i$'s belief about decision $j$ moves with shocks to $\theta_k$ and $E\left[x^*_j|\theta_k\right]$ captures how much decision $j$ itself moves with shocks to $\theta_k$. Proposition 2 below then shows that self $i$'s belief about decision $j$ moves less than decision $j$ itself in response to shocks to $\theta_k$, and is anchored towards the prior $E\left[x^*_j\right]$ (the unconditional mean of decision $j$).

**Proposition 2** *In response to $\theta_k$ shocks, each self $i$'s belief about other $x_j$, $E_i\left[x^*_j\right]$, is anchored:*

$$E\left[E_i\left[x^*_j\right]|\theta_k\right] = \lambda_{i,k} E\left[x^*_j|\theta_k\right] + \left(1-\lambda_{i,k}\right) E\left[x^*_j\right] \qquad \forall k,\ \forall i \neq j, \tag{10}$$

*where* $\lambda_{i,k} = \frac{\sigma^2_{\theta k}}{\sigma^2_{\theta_k}+\sigma^2_{i,k}} \in (0,1]$.

The degree of such anchoring is also determined by the precision of self $i$'s signal about $\theta_k$, $s_{i,k}$, and is parametrized by $\lambda_{i,k} = \frac{\sigma^2_{\theta k}}{\sigma^2_{\theta_k} + \sigma^2_{i,k}} \in (0, 1]$. Here, $s_{i,k}$ serves the role of an imperfect signal about how other decisions respond to $\theta_k$, which determines the size of the indirect effect of $\theta_k$ on $x_i$ through other $x_j$s. The noise in this signal then leads to the belief anchoring in (10).

**Narrow thinker's decision functions.** To facilitate the study of the narrow thinker's response to shocks to fundamentals, I define the narrow thinker's decision function as

$$x_i^{\text{Narrow}}\left(\vec{\theta}\right) \equiv E\left[x_i^*\left(\omega_i\right)|\vec{\theta}\right] \quad \forall i, \tag{11}$$

where $E\left[\cdot|\vec{\theta}\right]$ denotes the conditional expectation with respect to $\vec{\theta}$. By averaging over the realization of noises in signals, $x_i^{\text{Narrow}}\left(\vec{\theta}\right)$ captures the narrow thinker's decision as a function of fundamentals, and can be directly compared to

$$\left\{x_i^{\text{Standard}}\left(\vec{\theta}\right)\right\}_{i=1}^{N} = \arg\max_{\{x_i\}_{i=1}^{N}} u\left(x_1, \cdots, x_N, \vec{\theta}\right), \tag{12}$$

the standard, frictionless, decision function in which each decision is made with perfect knowledge of all $\vec{\theta}$. From the perspective of an econometrician who has data on fundamentals and decisions (but not different selves' signals), $x_i^{\text{Narrow}}\left(\vec{\theta}\right)$ defined in (11) is also the object of interest.

**Effective attenuation of interaction.** Now we study how the narrow thinker's decisions respond to shocks to each fundamental $\theta_k$, i.e. $\left\{\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k}\right\}_{i=1}^{N}$. For each decision $i$, the belief anchoring in (10) dampens the impact from other decisions $x_j$ to $x_i$ in response to the shock, and leads to an effective attenuation of interaction across decisions.

**Proposition 3** *In response to shocks to $\theta_k$, the narrow thinker's decisions can be characterized by*

$$\begin{pmatrix} \frac{\partial x_1^{\text{Narrow}}}{\partial \theta_k} & \frac{\partial x_2^{\text{Narrow}}}{\partial \theta_k} & \cdots & \frac{\partial x_N^{\text{Narrow}}}{\partial \theta_k} \end{pmatrix}' = \left(\mathbb{I}_N - \tilde{\Gamma}_k\right)^{-1} \Psi_k, \tag{13}$$

*where $\tilde{\Gamma}_k$ captures the effective interaction matrix and $\Psi_k$ captures the direct effect*

$$\tilde{\Gamma}_k = \begin{pmatrix} 1 & \lambda_{1,k} & \cdots & \lambda_{1,k} & \lambda_{1,k} \\ \lambda_{2,k} & 1 & \cdots & \lambda_{2,k} & \lambda_{2,k} \\ & & \cdots & & \\ \\ \lambda_{N,k} & \lambda_{N,k} & \cdots & \lambda_{N,k} & 1 \end{pmatrix} \circ \Gamma \quad \textit{and} \quad \Psi_k = \begin{pmatrix} \lambda_{1,k}\psi_{1,k} \\ \lambda_{2,k}\psi_{2,k} \\ \cdots \\ \\ \lambda_{N,k}\psi_{N,k} \end{pmatrix},$$

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

Proposition 3 is the main result of this subsection. 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 attenuated by a factor $\lambda_{i,k}$ between 0 and 1. That is, in response to shocks to $\theta_k$, because self $i$'s belief about decision $j$ is anchored, 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 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.

The vector $\Psi_k$ in (13) 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. The degree of the dampening is then also parametrized by $\lambda_{i,k}$, which depends on the signal-to-noise ratio of self $i$'s signal about $\theta_k$, $s_{i,k}$. To isolate the attenuation of interaction (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 abstract example in (6) and the consumer theory example without income effects in (8). Then, the direct effects of $\theta_k$ on decisions are maintained,[14] and the sole friction comes from the effective attenuation of interaction above and the accompanying dampening of indirect effects.

**Shock-specific attenuation and comparative statics.** One interesting feature is that the degree of the effective attenuation of interaction is allowed to be shock specific (the matrix $\tilde{\Gamma}_k$ depends on $k$). Thus, the analysis here allows the following possibility: if a fundamental $\theta_k$ is volatile, each self may pay "more attention" to $\theta_k$ and all selves' signal about $\theta_k$ will be more precise. Different selves then better coordinate in response to $\theta_k$, and there is less attenuation of interaction in response to $\theta_k$. In fact, in the context of the costly contemplation problem studied in Section 6 where the decision maker chooses optimally what information each decision is based upon, the above scenario arises endogenously in Proposition 12 in Appendix E.

### 3.3 Over- and Under-reaction

I then study how the belief anchoring and the attenuation of interaction above translate into concrete predictions about the narrow thinker's behavior. For example, I am interested in results about when the underlying friction leads to under-reaction, and when it leads to over-reaction.

As anticipated above when discussing the information structure, to isolate the friction of interest and get sharper results, I equate the number of fundamentals with the number of decisions ($M = N$), and impose that each self $i$ perfectly knows fundamental $\theta_i$ ($\sigma_{i,i}^2 = 0$). For instance, in the consumer theory examples in (8) and (9), it just means that each self $i$ perfectly knows the price

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

of the good she buys, $p_i$. In the abstract example in (6), it means that each self $i$ perfectly knows her local fundamental $\theta_i$. In fact, all the applications studied below will satisfy this restriction.

I can then study the "own-sensitivity" (e.g. own-price elasticity) and the "cross-sensitivity" (e.g. cross-price elasticity) for each decision. Specifically, for each self $i$, the own-sensitivity $\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_i}$ denotes how much $x_i$ responds to shocks to $\theta_i$, which she perfectly knows (I henceforth call $\theta_i$ as self $i$'s local fundamental). The cross-sensitivity $\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k}$ denotes how much $x_i$ responds to other fundamentals $\theta_k$, for $k \neq i$.

**Own-sensitivity.** We start from how each $x_i$ responds to shocks to its own local fundamental $\theta_i$, which she perfectly knows. From (5), we know such a response can be decomposed into a direct and an indirect effect.

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_i} = \underbrace{\psi_{i,i}}_{\text{Direct}} + \underbrace{\sum_{j \neq i} \gamma_{i,j} \frac{\partial E\left[E_i\left[x_j\right] | \vec{\theta}\right]}{\partial \theta_i}}_{\text{Indirect}}.$$

As each self $i$ knows her local fundamental $\theta_i$, the direct effect is maintained under narrow thinking. On the other hand, self $i$ has an imperfect understanding of other selves' decisions. The belief anchoring in Proposition 2 and the attenuation of interaction across decisions in Proposition 3 then dampen the indirect effect. 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 $x_i^{\text{Standard,Ind}}$ to denote the indirect effect on decision $i$ in (5), 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 $\frac{\partial x_i^{\text{Standard,Ind}}}{\partial \theta_i} > 0$. On the other hand, the indirect effect works in the opposition direction as the direct effect when $\frac{\partial x_i^{\text{Standard,Ind}}}{\partial \theta_i} < 0$.

**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* $i$,[15,16]

[15] To guarantee $I - \Gamma$ is invertible, we impose $\gamma \in \left(-1, \frac{1}{N-1}\right)$ in case 1), $\sum_{j \neq i} \gamma_{i,j} < 1$ $\forall i$ in case 2), and $\rho_i \Gamma_i < 1$ $\forall i$ in case 3).

[16] 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 4** *Suppose Assumption 1 holds,*

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

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_i} \leq \frac{\partial x_i^{\text{Standard}}}{\partial \theta_i} \quad \forall i;$$

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

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_i} \geq \frac{\partial x_i^{\text{Narrow}}}{\partial \theta_i} \quad \forall i.$$

The reason that additional conditions in Assumption 1 are required to establish Proposition 4 is because of the coexistence of opposing indirect effects. There could be some components of the indirect effect (e.g. through one decision $x_{j_1}$) that positively influence the optimal decision $i$ and there could be some components of the indirect effect (e.g. through another decision $x_{j_2}$) that negatively influence the optimal decision $i$. Dampening of each component (see Proposition 8 in Appendix B 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 own-sensitivities.

It is worth explaining the third condition in Assumption 1, "Substitutes with a single factor structure." The 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 (9). In fact, this condition is satisfied by many applications studied below, i.e. Corollaries 3 and 4. A main theme of the rest of the paper is then to show how, in concrete economic applications, narrow thinking can generate specific predictions about over- and under-reaction, depending on the directions of the direct and indirect effects.

**Cross-sensitivity.** I now turn to how $x_i$ responds to shocks to other fundamentals $\theta_k$, for $i \neq k$. For the abstract example in (6) and the consumer theory example without income effects in (8), $\theta_k$ has no direct effect on $x_i$. $x_i'$s response to $\theta_k$ is solely driven the indirect effect, i.e. changes in self $i$'s beliefs about other decisions. The dampening of indirect effects under narrow thinking then attenuates the cross-sensitivity. In the case that $\theta_k$ does have a direct effect on $x_i$ (e.g. the

---

any non-negative scalar.

consumer theory example with income effects in (9)), the dampening of this direct effect (because self $i$ may not perfectly know $\theta_k$) can also contribute to the attenuation of cross-sensitivity.

**Proposition 5** *If either 1) or 2) in Assumption 1 holds, the narrow thinker's cross-sensitivities are attenuated:*

$$\left|\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k}\right| \leq \left|\frac{\partial x_i^{\text{Standard}}}{\partial \theta_k}\right| \quad \forall i \neq k.$$

Similar to above, the attenuation of cross-sensitivity 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-sensitivity is dampened.

## 3.4 Discussion

**Comparison with rational inattention.** It is worth clarifying the difference between narrow thinking and rational inattention. When applying rational inattention to static multiple-decision problems (e.g. Koszegi and Matejka, 2018), different decisions are based on the same, though imperfect, information about the fundamentals. As a result, each decision is made with perfect knowledge of other decisions. 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. Specifically, in the environment considered here, the rationally inattentive decision maker's decision $i$ is given by $x_i^{\text{Standard}}\left(E\left[\vec{\theta}|\omega\right]\right)$, where $\omega$ is her imperfect, but common, signal. Narrow thinking, on the other hand, breaks such certainty equivalence: $x_i^*(\omega_i) \neq x_i^{\text{Standard}}\left(E\left[\vec{\theta}|\omega_i\right]\right)$.

The difference between two approaches is most sharply illustrated in the simple abstract example in (6). In this example, the direct effect of the fundamental on each decision $x_i$ is summarized by its local fundamental $\theta_i$. When each self $i$ of the narrow thinker perfectly knows its local fundamental $\theta_i$, the direct effect is the same as in the frictionless case, and the friction comes solely from each self's imperfect perception of other decisions and the dampening of indirect effects. On the other hand, under rational inattention and sparsity, the key friction is the decision maker's uncertainty about the fundamental, that is, the frictional direct effects. The decision maker nevertheless perfectly knows her other decisions when making a particular decision.

**Correlated fundamentals and rational confusion.** In the analysis above, I let different $\theta_k$s be uncorrelated. When different $\theta_k$s are correlated, an additional channel, "rational confusion," emerges: to the extent that $\theta_{k_1}$ and $\theta_{k_2}$ are correlated, each self can use signals about $\theta_{k_1}$ to forecast $\theta_{k_2}$ and vice versa. Given the interpretation that each self's imperfect knowledge about other selves' local fundamentals comes from cognitive frictions, one may not want to take into account such complicated rational confusion considerations. In fact, when different $\theta_k$s are correlated, this section's analysis can be interpreted as a characterization of the narrow thinker's behavior when such rational confusion is shut down. That is, each self $i$'s forecast about $\theta_k$ is based on her signal about $\theta_k$, $s_{i,k}$, solely. In this case, $E_i[\theta_k] = E[\theta_k | s_{i,k}]$. In Appendix, I also establish results about the behavior under narrow thinking allowing such rational confusion considerations, e.g. Proposition 9 in Appendix D.

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

**Proposition 6** *On average, each narrow thinker's decision coincides with the frictionless one:*

$$E\left[x_i^{\text{Narrow}}\right] = E\left[x_i^{\text{Standard}}\right] \quad \forall i, \tag{14}$$

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

**The boundary of a self and a potential experiment test.** In the above analysis, each self is in charge of one decision (e.g. purchasing one good). More generally, one should define the boundary of a self as a group of decisions made based on the same information. 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. Using the language of this paper, in the first case, different decisions are based on different information and made by different selves. Each decision is then made with an imperfect understanding of other decisions. In the second case, different decisions are based on the same information and made by the same self. Each decision is then made with perfect knowledge of other decisions. In sum, as Kahneman (2011) points out, "we tend to make decisions as problems arise."

This discussion also alludes to a potential experimental test of narrow thinking. One could start from a multiple decision problem, and compare two cases. In the first case, different decisions are made separately, potentially based on different information due to bounded recall and sequential retrieval from memory. In the second case, all decisions are made together, based on the same information. Narrow thinking then predicts that, only in the first case, the decision maker's behavior deviates from that of the frictionless benchmark.

# 4 A Bounded Recall Interpretation in Sequential Settings

The above analysis of the decision problem under narrow thinking is through the lens of a static, incomplete-information, game among multiple selves, as the multiple-decision problems studied in the paper are traditionally treated as static (e.g. demand of multiple goods). One may naturally wonder how to interpret the above analysis about the narrow thinker's behavior if the sequential nature of the decision-making process is modeled explicitly. This is the focus of this section. I first explain how to interpret the narrow thinker's information structure introduced above as a particular form of bounded recall (or selective retrieval from memory), and how the above analysis is observationally similar to other forms of bounded recall. I then discuss a particular feature of my approach: because of the equivalence to a static incomplete information game, the narrow thinker's behavior is insensitive to the exact order of the decisions. My approach is hence particularly useful when the analyst does not want to take a stand on the exact order of decisions. I finally compare the optimal behavior under narrow thinking with that under perfect recall.

**Interpreting the narrow thinker in a sequential setting.** To illustrate the main points clearly, in this section I will work with the symmetric version of the two-decision case of the abstract example in (7). That is, the optimal decision rule for each self $i \in \{1, 2\}$ is given by

$$x_i^* (\omega_i) = E_i \left[\theta_i + \gamma x_{-i}^* (\omega_{-i})\right], \tag{15}$$

where $\theta_i \sim \mathcal{N}\left(\bar{\theta}, \sigma_\theta^2\right)$ and $E_i [\cdot] = E [\cdot|\omega_i]$ captures self $i$'s belief when decision $i$ is made. For the narrow thinker studied in the previous section, the information of self $i \in \{1, 2\}$ is given by $\omega_i = \{\theta_i, s_{i,-i}\}$, where $s_{i,-i} = \theta_{-i} + \epsilon_{i,-i}$, $\epsilon_{i,-i} \sim \mathcal{N}(0, \sigma^2)$ and $\sigma^2$ captures the size of the noise in each self $i$'s signal about the other $\theta_{-i}$.[17] As discussed in (8), one can also interpret each decision $i$ in (15) as purchasing good $i$, the fundamental $\theta_i$ as a negative multiple of the price of good $i$, and the interaction $\gamma$ as the complementarity or substitutability between goods.

How do we interpret the above narrow thinker's information structure if one wants to take an explicit stand on the sequential nature of the decision-making process? For example, consider the case that decision 1 is made before decision 2. For the above narrow thinker, when her self 1 decides on $x_1$, she perfectly knows her local fundamental $\theta_1$, but only receives a noisy signal about the future $\theta_2$. When her self 2 decides on $x_1$, she perfectly knows her local fundamental $\theta_2$, 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 $\theta_1$ ($s_{2,1}$) and $\theta_2$. 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 fundamental (but not directly from a noisy signal

[17]Here and also for the rest of this section, all noises and fundamentals are independent from each other.

about her past endogenous decision).

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 $\omega_1 = \{\theta_1, s_{1,2} = \theta_2 + \epsilon_{1,2}\}$ and $\omega_2 = \{\theta_2, s_{2,1} = \theta_1 + \epsilon_{2,1}\}$. To characterize her optimal behavior, because of the equivalence to a static incomplete information game, the analyst does not need the additional information about the order of decisions. This property is particularly useful for many applications studied in the next section (e.g. demand of multiple goods), as the available data (e.g. demand data) typically do not involve the order of decisions.

**Observational equivalence with other forms of bounded recall.** A natural question is how different the above narrow thinker's behavior is from the behavior of 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 $\omega_1 = \left\{\theta_1, s_{1,2}^{\text{BR}} = \theta_2 + \epsilon_{1,2}^{\text{BR}}\right\}$, where $\epsilon_{1,2}^{\text{BR}} \sim \mathcal{N}\left(0, \sigma_{\text{BR}}^2\right)$. That is, she perfectly knows $\theta_1$, and receives a noisy signal about the future $\theta_2$, and $\sigma_{\text{BR}}^2$ captures the size of the noise. On the other hand, her self 2's information is given by $\omega_2 = \left\{\theta_2, s_{2,1}^{\text{BR}} = x_1 + \epsilon_{2,1}^{\text{BR}}\right\}$, where $\epsilon_{2,1}^{\text{BR}} \sim \mathcal{N}\left(0, \Sigma_{\text{BR}}^2\right)$. That is, she perfectly knows $\theta_2$, but cannot perfectly recall her past decision $x_1$. Instead, she receives a noisy signal directly about $x_1$ and $\Sigma_{\text{BR}}^2$ captures the size of the noise. Similarly, for the case that the decision 2 is made before the decision 1, we have $\omega_2 = \left\{\theta_2, s_{2,1}^{\text{BR}} = \theta_1 + \epsilon_{2,1}^{\text{BR}}\right\}$ and $\omega_1 = \left\{\theta_1, s_{1,2}^{\text{BR}} = x_2 + \epsilon_{1,2}^{\text{BR}}\right\}$, where $\epsilon_{2,1}^{\text{BR}} \sim \mathcal{N}\left(0, \sigma_{\text{BR}}^2\right)$ and $\epsilon_{1,2}^{\text{BR}} \sim \mathcal{N}\left(0, \Sigma_{\text{BR}}^2\right)$.

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 ($x_i^{\text{Narrow}}\left(\vec{\theta}\right)$ in (11)), we define $x_i^{\text{BR}}\left(\vec{\theta}\right) \equiv E\left[x_i^*\left(\omega_i\right)|\vec{\theta}\right]$, where $E\left[\cdot|\vec{\theta}\right]$ averages over not only the realization of noises in signals but also the potential order of decisions.[18]

In fact, without additional knowledge about the order of decisions, the own- and cross- sensitivities of the narrow thinker and those of the decision maker (BR) are observationally equivalent.[19]

**Lemma 3** *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- and cross- sensitivities:*

$$\frac{\partial x_i^{Narrow}}{\partial \theta_i} = \frac{\partial x_i^{BR}}{\partial \theta_i} \quad and \quad \frac{\partial x_i^{Narrow}}{\partial \theta_{-i}} = \frac{\partial x_i^{BR}}{\partial \theta_{-i}} \quad \forall i \in \{1, 2\}. \tag{16}$$

[18] 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.

[19] 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 3 and beyond.

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 receives noisy signals about the first $N-1$ decisions: $s_{N,i}^{BR} = x_i + \epsilon_{N,i}^{BR}$, $i \in \{1, \cdots, N-1\}$. As each decision is a function of all fundamentals $(\theta_1, \cdots, \theta_N)$, 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 5.

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.

**Comparison with standard sequential decisions.** In Appendix B, I also 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. 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: as the later self frictionlessly reacts to the earlier decision, the earlier self's decision can be characterized by the standard, frictionless, decision function. For example, if decision 1 comes first, we have $x_1^*(\omega_1) = x_1^{\text{Standard}}(\theta_1, E_1[\theta_2])$. 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 behavior under perfect recall, driven by uncertainty about fundamentals alone, is limited. In Lemma 4 in Appendix B, I show that the own- and cross- sensitivities under perfect recall deviate less from their frictionless counterparts than those under narrow thinking (averaged over the realization of the potential order of decisions).

**The road ahead.** 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. It is particularly useful when the analyst does not want to take a stand on this exact order. In the next section, we will see how the approach can deliver concrete predictions in specific economic applications.

# 5 Applications

In this Section, I study the implications of narrow thinking for concrete economic applications. Based on Proposition 4 above, I lay out narrow thinking's predictions about whether the decision maker over- or under-reacts in a number of specific economic environments. A general rule is: when the indirect effect works in the same direction as the direct effect, a dampening of the indirect effect under narrow thinking often 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 often leads to over-reaction. The analysis then also contrasts with the often-held belief that noises in the decision maker's mental representation of the world typically lead to under-reaction.

Specifically, I start from the classical consumer problem of demanding multiple goods under narrow thinking. To tease out the mechanism, I solve two polar cases. In the first, as (8) above, the decision maker's utility is quasi-linear and there are no income effects. In this case, the interaction across different consumption decisions comes from the complementarity/substitutability embedded in the utility function (i.e. the cross-derivatives of the utility function). In the second, as (9) above, the decision maker's utility is separable but there are income effects. The interaction across different decisions now comes from the budget constraint. Using the general principle above, I will explain why narrow thinking leads to excess smoothness to own price shocks in the first case, but excess sensitivity in the second case. The second case is also of independent interest: as different selves of the narrow thinker have different beliefs about the marginal value of money, the fungibility principle is violated. She behaves as if the money allocated to one good cannot perfectly substitute for the money allocated to another. I will further explain how this violation leads to a "smooth" model of mental accounting under narrow thinking.

Finally, I further show how narrow thinking can provide a unified framework for other empirical examples of under-reaction (e.g. excess smoothness to taste shocks and small wage elasticity of daily labor supply) and over-reaction (e.g. the label effects and excess sensitivity to income shocks).

## 5.1 Demand Elasticities under Narrow Thinking

**Set up**. I first set up a general consumer theory problem under narrow thinking. The decision maker's utility depends on her consumption of $N$ goods, $(x_1, \cdots, x_N)$, and the numeraire $y$ (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).[20] Along with the discussion about constraint problems in Subsection 2.3, I always let the last decision, $y$, be made with perfect knowledge about prices and other decisions. Note that I let $\tilde{u}$ be well defined for all $y \in \mathbb{R}$. This allows the possibility that the "residual decision" $y$ is negative and guarantees that one can find a $y$ such that the budget constraint is satisfied.

As the budget constraint always binds in the optimum, one can use it to substitute $y$:

$$u(x_1, \cdots, x_N, \vec{p}) = \tilde{u}\left(x_1, \cdots, x_N, w - \sum_{i=1}^{N} p_i x_i\right). \tag{17}$$

This is then nested in the unconstrained problem in (2), with $\vec{\theta} = \vec{p}$.

**The case without income effects.** In this and the next subsections, I then focus on two polar cases to tease out the mechanism. I refer the reader to Appendix C for the general case. Here, the decision maker's utility takes a common form, namely CES with quasi-linear utility. Specifically, her utility in (17) is given by

$$u(x_1, \cdots, x_N, \vec{p}) = \frac{\left(\sum_{i=1}^{N} \frac{1}{N} x_i^{\frac{\varrho-1}{\varrho}}\right)^{\frac{\varrho(1-\kappa)}{\varrho-1}}}{1-\kappa} + w - \sum_{i=1}^{N} p_i x_i, \tag{18}$$

where $\rho > 0$ captures how substitutable each pair of goods is, and $\kappa > 0$ captures the rate at which the marginal value of consumption moves with respect to the "total consumption," $X = \left(\sum_{i=1}^{N} \frac{1}{N} x_i^{\frac{\varrho-1}{\varrho}}\right)^{\frac{\varrho}{\varrho-1}}$. The utility function in (18) is strictly concave over $\vec{x}$.[21]

**Information.** Following the information structure considered in Section 3, it is natural to consider the following narrow thinker here: when self $i$ decides on the consumption $x_i$, she perfectly knows its price $p_i$, but only receives noisy signals about other prices. The information structure is also consistent with the "What You See is All There Is" principle raised in Kahneman (2011).

[20] For notation simplicity, I normalize the price of the last good $y$ is normalized to 1. In fact, as long as its price is common knowledge across different selves, this normalization is without loss of generality.

[21] In the current CES context, I also let $\bar{x}_i$ be equal across $i$. This will lead to the symmetric optimal consumption decision rules in (19). This scenario arises when $\bar{p}_i = \bar{p}_j$ for each $i \neq j$.

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 $i \in \{1, \cdots, N\}$ of the narrow thinker, who is in charge of purchasing good $i$, perfectly knows $p_i \sim \log \mathcal{N}\left(\log \bar{p}_i, \sigma^2_{p_i}\right)$, but receives a noisy signal about each of the other $p_k$: $s_{i,k} = p_k \epsilon_{i,k}$, with $\epsilon_{i,k} \sim \log \mathcal{N}\left(0, \sigma^2_{i,k}\right)$ and $\sigma^2_{i,k} > 0$. All $\epsilon$s and $p$s are independent from each other.

**Log-linearization.** As I am mostly interested in the response to small temporary price shocks, I will work with log-linearized optimal decision rules throughout. This approximation allows me to analytically characterize the narrow thinker's behavior and applies the general result in Section 3. Such an approximation is standard in the applied literature and provides simple and interpretable formula for commonly used utility functions (e.g. CES and CRRA utility). 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: $\{\bar{x}_i\}_{i=1}^N = \arg\max_{\{x_i\}_{i=1}^N} u\left(x_1, \cdots, x_N, \bar{p}_1, \cdots \bar{p}_N\right)$. I then use a hat over a variable to denote its log-deviation from this point, e.g. $\hat{x}_i = \log \frac{x_i}{\bar{x}_i}$.

**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 (18) and taking first order condition of (3), we have, for each $i$, $E_i\left[\frac{\partial u}{\partial x_i}\left(x_i^*\left(\omega_i\right), \vec{x}^*_{-i}\right)\right] = E_i\left[\frac{1}{N} x_i^*\left(\omega_i\right)^{-\frac{1}{\varrho}} X^{\frac{1}{\varrho}-\kappa}\right] = p_i$. I then log-linearize the previous expression and arrive at the optimal consumption rules in (8) above, i.e.

$$\hat{x}_i^*\left(\omega_i\right) = E_i\left[BR_i\left(\vec{x}^*_{-i}; \vec{p}\right)\right] \equiv -\psi \hat{p}_i + \sum_{j \neq i} \gamma E_i\left[\hat{x}_j^*\left(\omega_j\right)\right], \tag{19}$$

where $\psi = \frac{1}{\frac{1-\frac{1}{N}}{\varrho} + \frac{\kappa}{N}} > 0$ and $\gamma = \frac{\left(\frac{1}{\varrho} - \kappa\right)\frac{1}{N}}{\frac{1-\frac{1}{N}}{\varrho} + \frac{\kappa}{N}} \in \left(-1, \frac{1}{N-1}\right)$.

The first term in (19) captures the direct effect of price changes on $x_i$, that is, the effect of $\vec{p}$ holding other decisions fixed. $\psi$ then parametrizes the size of such an effect. In the quasi-linear environment here, such a direct effect on $x_i$ depends solely on $p_i$. 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 (19) captures the indirect effect on $x_i$, that is, other consumption decisions' impact on $x_i$. A positive (negative) $\gamma$ means that each pair of goods are complements (substitutes),[22] and that the optimal consumption of good $i$ increases (decreases) with self $i$'s belief about each of the other consumption $x_j$.

**Narrow thinker's demand.** The main question of interest is how the narrow thinker's consumption responds to price changes. Similar to (11), for each $i$, I define the narrow thinker's

[22] In the quasi-linear context here, there is no difference between gross and net complements (substitutes).

(log) demand as a function of (log) prices:

$$\hat{x}_i^{\text{Narrow}}(\hat{p}_1, \cdots, \hat{p}_N) \equiv E\left[\hat{x}_i^*(\omega_i) | \hat{p}_1, \cdots, \hat{p}_N\right] \quad \forall i, \tag{20}$$

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

**Own-price demand elasticities.** I now translate the general insights about behavioral under narrow thinking studied in Section 3 into predictions about own- and the cross-price elasticities in the current context. We start from own-price demand elasticities. In fact, in the current example, the indirect effect of $p_i$ on $x_i$ is always in the same direction as the direct effect. As Proposition 4 above shows, narrow thinking then attenuates the own-price demand elasticities.[23]

**Corollary 1** *For each $i$, the narrow thinker's consumption $x_i$ decreases (increases) less in response to positive (negative) shocks to $p_i$ :*

$$\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} \le \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} < 0 \quad \forall i. \tag{21}$$

*Moreover, the inequality is strict when $\gamma \neq 0$ (when the indirect effect is non-zero).*

To understand (21), let us first consider the complements case with $\gamma > 0$. From (19), we know an increase in $p_i$ will have a negative direct effect on $x_i$. As different goods are complements, the consumption of other goods $x_j$ will also decrease. Such a decrease will further decrease $x_i$, generating a negative indirect effect on $x_i$. Under narrow thinking, the indirect effect is dampened, and $x_i$ 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 $x_j$ will now increase. Such an increase will then further decrease $x_i$, again generating a negative indirect effect on $x_i$. Under narrow thinking, the indirect effect 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 is always in the same direction as the direct effect. The dampening of the indirect effect then leads to excess smoothness, i.e. the attenuation of own-price elasticities in Corollary 1.

**Cross-price demand elasticities.** From condition (19), we know that cross-price demand elasticities are driven solely by changes of each self's beliefs about the others' decisions, i.e. indirect

[23] When $\gamma > 0$, the current environment falls into case 2) in Assumption 1. When $\gamma < 0$, the current environment falls into case 3) in Assumption 1, with $\rho_i = 1$, $\Gamma_i = \gamma$ and $\mathit{\Delta}_i = 0$.

effects. As in Proposition 5 about cross-sensitivity, the dampening of indirect effects leads to an attenuation of cross-price demand elasticities under narrow thinking.

**Corollary 2** *All of the narrow thinker's cross-price demand elasticities are attenuated:*

$$\left|\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_k}\right| \leq \left|\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_k}\right| \quad \forall i \neq k,$$

*when at least one of the following two conditions hold: 1) Symmetric information, i.e.* $\lambda_{i,k} \equiv \frac{\sigma_{p_k}^2}{\sigma_{p_k}^2+\sigma_{i,k}^2} = \lambda$ *for all* $i \neq k$*. 2) Complements, i.e.* $\gamma > 0$*.*

**Slutsky asymmetry.** Under narrow thinking, the Slutsky matrix can be asymmetric. For example, when $N = 2$, 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.[24]

**Testable predictions.** The above discussion also points out testable differences between the narrow thinker's demand and the demand under standard consumer theory. First, the Slutsky matrix under narrow thinking can be asymmetric. Second, in the same spirit of Proposition 6 in Section 3, the narrow thinker's frictional behavior studied above is about the response to price shocks. The narrow thinker's demand elasticity estimated based on temporary price shocks can then differ from the one estimated based on persistent price differences. This is different from the standard consumer theory. Appendix C provides further discussion along this line.

## 5.2 Violation of the Fungibility Principle and a "Smooth" Model of Mental Accounting

I now turn to the second case with income effects but separable utility, polar to the first. The interaction across different decisions now comes from the budget constraint. In this context, as different selves have different information, they have different beliefs about the marginal value of money. The narrow thinker then violates the fungibility principle: she behaves as if the money allocated to one good cannot perfectly substitute for the money allocated to another. Such a violation then generates mental accounting-type behavior which has "no agreed-upon model" (Farhi and Gabaix, 2015),[25] e.g. excess sensitivity to own-price changes (Hastings and Shapiro, 2013).[26]

[24] 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 $i, k$, $\frac{\partial x_i^{\text{Narrow}}}{\partial p_k} = \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} \frac{\bar{x}_i}{\bar{p}_k}$.

[25] By mental accounting-type behavior, I mean behavior akin to that generated by an explicit mental budget as in Heath and Soll (1996). An explicit mental budget means, for example, that the consumer allocates exactly 100 dollars to food spending. The discussion after Corollary 3 will elaborate further.

[26] Hastings and Shapiro (2013) find that, when gasoline prices rise, consumers substitute to lower octane gasoline, to an extent that cannot be explained by neoclassical effects.

To illustrate how narrow thinking can generate such excess sensitivity, consider an increase in 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 response of other consumption is limited, the indirect effects from smoothing out other consumption are dampened, and the food consumption will decrease more. By providing a model of the violation of the fungibility principle without relying on the explicit mental budgeting model in Heath and Soll (1996), the narrow thinking approach to mental accounting also generates new testable predictions.

**Environment.** In the consumer problem set up in (17), I let the consumer's utility be

$$\tilde{u}\left(x_1, \cdots, x_N, y\right) = \sum_{i=1}^{N} v_i\left(x_i\right) + h\left(y\right). \tag{22}$$

In (22), $v_i\left(x_i\right) = \frac{x_i^{1-\kappa_i}}{1-\kappa_i}$ captures the consumer's utility from consuming good $i$, where $\kappa_i > 0$ 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.[27] $h\left(y\right)$, a strictly concave function on $\mathbb{R}$,[28] captures the consumer's utility from the last decision, which can be interpreted as utility from saving or the value of money. The decision maker is subject to the budget constraint: $\sum_{i=1}^{N} p_i x_i + y \leq w$, where $p_i$ is the price of the good $i$ and $w$ is her total wealth. As discussed above, I always let the last decision, $y$, be made with perfect knowledge about all the fundamentals and other decisions, which guarantees that the budget constraint is satisfied. Also same as above, each self $i \in \{1, \cdots, N\}$ of the narrow thinker, who decides on the consumption $x_i$, perfectly knows its price $p_i$, but only receives noisy signals about other prices. The distributions of prices and signals are as specified above.

**Optimal decisions.** The optimal decision for each self $i \in \{1, \cdots, N\}$ must satisfy

$$\underbrace{v_i^{'}\left(x_i^*\left(\omega_i\right)\right)}_{\text{marginal utility of consuming good i}} = p_i E_i\left[\underbrace{h'\left(y^*\right)}_{\text{marginal value of monery}}\right]. \tag{23}$$

Condition (23) means that, from each self $i$'s perfective, 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 is made based on perfect knowledge, and the standard perturbation argument holds between the consumption $x_i$ and $y$.

[27]Holding the spending share for good $i$ fixed, a higher $\kappa_i$ means a less elastic Marshallian demand for good $i$ (under both the standard consumer theory and narrow thinking).

[28]This allows the possibility that the "residual decision" $y$ is negative and guarantees that the budget constraint will always be satisfied.

I then log-linearize the above condition and the budget constraint, and use a hat over a variable to denote its log-deviation. The optimal decision rule (23) for each $i$ then becomes

$$-\kappa_i \hat{x}_i^* (\omega_i) = \hat{p}_i - \kappa_h E_i [\hat{y}^*], \tag{24}$$

where $-\kappa_i \hat{x}_i^* (\omega_i)$ captures the marginal utility of consuming good $i$, $-\kappa_h E_i [\hat{y}^*]$ captures self $i$'s belief about the marginal value of money, and $\kappa_h = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$ captures the rate at which the marginal value of money moves with respect to $y$ (at the point of log-linearization).

The log-linearized budget constraint is given by

$$\sum_{i=1}^{N} \mu_i (\hat{x}_i^* (\omega_i) + \hat{p}_i) + \mu_y \hat{y}^* = 0, \tag{25}$$

where $\mu_i = \frac{\bar{p}_i \bar{x}_i}{w}$ and $\mu_y = \frac{\bar{y}}{w}$ are the spending share of good $i$ and $y$ at the point of log-linearization.

To see how the income effects emerge in the current environment, note that an shock to any $\hat{p}_k$ directly influences the budget constraint (25) and thus $\hat{y}^*$. As a result, each self $i$'s belief about the marginal value of money, $-\kappa_h E_i [\hat{y}^*]$, will also change. Such change will then generate a direct effect of $\hat{p}_k$ on each consumption $\hat{x}_i$. This channel is muted in the quasi-linear case above, as the marginal value of money is a constant there. Substituting (25) into (24), we then arrive at the optimal consumption rules in (9) above.[29]

**Violation of the fungibility principle.** From (24), one can see that the fungibility principle is violated under narrow thinking. As different selves have different information, they hold different beliefs about the marginal value of money, $-\kappa_h E_i [\hat{y}^*]$. Then, the marginal value of spending an additional unit of money on each good $i$,$-\kappa_i \hat{x}_i^* (\omega_i) - \hat{p}_i = -\kappa_h E_i [\hat{y}^*]$, then also differs.

**Excess sensitivity to own-price changes.** I then formalize how narrow thinking can generate mental accounting-type behavior, i.e. excess sensitivity to own-price changes.[30] Similar to condition (20), for each $i$, I define the narrow thinker's (log) demand function as $\hat{x}_i^{\text{Narrow}} (\hat{p}_1, \cdots, \hat{p}_N) \equiv E [\hat{x}_i^* (\omega_i) | \hat{p}_1, \cdots, \hat{p}_N]$, averaging over the realization of noises in signals.

**Corollary 3** *For each good $i$ such that $\kappa_i > 1$, the narrow thinker's consumption $x_i$ decreases*

[29] In (9), we have $\psi_{i,i} = \frac{\frac{1}{\kappa_i} + \frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}{1+\frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}$, $\psi_{i,k} = \frac{\frac{\kappa_h}{\kappa_i}\frac{\mu_k}{\mu_y}}{1+\frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}$ and $\gamma_{i,j} = -\frac{\frac{\kappa_h}{\kappa_i}\frac{\mu_j}{\mu_y}}{1+\frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}$, for all $j, k \neq i$. The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\frac{\kappa_h}{\kappa_i}}{1+\frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}$ and $\Gamma_i = \Delta_i = \frac{\mu_i}{\mu_y}$. As a result, the Corollary 3 directly follows.

[30] Similar to Corollary 2, in the current environment, one can also establish that all of the narrow thinker's cross-price demand elasticities are attenuated under narrow thinking when $\lambda_{i,k} = \lambda$ for all $i \neq k$.

*(increases) more in response to positive (negative) shocks to* $p_i$ :

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} < \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} < 0.$$

To see the mechanism behind the excess sensitivity, note when $\kappa_i > 1$, 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). Through the decrease of $x_j$, the indirect effect of an increase in $p_i$ on $x_i$ 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, as Proposition 4 above, a dampening of the indirect effect then leads to over-reaction in this environment. This contrasts with the case in Corollary 1, where the indirect effect of an increase in $p_i$ on $x_i$ is negative and works in the same direction as the direct effect. As a result, a dampening of the indirect effect leads to under-reaction there.

**A "smooth" model of mental accounting and other implications.** First, the excess sensitivity in Corollary 3 does not require the consumers to have an explicit mental budget as in Heath and Soll (1996). An explicit mental budget means, for example, that the consumer allocates exactly 100 dollars to food spending. Narrow thinking nevertheless brings the consumer's demand elasticity closer to the case of explicit mental budgeting. That is, if we use $\frac{\partial \hat{x}_i^{\text{Explicit}}}{\partial \hat{p}_i}$ to denote the own-price demand elasticity in the explicit mental budgeting, we have $\frac{\partial \hat{x}_i^{\text{Explicit}}}{\partial \hat{p}_i} < \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} < \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i}$.[31] In this sense, the narrow thinker exhibits mental accounting-*type* behavior, and the narrow thinking approach can be viewed as a "smooth version" of the explicit mental budgeting model. The degree of the frictional behavior under narrow thinking is then summarized by the signal-to-noise ratio of each self's signals about other prices, i.e. $\left\{\lambda_{i,k} \equiv \frac{\sigma_{p_k}^2}{\sigma_{p_k}^2 + \sigma_{i,k}^2}\right\}_{i,k=1}^{N}$.

Second, when $\kappa_i < 1$, instead, the narrow thinker's consumption $x_i$ decreases less in response to an increase in $p_i$. That is, for all $i$, $\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} < \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i}$.[32] 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 $x_j$ after an increase in $p_i$ will then further decrease $x_i$. This scenario falls into the case that 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. 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

[31]See the proof of Corollary 3 for a proof of this statement.

[32]In fact, in this case, it is possible that $\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} > 0$.

possible that the consumer's demand with respect to gasoline is less elastic (has a higher $\kappa_i$). As a result, the empirical finding is line with the prediction in Corollary 3.

Finally, the frictional behavior under narrow thinking is about the response to temporary price shocks. The average allocation of funds across different goods under narrow thinking, nevertheless, may still be consistent with standard consumer theory. This is in line with Proposition 6 studied above. Such differential predictions in response to shocks versus on average are a unique testable prediction under narrow thinking. The difference also sheds light on a key issue: when the decision maker has a narrow bracket and when she has a broad bracket. The narrow thinker has a narrow bracket, that is, she neglects the interaction across decisions, in response to shocks. On the other hand, she has a broad bracket, that is, she takes into account the impact of other decisions on each decision, on average.

## 5.3 Other Under-reaction Examples

**Excess smoothness to taste shocks.** One often mentioned form of under-reaction is excess smoothness to taste shocks, which is also connected to mental accounting. 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 illustrate how narrow thinking can generate excess smoothness to taste shocks, consider a positive taste shock to food. Under standard consumer theory, the decision maker can coordinate all her decisions by decreasing other consumption to help increase food consumption, in response to the positive taste shock. Under narrow thinking, however, other consumption will not decrease as much, the indirect effect from such a coordinated response is dampened, and food consumption will increase less.

To formalize, consider an environment similar to the one in Section 5.2 (consumer theory with income effects). The decision maker's utility is given by

$$\sum_{i=1}^{N} \varphi_i v_i\left(x_i\right) + h\left(y\right), \tag{26}$$

where $v_i\left(x\right)$ and $h\left(y\right)$ are defined as in Section 5.2. Here I introduce taste shocks, and $\varphi_i \sim \log\mathcal{N}\left(\log\bar{\varphi}_i, \sigma^2_{\varphi_i}\right)$ parametrizes the taste for good $i$. The decision maker needs to satisfy the budget constraint, $\sum_{i=1}^{N} p_i x_i + y \leq w$. As above, I let the last decision $y$ be made with perfect knowledge about all the fundamentals and other decisions, which guarantees that the budget constraint is satisfied. As I am interested in response to taste shocks, I treat $w$ and $p$s as constants.

Similar to the case 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 the taste $\varphi_i$, but receives a noisy signal about each of the other $\varphi_k$: $s_{i,k} = \varphi_k \epsilon_{i,k}$, with $\epsilon_{i,k} \sim \log \mathcal{N}\left(0, \sigma_{i,k}^2\right)$ and $\sigma_{i,k}^2 > 0$. All $\epsilon$s and $\varphi$s are independent from each other.

I use a hat over a variable to denote its log-deviation from the point of log-linearization. Similar to conditions (11) and (20), for all $i$, I define the narrow thinker's (log) demand function as $\hat{x}_i^{\text{Narrow}}\left(\hat{\varphi}_1, \cdots, \hat{\varphi}_N\right) \equiv E\left[\hat{x}_i^*\left(\omega_i\right) | \hat{\varphi}_1, \cdots, \hat{\varphi}_N\right]$. Compared to the standard case when each decision is made with perfect knowledge of all fundamentals (indexed by the superscript *Standard*, as above), one can then establish excess smoothness to taste shocks under narrow thinking.

**Corollary 4** *For each good $i$, the narrow thinker's consumption $x_i$ increases (decreases) less in response to positive (negative) taste shocks to $\varphi_i$ :*

$$\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{\varphi}_i} > \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\varphi}_i} > 0. \tag{27}$$

To see the mechanism behind excess smoothness, note an increase in taste $\varphi_i$ always increases $x_i$ (a positive direct effect) and decreases the consumption of other goods $x_j$. The 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. In this case, as Proposition 4 shows, the dampening of the indirect effect under narrow thinking leads to under-reaction. Specifically, under narrow thinking, other $x_j$ decreases less, and the consumer increases $x_i$ less. As a result, the narrow thinker's consumption exhibits excess smoothness to taste shocks.[33]

**Comfort zones.** Now I turn to the context of time management and offer a novel theory of "comfort zones:" the narrow thinker under-reacts to shocks to the attractiveness of each of her activity. For a concrete example of the above comfort zones behavior, consider an engineering student who decides on how much time she will spend on an economics class. She realizes the economics class has a great professor, that is, the $\varphi_i$ of spending time on the economics class is high. She should not only spend more time on economics (a positive direct effect), but also spend less time on engineering classes, which helps her further increase the amount of time spent on economics (a positive indirect effect). Under narrow thinking, however, she is concerned that she will not decrease the amount of time spent on engineering classes as much. The dampening of indirect effects then prevents the engineering student from spending significantly more time on economics and leads the engineering student to stay within her comfort zone. Corollary 7 in Appendix D provides a formalization of this comfort zone behavior.

[33] Interestingly, the excess smoothness to taste shocks to $\varphi_i$ does not require that the decision maker's utility with respect to the good $i$ has a high $\kappa_i$ (recall $v_i\left(x_i\right) = \frac{x_i^{1-\kappa_i}}{1-\kappa_i}$). This is because an increase in $\varphi_i$ always decreases the consumption of other goods $x_j$.

**The small wage elasticity of daily labor supply.** Another well documented empirical example of under-reaction is the small wage elasticity of daily labor supply (Camerer et al., 1997, Crawford and Meng, 2011, Farber, 2015, Thakral and To, 2017). In the standard labor supply theory, when the wage on a particular day increases, the decision maker should coordinate her behavior by increasing her labor supply on the day of wage increase (a positive direct effect) and decreasing her labor supply on other days. Such a coordinated response then further increases the elasticity of daily labor supply (a positive indirect effect). Under narrow thinking, however, labor supply on other days will not be as responsive, the accompanying dampening of indirect effects will prevent a large increase in labor supply on the day of the wage increase. Corollary 8 in Appendix D formalizes how narrow thinking generates a small wage elasticity of daily labor supply.

A few additional predictions of the narrow thinking approach emerge. First, as the narrow thinking approach does not require the decision maker to have an explicit daily income target, the narrow thinker's behavior can be consistent with the empirically documented positive, but small, wage elasticity of daily labor supply. This avoids the difficulty raised by Farber (2015), who points out that the income targeting model and the reference dependence model often predict a negative wage elasticity of daily labor supply, inconsistent with the empirical evidence. Second, in line with Proposition 6, 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. Third, the narrow thinking approach's prediction is consistent the finding in Camerer et al. (1997) and Farber (2015) that wage elasticity of daily labor supply increases with the taxi driver's experience. More experience is akin to an increase in cognitive capacity $\tau$ in Section 6, helping the decision maker to coordinate her multiple selves.

## 5.4 Other Over-reaction Examples

**Excess sensitivity to temporary income shocks.** One often mentioned form of over-reaction is excess sensitivity to temporary income shocks. As Stephens Jr and Unayama (2011), Parker (2017) and Kueng (2018) document, such excess sensitivity cannot be fully explained by the existence of liquidity constraints. As an example of such behavior, Thaler (1999) mentioned his own experience after earning a speaking fee for a conference in Switzerland. He spent excessively on hotels and meals for an additional week of vacation in Switzerland. He said he would not spend so much on the vacation without the speaking fee. Such behavior is inconsistent with the standard consumption smoothing behavior: after the positive income shock in Switzerland, the decision maker should not only increase her current consumption in Switzerland (a positive direct effect), but also increase consumption at other points in time. Such increase in other consumption then prevents a large increase in her current consumption (a negative indirect effect). Under narrow thinking, however, consumption at other points in time may not be as responsive to current temporary income shocks. As a result, the narrow thinker increases her current consumption more. This scenario falls into the case that the indirect effect works in the opposite direction to the direct effect. Corollary 9 in Appendix D formalizes how narrow thinking leads to over-reaction.

**The label effect.** A similar mechanism can also explain another form of over-reaction, "the label effect:" consumption decisions can be sensitive to the label attached to the consumer's budget (Beatty et al., 2014, Benhassine et al., 2015, Abeler and Marklein, 2016, Hastings and Shapiro, 2017). For example, Beatty et al. (2014) study the UK Winter Fuel Payment program. Despite its label, the program 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 much more after receiving the Winter Fuel Payment than after receiving cash. Such behavior is inconsistent with standard consumer theory: after the Winter Fuel Payment, the decision maker should not only increase her fuel consumption (a positive direct effect), but also other consumption. Such increase in other consumption then prevents a large increase in her fuel consumption (a negative indirect effect). Under narrow thinking, however, other consumption may not be as responsive, and the decision maker increases her fuel consumption more. Corollary 10 in Appendix D provides a formalization.

# 6 Endogenous Narrow Thinking: Costly Contemplation

The previous analysis lets different decisions be made based on different, but exogenous, information. In this section, I try to endogenize such information, in a "costly contemplation" problem.[34] 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. As a result, it is optimal for different decisions' signals to take different forms. The analysis also provides a framework to study the optimal information choice problem at the *decision*-level, going beyond the standard rational inattention paradigm.

## 6.1 Set up

In this section, for notation clarity, I use bold letters to denote random variables and normal letters to denote their realizations. Let $(S, \mathcal{F}, P)$ be the underlying probability space. The decision

[34] As discussed in Section 2.3, there are multiple cognitive frictions justifying narrow thinking. The analysis in this section is complementary to other justifications discussed there.

maker's utility is given by $u\left(\vec{x},\vec{\theta}\right)$, where $u$ is twice continuously differentiable and strictly concave over $x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ and $\mathcal{X}_i$, a convex set on $\mathbb{R}$, denotes the set of possible decision $x_i$. The payoff relevant fundamental, $\vec{\theta}$, is the realization of an exogenously drawn random vector $\boldsymbol{\vec{\theta}} : S \to \Theta$, where $\Theta \subseteq \mathbb{R}^M$ denotes the set of possible fundamental.

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 $\boldsymbol{\omega_i} : S \to \Omega_i$, where $\Omega_i$ denotes the set of possible signal realizations for decision $i$. Different from Section 2, $\boldsymbol{\omega_i}$, which summarizes how decision $i$'s signal is generated, is chosen endogenously from a set of random vectors $\boldsymbol{\Omega_i}$.

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

$$\max_{\{\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}, x_i(\cdot)\}_{i=1}^N} E\left[u\left(x_1(\omega_1), \cdots, x_N(\omega_N), \vec{\theta}\right)\right] \tag{28}$$

$$s.t. \quad \sum_{i=1}^N I\left(\boldsymbol{\omega_i}; \boldsymbol{\vec{\theta}}\right) \leq \tau. \tag{29}$$

In the cognitive constraint (29), $I\left(\boldsymbol{\omega_i}; \boldsymbol{\vec{\theta}}\right)$ denotes the mutual information between decision $i$'s signal $\boldsymbol{\omega_i}$ and the fundamental $\boldsymbol{\vec{\theta}}$, which equals to the entropy reduction $H\left(\boldsymbol{\vec{\theta}}\right) - H\left(\boldsymbol{\vec{\theta}}|\boldsymbol{\omega_i}\right)$. It captures the cognitive cost for decision $i$. (29) then means the sum of cognitive costs used by all decisions $i$ cannot surpass the decision maker's total cognitive capacity, $\tau$.

The above costly contemplation problem can be decomposed into two sub-problems. The first is about how decisions are made *given* the chosen information $\{\boldsymbol{\omega_i}\}_{i=1}^N$. 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 $\{\boldsymbol{\omega_i} \in \boldsymbol{\Omega_i}\}_{i=1}^N$ for each decision $i$, subject to the cognitive constraint in (29). One can henceforth interpret the costly contemplation problem in (28) as follows. The decision maker first chooses $\{\boldsymbol{\omega_i}\}_{i=1}^N$, i.e. how each self $i$'s signal is generated, subject to the cognitive constraint in (29). Given the information structure $\{\boldsymbol{\omega_i}\}_{i=1}^N$, different selves play the equivalent incomplete information Bayesian game defined in Proposition 1.

It is worth highlighting the difference of the costly contemplation problem 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. It is also worth noting that, to be parallel with the rational inattention literature, I let the cognitive cost for decision $i$ be equal to the mutual information $I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right)$. In fact, most results in this section can be generalized to the case that the cognitive cost for decision $i$ is an arbitrary continuously differentiable convex function of the mutual information.

## 6.2 Revisiting the Abstract Example

Let me start with the abstract example in (6). To illustrate, let me first consider the $N = 2$ case. Same as (6), the decision maker's utility can be written as $u\left(x_1, x_2, \vec{\theta}\right) = -\frac{1}{2}\left(x_1 - \theta_1\right)^2 - \frac{1}{2}\left(x_2 - \theta_2\right)^2 + \gamma x_1 x_2$, where $\gamma \equiv \gamma_{1,2} = \gamma_{2,1}$. For $i \in \{1, 2\}$, $\boldsymbol{\theta_i} \sim \mathcal{N}\left(\bar{\theta}_i, \sigma^2_{\theta_i}\right)$ and is independent from each other. At the information side, I do not directly impose that each self $i$ has perfect knowledge of her local fundamental $\boldsymbol{\theta_i}$. Instead, I let the decision maker choose endogenously the precision of each self's signal about $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$.

Specifically, each potential signal $\boldsymbol{\omega_i} = \{\boldsymbol{s_{i,1}}, \boldsymbol{s_{i,2}}\} \in \boldsymbol{\Omega_i}$ for decision $i$ consists of a noisy signal about $\boldsymbol{\theta_1}$, $\boldsymbol{s_{i,1}} = \boldsymbol{\theta_1} + \boldsymbol{\epsilon_{i,1}}$, and a noisy signal about $\boldsymbol{\theta_2}$, $\boldsymbol{s_{i,2}} = \boldsymbol{\theta_2} + \boldsymbol{\epsilon_{i,2}}$. All $\boldsymbol{\epsilon}$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 (29).

**Proposition 7** *In the optimum of the costly contemplation problem in (28):*

$$\left(\sigma^*_{1,1}\right)^2 < \left(\sigma^*_{2,1}\right)^2 \quad \textit{and} \quad \left(\sigma^*_{2,2}\right)^2 < \left(\sigma^*_{1,2}\right)^2,$$

*where $\sigma^*_{i,k}$ is the variance of the noise of self $i$'s signal about $\boldsymbol{\theta_k}$ in the optimum.*

Proposition 7 means that, in the optimum, self 1's signal about her local fundamental $\boldsymbol{\theta_1}$ is more precise than self 2's signal about $\boldsymbol{\theta_1}$. Similarly, self 2's signal about her local fundamental $\boldsymbol{\theta_2}$ is more precise than self 1's signal about $\boldsymbol{\theta_2}$. Even though the set of potential signals for two decisions is the same, i.e. $\boldsymbol{\Omega_1} = \boldsymbol{\Omega_2}$, it is optimal to choose different signals for different decisions. Specifically, as $\boldsymbol{\theta_i}$ directly influences self $i$'s optimal decision rule, it is optimal for self $i$ to have a more precise signal about $\boldsymbol{\theta_i}$ than the other self. This also justifies the information structure used in Section 3.3, in which self $i$ has a more precise signal about her local fundamental $\boldsymbol{\theta_i}$ than other selves. As the optimal decision rules in this abstract example are the same as those in the quasi-linear consumer theory context in Section 5.1, Proposition 7 also applies to the quasi-linear consumer theory context (barring an approximation). In that context, it means that each self has a more precise signal about the good she buys than other selves.

It is worth noting that Proposition 7 can be easily extended to the $N-$decision case, as long as the utility $u$ is symmetric across $i$ (Proposition 10 in Appendix E). That is, in the optimum, self

$i$'s signal about her local fundamental $\boldsymbol{\theta_i}$ is more precise than other selves' signals about it. The analysis is more complicated with asymmetric utility functions. What can be established there is a limit result: when the cognitive capacity $\tau$ is small enough, it is optimal for each self $i$ to only receive a signal about her local fundamental $\boldsymbol{\theta_i}$. That is, in the optimum, self $i$'s signals about other fundamentals are completely uninformative. This is Proposition 11 in the Appendix E.

Finally, in Lemma 5 and Propositions 13 and 14 in Appendix E, I also study a more general costly contemplation problem. There, I study general utility functions and allow the potential signals to depend on the fundamental flexibly.[35] Though the analysis there is more involved, the general insight remains to the same: as each self is "interested in" different parts of the fundamental, it is optimal for different selves' signals to take different forms.

# 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. In response to shocks to the fundamental, it is as if each decision is made caring less about the influence of other decisions. I then show how narrow thinking can provide a unified explanation to seemingly disparate behavioral phenomena, including both under-reaction and over-reaction. A general principle is: when the indirect effect works in the same direction as the direct effect, 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.

Finally, let me outline two potential avenues for future research. A first possibility is to explore the general equilibrium implications of narrow thinking. A second possibility is to conduct an experimental test of narrow thinking, along the line of the potential experiment I discussed at the end of Section 3.

[35]Note that in this subsection, consistent with the rest of the paper, I restrict each potential $\boldsymbol{\omega_i} \in \boldsymbol{\Omega_i}$ to have a particular form: each $\boldsymbol{\omega_i}$ consists of $N$ noisy signals, one for each $\boldsymbol{\theta_k}$, $k \in \{1, \cdots, N\}$.

# Appendix A: Proofs

**Proof of Lemma 1.** If the optimum of (2) is not unique,[36] consider two solutions of (2), $\{x_1^*(\cdot), \cdots, x_N^*(\cdot)\}$ and $\{y_1^*(\cdot), \cdots, y_N^*(\cdot)\}$, such that they differ with a non-zero probability. Now, consider $\{z_1^*(\cdot), \cdots, z_N^*(\cdot)\}$, 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 $\vec{x}$, we have $E\left[u\left(z_1^*(\omega_1), \cdots, z_N^*(\omega_N), \vec{\theta}\right)\right] > E\left[u\left(x_1^*(\omega_1), \cdots, x_N^*(\omega_N), \vec{\theta}\right)\right]$ and $E\left[u\left(z_1^*(\omega_1), \cdots, z_N^*(\omega_N), \vec{\theta}\right)\right] > E\left[u\left(y_1^*(\omega_1), \cdots, y_N^*(\omega_N), \vec{\theta}\right)\right]$. This contradicts with the optimality of $\{x_1^*(\cdot), \cdots, x_N^*(\cdot)\}$ and $\{y_1^*(\cdot), \cdots, y_N^*(\cdot)\}$.

**Proof of Lemma 2.** 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 $\{x_1^*(\cdot), \cdots, x_N^*(\cdot)\}$ that satisfies the decision-by-decision optimality condition in (3) but is not the optimum in (2). Let me use $\{y_1^*(\cdot), \cdots, y_N^*(\cdot)\}$ to denote the optimum in (2). We then have

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

We then define

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

From the decision-by-decision optimality condition in (3) and the fact that $u$ is twice continuously differentiable, we have, for all $i$, $E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(\omega_1), \cdots, x_N^*(\omega_N), \vec{\theta}\right) |\omega_i\right] = 0$. Moreover, we have

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

Now, using law of iterated expectations, we have, for each $i$,

$$\begin{aligned} & E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(\omega_1), \cdots, x_N^*(\omega_N), \vec{\theta}\right)\left(y_i^*(\omega_i) - x_i^*(\omega_i)\right)\right] \\ = & E\left[E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(\omega_1), \cdots, x_N^*(\omega_N), \vec{\theta}\right)\left(y_i^*(\omega_i) - x_i^*(\omega_i)\right)|\omega_i\right]\right] \\ = & E\left[E\left[\frac{\partial u}{\partial x_i}\left(x_1^*(\omega_1), \cdots, x_N^*(\omega_N), \vec{\theta}\right)|\omega_i\right]\left(y_i^*(\omega_i) - x_i^*(\omega_i)\right)\right] = 0. \end{aligned}$$

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

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

[36] 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 2.

**Proof of Proposition 2 and Proposition 3.** For notation simplicity, I normalize the mean of each $\theta_i$, $\bar{\theta}_i$, to be zero. Based on Lemma 1 and Lemma 2, 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,

$$x_i^*(\omega_i) = \sum_{k=1}^{M} \alpha_{i,k} s_{i,k}. \tag{30}$$

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 $\lambda_{i,k} = \frac{\sigma^2_{\theta k}}{\sigma^2_{\theta_k} + \sigma^2_{i,k}} \in (0,1]$. We then have

$$E_i\left[x_j^*\right] = \sum_{k=1}^{N} \lambda_{i,k} \alpha_{j,k} s_{i,k}. \tag{31}$$

Together with the optimal decision rule in (5) and the guess in (30), we have, for all $i$,

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

For the guess in (30) to be valid, we then need to have, for all $i, k$,

$$\alpha_{i,k} = \psi_{i,k} \lambda_{i,k} + \sum_{j \neq i} \lambda_{i,k} \gamma_{i,j} \alpha_{j,k}. \tag{32}$$

(32) is satisfied when

$$\begin{pmatrix} \alpha_{1,k} \\ \alpha_{2,k} \\ \cdots \\ \\ \alpha_{N,k} \end{pmatrix} = \left( \mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,k} & \cdots & \lambda_{1,k} & \lambda_{1,k} \\ \lambda_{2,k} & 1 & \cdots & \lambda_{2,k} & \lambda_{2,k} \\ & & \cdots & & \\ \\ \lambda_{N,k} & \lambda_{N,k} & \cdots & \lambda_{N,k} & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} \lambda_{1,k}\psi_{1,k} \\ \lambda_{2,k}\psi_{2,k} \\ \cdots \\ \\ \lambda_{N,k}\psi_{N,k} \end{pmatrix}.$$

This verifies that the guess in (30) indeed characterizes the narrow thinker's optimal decision rules. Because different $\theta_i$s are independent, Proposition 2 then follows immediately from (30) and (31). To prove Proposition 3, note that based on the definition in (11), we then have, for all $i$,

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

Taking partial derivative with respect to each $\theta_k$ then leads to Proposition 3.

*Comment.* In the proof, one may wonder why $\left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,k} & \cdots & \lambda_{1,k} & \lambda_{1,k} \\ \lambda_{2,k} & 1 & \cdots & \lambda_{2,k} & \lambda_{2,k} \\ & & \cdots & & \\ \\ \lambda_{N,k} & \lambda_{N,k} & \cdots & \lambda_{N,k} & 1 \end{pmatrix} \circ \Gamma\right)$ is invertible. Note that the optimal decision rule (5) comes from the decision specific optimality conditions in (3), so $\gamma_{i,j} = -\frac{\partial^2 u}{\partial x_i \partial x_j}\left(\frac{\partial^2 u}{\partial x_i^2}\right)^{-1}$,[37] where $u$ is a strictly concave function. Condition (32) can then be re-written as

$$\lambda_{i,k}^{-1} u_{i,i}\alpha_{i,k} + \sum_{j\neq i} u_{i,j}\alpha_{j,k} = \lambda_{i,k}^{-1} 1_{i=k} \quad \forall i,k,$$

where $u_{i,j} = \frac{\partial^2 u}{\partial x_i \partial x_j}$. To prove $\left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,k} & \cdots & \lambda_{1,k} & \lambda_{1,k} \\ \lambda_{2,k} & 1 & \cdots & \lambda_{2,k} & \lambda_{2,k} \\ & & \cdots & & \\ \\ \lambda_{N,k} & \lambda_{N,k} & \cdots & \lambda_{N,k} & 1 \end{pmatrix} \circ \Gamma\right)$ is invertible is then equivalent to prove $\left(\begin{pmatrix} \lambda_{1,k}^{-1} & 1 & \cdots & 1 & 1 \\ 1 & \lambda_{2,k}^{-1} & \cdots & 1 & 1 \\ & & \cdots & & \\ \\ 1 & 1 & \cdots & 1 & \lambda_{N,k}^{-1} \end{pmatrix} \circ U\right)$ is invertible, where $U(i,j) = u_{i,j}$ is a negative definite matrix (as $u$ is strictly concave over $x$). Then note that,

$$\begin{pmatrix} \lambda_{1,k}^{-1} & 1 & \cdots & 1 & 1 \\ 1 & \lambda_{2,k}^{-1} & \cdots & 1 & 1 \\ & & \cdots & & \\ \\ 1 & 1 & \cdots & 1 & \lambda_{N,k}^{-1} \end{pmatrix} \circ U = U + diag\left\{\left(\lambda_{1,k}^{-1} - 1\right)u_{1,1}, \cdots, \left(\lambda_{N,k}^{-1} - 1\right)u_{N,N}\right\}$$

[37] In the case that (5) comes from a linear approximation, the derivatives are evaluated at the point of linearization.

is also negative definite. As a result,

$$\left(\begin{pmatrix} \lambda_{1,k}^{-1} & 1 & \cdots & 1 & 1 \\ 1 & \lambda_{2,k}^{-1} & \cdots & 1 & 1 \\ & & \cdots & & \\ & & & & \\ 1 & 1 & \cdots & 1 & \lambda_{N,k}^{-1} \end{pmatrix} \circ U\right)$$

and thus

$$\left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,k} & \cdots & \lambda_{1,k} & \lambda_{1,k} \\ \lambda_{2,k} & 1 & \cdots & \lambda_{2,k} & \lambda_{2,k} \\ & & \cdots & & \\ & & & & \\ \lambda_{N,k} & \lambda_{N,k} & \cdots & \lambda_{N,k} & 1 \end{pmatrix} \circ \Gamma\right)$$

is invertible.

**Proof of Proposition 4.** We prove the Proposition 4 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 (13), we can express the own-sensitivity as

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

with $\frac{\partial x_i^{\text{Standard,Ind}}}{\partial \theta_i} = \gamma (N-1) \frac{\Psi + \gamma \psi}{1 - \gamma^2 (N-1) - \gamma (N-2)}$. Using the fact that $\lambda \in [0,1)$, $\psi, \Psi > 0$ and $\gamma \in \left(-1, \frac{1}{N-1}\right)$,[38] Proposition 4 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 (13), we have

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_i} = \psi_{i,i} + \sum_{j \neq i} \gamma_{i,j} \psi_{j,i} + \sum_{j \neq i} \gamma_{i,j} \sum_{l \neq j} \lambda_{j,i} \gamma_{j,l} \psi_{l,i} + \cdots . \tag{33}$$

As each term in (33) 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

[38] This means $1 - \lambda \gamma^2 (N-1) - \lambda \gamma (N-2) > 0$.

all $i$. Based on the proof of Proposition 3, 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

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

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

**Proof of Proposition 5.** We prove the Proposition 5 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 (13), we can express the cross-sensitivity as

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k} = \frac{\lambda\Psi + \lambda\gamma\psi}{1-\lambda\gamma^2(N-1) - \lambda\gamma(N-2)} \quad \forall i.$$

Using the fact that $\lambda \in [0,1)$, $\psi, \Psi > 0$ and $\gamma \in \left(-1, \frac{1}{N-1}\right)$,[39] Proposition 5 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 (13), we have

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k} = \lambda_{i,k}\psi_{i,k} + \sum_{j\neq i}\lambda_{i,k}\gamma_{i,j}\psi_{j,k} + \sum_{j\neq i}\lambda_{i,k}\gamma_{i,j}\sum_{l\neq j}\lambda_{j,k}\gamma_{j,l}\psi_{l,k} + \cdots.$$

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

**Proof of Proposition 6.** Taking an unconditional expectation, averaging over the realization of all fundamentals and signals, of condition (5), we have

$$E\left[x_i^{\text{Narrow}}\right] = E\left[\sum_{1\leq k\leq M}\psi_{i,k}\theta_k\right] + \sum_{j\neq i}\gamma_{i,j}E\left[x_j^{\text{Narrow}}\right] \quad \forall i,$$

[39] This means $1-\lambda\gamma^2(N-1) - \lambda\gamma(N-2) > 0$.

where the law of iterated expectation is used. The above condition also holds when each self perfectly knows all the fundamental:

$$E\left[x_i^{\text{Standard}}\right] = E\left[\sum_{1\le k\le M}\psi_{i,k}\theta_k\right] + \sum_{j\neq i}\gamma_{i,j}E\left[x_j^{\text{Standard}}\right] \quad \forall i.$$

As $(\mathbb{I}_N - \Gamma)$ is invertible. We then have $E\left[x_i^{\text{Narrow}}\right] = E\left[x_i^{\text{Standard}}\right] \quad \forall i.$

**Proof of Lemma 3.** Under narrow thinking, from Proposition 3, we have, for $i \in \{1, 2\}$,

$$\frac{\partial x_i^{\text{Narrow}}}{\partial\theta_i} = 1 + \frac{\lambda\gamma^2}{1-\lambda\gamma^2} \quad \text{and} \quad \frac{\partial x_i^{\text{Narrow}}}{\partial\theta_{-i}} = \frac{\lambda\gamma}{1-\lambda\gamma^2}. \tag{34}$$

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

$$\begin{aligned} x_1^*(\omega_1) &= \alpha_1\theta_1 + \alpha_2 s_{1,2}^{\text{BR}} = \alpha_1\theta_1 + \alpha_2\left(\theta_2 + \epsilon_{1,2}^{\text{BR}}\right), \\ x_2^*(\omega_2) &= \beta_2\theta_2 + \beta_1 s_{2,1}^{\text{BR}} = \beta_2\theta_2 + \beta_1\left(x_1 + \epsilon_{2,1}^{\text{BR}}\right). \end{aligned}$$

From self 1's optimality, we then have

$$x_1^*(\omega_1) = \theta_1 + E_1\left[\gamma x_2^*(\omega_2)\right] = \theta_1 + \gamma E_1\left[\beta_2\theta_2 + \beta_1 x_1\right].$$

As a result,

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

where $\lambda_1^{\text{BR}} = \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_{\text{BR}}^2}$. From self 2's optimality, we then have

$$\begin{aligned} x_2^*(\omega_2) &= \theta_2 + \gamma E_2\left[\alpha_1\theta_1 + \alpha_2\left(\theta_2 + \epsilon_{1,2}^{\text{BR}}\right)\right], \\ &= \left(1+\gamma\alpha_2\right)\theta_2 + \gamma E_2\left[\alpha_1\theta_1 + \alpha_2\epsilon_{1,2}^{\text{BR}}\right], \\ &= \left(1+\gamma\alpha_2\right)\theta_2 + \lambda_2^{\text{BR}}\gamma\left(x_1 + \epsilon_{2,1}^{\text{BR}} - \alpha_2\theta_2\right), \end{aligned}$$

where $\lambda_2^{\text{BR}} = \frac{\alpha_1^2 Var(\theta_1)+\alpha_2^2\sigma_{\text{BR}}^2}{\alpha_1^2 Var(\theta_1)+\alpha_2^2\sigma_{\text{BR}}^2+\Sigma_{\text{BR}}^2}$. As a result,

$$\beta_1 = \lambda_2^{\text{BR}}\gamma \;\; \text{and} \;\; \beta_2 = 1 + \gamma\left(1-\lambda_2^{\text{BR}}\right)\alpha_2.$$

Together, we have

$$\alpha_1 = 1 + \frac{\lambda_2^{\mathrm{BR}}\gamma^2}{1-\lambda_2^{\mathrm{BR}}\gamma^2} \quad \text{and} \quad \alpha_2 = \frac{\lambda_1^{\mathrm{BR}}\gamma}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}$$
$$\beta_1 = \lambda_2^{\mathrm{BR}}\gamma \quad \text{and} \quad \beta_2 = \frac{1-\gamma^2\lambda_2^{\mathrm{BR}}}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}.$$

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

$$x_2^*(\omega_2) = \alpha_1\theta_2 + \alpha_2\left(\theta_1 + \epsilon_{2,1}^{\mathrm{BR}}\right),$$
$$x_1^*(\omega_1) = \beta_2\theta_1 + \beta_1\left(x_2 + \epsilon_{1,2}^{\mathrm{BR}}\right).$$

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

$$\frac{\partial x_i^{\mathrm{BR}}}{\partial \theta_i} = \frac{1}{2}\left(\alpha_1+\beta_2+\beta_1\alpha_2\right) = 1 + \frac{\gamma}{2}\left(\frac{\lambda_2^{\mathrm{BR}}\gamma}{1-\lambda_2^{\mathrm{BR}}\gamma^2} + \frac{\lambda_1^{\mathrm{BR}}\gamma}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}\right) \tag{35}$$
$$\frac{\partial x_i^{\mathrm{PR}}}{\partial \theta_{-i}} = \frac{1}{2}\left(\alpha_2+\beta_1\alpha_1\right) = \frac{1}{2}\left(\frac{\lambda_2^{\mathrm{BR}}\gamma}{1-\lambda_2^{\mathrm{BR}}\gamma^2} + \frac{\lambda_1^{\mathrm{BR}}\gamma}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}\right). \tag{36}$$

Compared to the formula about own- and cross- sensitivities under narrow thinking in (34), to prove Lemma 3, we only need to prove that, for any $\lambda \in (0,1)$, we can find $\sigma_{BR}^2, \Sigma_{BR}^2 > 0$ such that

$$\frac{\lambda\gamma}{1-\lambda\gamma^2} = \frac{1}{2}\left(\frac{\lambda_2^{\mathrm{BR}}\gamma}{1-\lambda_2^{\mathrm{BR}}\gamma^2} + \frac{\lambda_1^{\mathrm{BR}}\gamma}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}\right).$$

This is true as

$$\lim_{\sigma_{BR}^2\to 0, \Sigma_{BR}^2\to 0} = \frac{1}{2}\left(\frac{\lambda_2^{\mathrm{BR}}\gamma}{1-\lambda_2^{\mathrm{BR}}\gamma^2} + \frac{\lambda_1^{\mathrm{BR}}\gamma}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}\right) = \frac{\gamma}{1-\gamma^2}$$

and

$$\lim_{\sigma_{BR}^2\to +\infty, \Sigma_{BR}^2\to +\infty} = \frac{1}{2}\left(\frac{\lambda_2^{\mathrm{BR}}\gamma}{1-\lambda_2^{\mathrm{BR}}\gamma^2} + \frac{\lambda_1^{\mathrm{BR}}\gamma}{1-\gamma^2\left(\lambda_2^{\mathrm{BR}}\left(1-\lambda_1^{\mathrm{BR}}\right)+\lambda_1^{\mathrm{BR}}\right)}\right) = 0.$$

**Proof of Corollary 1.** When $\gamma > 0$, the current environment falls into case 2) in Assumption 1. When $\gamma < 0$, the current environment falls into case 3) in Assumption 1, with $\rho_i = 1$, $\Gamma_i = \gamma$ and $\mathit{\Delta}_i = 0$. Corollary (1) then follows from Proposition 4.

**Proof of Corollary 2.** Two conditions in Corollary 2 correspond to 1) and 2) in Assumption 1. The result then follows directly from Proposition 5.

**Proof of Corollary 3.** The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\frac{\kappa_h}{\kappa_i}}{1+\frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}$ and $\Gamma_i = \Delta_i = \frac{\mu_i}{\mu_y}$. As a result, the Corollary directly follows. To see this directly, we can have explicit formula for own-price and cross-price elasticities.

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

$$\frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_k} = \frac{\kappa_h}{\kappa_k} \frac{\mu_k \frac{1-\kappa_k}{\kappa_k}}{\left(\sum_{j\neq k}\mu_j \frac{\mu_y \lambda_{j,k}}{\mu_y\kappa_k + (1-\lambda_{j,k})\kappa_h\mu_j} + \frac{\mu_k}{\kappa_k}\right)\kappa_h + \mu_y} - \frac{1}{\kappa_k} \quad \forall k, \tag{37}$$

where $\lambda_{i,k} = \frac{\sigma_{p_k}^2}{\sigma_{p_k}^2 + \sigma_{i,k}^2}$. Note that $\sum_{j\neq k}\mu_j \frac{\mu_y\lambda_{j,k}}{\mu_y\kappa_j + (1-\lambda_{j,k})\kappa_h\mu_j}$ is smaller than its standard counterpart (all $\lambda$s are 1). As a result, when $\kappa_k > 1$, $\frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_k} < \frac{\partial \hat{x}_k^{\text{Standard}}}{\partial \hat{p}_k} < 0$. On the other hand, when $\kappa_k < 1$, we have $\frac{\partial \hat{x}_k^{\text{Standard}}}{\partial \hat{p}_k} < \frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_k}$.

These formula also help me prove the statement in the main text that, "narrow thinking nevertheless brings the consumer's demand elasticity closer to the case of explicit mental budgeting." For a decision maker with explicit mental budgets (indexed by $MB$), $p_k x_k^{MB} = M_k \quad \forall k$. As a result, $\frac{\partial \hat{x}_k^{\text{MB}}}{\partial \hat{p}_k} = -1$. Then note that, from (37), we have $\frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_k} = \frac{\partial \hat{x}_k^{\text{MB}}}{\partial \hat{p}_k} - \frac{1-\kappa_k}{\kappa_k}\left(\frac{\left(\sum_{j\neq k}\mu_j \frac{\mu_y\lambda_{j,k}}{\mu_y\kappa_j + (1-\lambda_{j,k})\kappa_h\mu_j}\right)\kappa_h + \mu_y}{\left(\sum_{j\neq k}\mu_j \frac{\mu_y\lambda_{j,k}}{\mu_y\kappa_j + (1-\lambda_{j,k})\kappa_h\mu_j} + \frac{\mu_k}{\kappa_k}\right)\kappa_h + \mu_y}\right)$. As $\sum_{j\neq k}\mu_j \frac{\mu_y\lambda_{j,k}}{\mu_y\kappa_j + (1-\lambda_{j,k})\kappa_h\mu_j}$ is smaller than its standard counterpart (all $\lambda$s are 1), narrow thinking moves $\frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_k}$ closer to $\frac{\partial \hat{x}_k^{\text{MB}}}{\partial \hat{p}_k}$, no matter whether $\kappa_k > 1$ or $\kappa_k < 1$.

**Proof of Corollary 4.** In the proof, for notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. The optimality condition for each decision $i$ and the budget constraint become

$$\varphi_i - \kappa_i x_i^*(\omega_i) = -\kappa_h E_i[y^*] \quad \forall i, \tag{38}$$

$$\sum_{i=1}^{N} \mu_i x_i^*(\omega_i) + \mu_y y^* = 0, \tag{39}$$

where, as in Section 5.2, $\kappa_h = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$. Substituting $y^*$ in (38) based on (39), we have

$$x_i^*(\omega_i) = \frac{\frac{1}{\kappa_i}}{1 + \frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}\varphi_i - E_i\left[\sum_{j\neq i}\frac{\frac{\kappa_h}{\kappa_i}\frac{\mu_j}{\mu_y}}{1 + \frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}} x_j^*(\omega_j)\right].$$

The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\frac{\kappa_h}{\kappa_i}\frac{1}{\mu_y}}{1+\frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}$, $\Gamma_i = \mu_i$ and $\Delta_i = 0$. The Corollary then follows from Proposition 4.

**Proof of Proposition 7.** As discussed in the main text, the costly contemplation problem in (28) can be divided into two subproblems, the optimal information choice subject to the cognitive constraint in (29), and the optimal decisions *given* the chosen information. From condition (7), given any chosen information $\{\boldsymbol{\omega_i}\}_{i=1}^{2}$, the optimal decision rule $\{x_i^*(\cdot)\}_{i=1}^{2}$ can be characterized by

$$E\left[x_i^*(\omega_i) - \theta_i - \gamma x_{-i}^*(\omega_{-i})\,|\omega_i\right] = 0 \quad \forall i, \omega_i \in \Omega_i. \tag{40}$$

Using law of iterated expectations, we henceforth have

$$\frac{1}{2}E\left[\left[x_i^*(\omega_i)\right]^2 - \theta_i x_i^*(\omega_i) - \gamma x_i^*(\omega_i)\, x_{-i}^*(\omega_{-i})\right] = 0 \quad \forall i \in \{1,2\}.$$

Substituting into the decision maker's utility function, the optimal information choice in (29) is then equivalent to

$$\max_{\{\boldsymbol{\omega_i} \in \boldsymbol{\Omega_i}\}_{i=1}^{2}} \frac{1}{2}E\left[\theta_1 x_1^*(\omega_1) + \theta_2 x_2^*(\omega_2)\right] \tag{41}$$
$$s.t. \quad x_i^*(\omega_i) \text{ satisfy (40)}$$
$$\sum_{i=1}^{2} I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right) \leq \tau.$$

Now, given the $\boldsymbol{\Omega_i}$ specified in the main text, any $\boldsymbol{\omega_i} = \{\boldsymbol{s_{i,1}}, \boldsymbol{s_{i,2}}\}$ takes the form of $\boldsymbol{s_{i,1}} = \boldsymbol{\theta_1} + \boldsymbol{\epsilon_{i,1}}$ and $\boldsymbol{s_{i,2}} = \boldsymbol{\theta_2} + \boldsymbol{\epsilon_{i,2}}$, with $\boldsymbol{\epsilon_{i,1}} \sim N\left(0, \sigma_{i,1}^2\right)$, $\boldsymbol{\epsilon_{i,2}} \sim N\left(0, \sigma_{i,2}^2\right)$ and all $\boldsymbol{\epsilon}$s and $\boldsymbol{\theta}$s are independent from each other. Similar to the proof of Proposition 3, we have

$$\begin{pmatrix} \frac{\partial E[x_1^*(\omega_1)|\theta_1,\theta_2]}{\partial \theta_1} \\ \frac{\partial E[x_2^*(\omega_2)|\theta_1,\theta_2]}{\partial \theta_1} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,1} \\ \lambda_{2,1} & 1 \end{pmatrix} \circ \Gamma\right)^{-1} \begin{pmatrix} \lambda_{1,1} \\ 0 \end{pmatrix},$$

and

$$\begin{pmatrix} \frac{\partial E[x_1^*(\omega_1)|\theta_1,\theta_2]}{\partial \theta_2} \\ \frac{\partial E[x_2^*(\omega_2)|\theta_1,\theta_2]}{\partial \theta_2} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,2} \\ \lambda_{2,2} & 1 \end{pmatrix} \circ \Gamma\right)^{-1} \begin{pmatrix} 0 \\ \lambda_{2,2} \end{pmatrix},$$

where $\lambda_{i,j} = \frac{\sigma_{\theta_j}^2}{\sigma_{\theta_j}^2 + \sigma_{i,j}^2} \in (0,1]$. The problem in (41) then becomes

$$\max_{\{0 \leq \lambda_{i,j} \leq 1\}_{1 \leq i,j \leq 2}} g\left(\{\lambda_{i,j}\}_{1 \leq i,j \leq 2}\right) \equiv \frac{1}{2}\frac{\lambda_{1,1}}{1 - \lambda_{1,1}\lambda_{2,1}\gamma^2}\sigma_{\theta_1}^2 + \frac{1}{2}\frac{\lambda_{2,2}}{1 - \lambda_{2,2}\lambda_{1,2}\gamma^2}\sigma_{\theta_2}^2 \tag{42}$$
$$s.t.\ h\left(\{\lambda_{i,j}\}_{1 \leq i,j \leq 2}\right) \equiv \sum_{1 \leq i,j \leq 2} \frac{1}{2}log_2\left(\frac{1}{1 - \lambda_{i,j}}\right) \leq \tau,$$

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

$\frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,2}}\right)$.

Now we prove Proposition 7. If, in the optimum of the costly contemplation problem, we have $\left(\sigma^*_{1,1}\right)^2 \geq \left(\sigma^*_{2,1}\right)^2$. This means $\lambda^*_{1,1} \leq \lambda^*_{2,1}$, where $\lambda^*_{i,j} = \frac{\sigma^2_{\theta_j}}{\sigma^2_{\theta_j}+\left(\sigma^*_{i,j}\right)^2} \in (0,1]$. As a result, $\frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{1,1}} > \frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{2,1}}$ and $\frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{1,1}} \leq \frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{2,1}}$. This is inconsistent with the first order condition of (42):

$$\frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{1,1}} / \frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{1,1}} = \frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{2,1}} / \frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{2,1}}.$$

Therefore, $\left(\sigma^*_{1,1}\right)^2 < \left(\sigma^*_{2,1}\right)^2$. Similarly, we can prove $\left(\sigma^*_{2,2}\right)^2 < \left(\sigma^*_{1,2}\right)^2$.

# Appendix B: Optimal Behavior under Narrow Thinking

## Dampening of Indirect Effects.

Consider the optimal decision rules in (5). As defined above, each decision $i$'s response to shocks to fundamentals can be decomposed into a direct and an indirect effect. As each self has an imperfect understanding of other selves' decisions, the indirect effect should be dampened under narrow thinking in response to shocks to fundamentals.

To formalize the above intuition, 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 (5), we have

$$
\begin{aligned}
x_i^*(\omega_i) &= \sum_{k=1}^{M} \psi_{i,k} E_i[\theta_k] + \sum_{j \neq i} \gamma_{i,j} E_i\left[x_j^*\right] \\
&= \sum_{k=1}^{M} \psi_{i,k} E_i[\theta_k] + \sum_{j \neq i} \gamma_{i,j} \left( \sum_{k=1}^{M} \psi_{j,k} E_i[E_j[\theta_k]] \right) + \sum_{j \neq i} \gamma_{i,j} \left( \sum_{l \neq j} \gamma_{j,l} E_i\left[E_j\left[x_l^*\right]\right] \right) \\
&= \underbrace{\sum_{k=1}^{M} \psi_{i,k} E_i[\theta_k]}_{\text{Direct}} + \underbrace{\sum_{j \neq i} \gamma_{i,j} \left( \sum_{k=1}^{M} \psi_{j,k} E_i[\theta_k] \right) + \sum_{j \neq i} \gamma_{i,j} \left( \sum_{l \neq j} \gamma_{j,l} \left( \sum_{k=1}^{M} \psi_{l,k} E_i[E_j[\theta_k]] \right) \right) + \cdots}_{\text{Indirect}}.
\end{aligned}
\tag{43}
$$

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 $\theta$s, 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 (11), averaging over the realization of noises in signals, one can then define each part of the indirect effects as a function of fundamentals: $x_i^{\text{Ind},+,\text{Narrow}}(\theta_1, \cdots, \theta_M) \equiv E\left[x_i^{\text{Ind},+}(\omega_i) | \theta_1, \cdots, \theta_M\right]$ and $x_i^{\text{Ind},-,\text{Narrow}}(\theta_1, \cdots, \theta_M) \equiv E\left[x_i^{\text{Ind},-}(\omega_i) | \theta_1, \cdots, \theta_M\right]$. One can then establish:

**Proposition 8** *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 $\theta_k$,*

$$
\left| \frac{\partial x_i^{\text{Ind},+,\text{Narrow}}}{\partial \theta_k} \right| \leq \left| \frac{\partial x_i^{\text{Ind},+,\text{Standard}}}{\partial \theta_k} \right| \quad \textit{and} \quad \left| \frac{\partial x_i^{\text{Ind},-,\text{Narrow}}}{\partial \theta_k} \right| \leq \left| \frac{\partial x_i^{\text{Ind},-,\text{Standard}}}{\partial \theta_k} \right| \quad \forall i, k
$$

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

**Proof of Proposition 8.** 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

$$
E_{i_1}\left[E_{i_2}\left[\cdots E_{i_m}[\theta_k]\right]\right] = \lambda_{i_1,k} \cdots \lambda_{i_m,k} s_{i_1,k},
\tag{44}
$$

and

$$
E\left[E_{i_1}\left[E_{i_2}\left[\cdots E_{i_m}[\theta_k]\right]\right] | \vec{\theta}\right] = \lambda_{i_1,k} \cdots \lambda_{i_m,k} \theta_k.
\tag{45}
$$

From conditions (43) and (45), we have

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k} = \lambda_{i,k}\psi_{i,k} + \sum_{j\neq i} \lambda_{i,k}\gamma_{i,j}\psi_{j,k} + \sum_{j\neq i} \lambda_{i,k}\gamma_{i,j} \sum_{l\neq j} \lambda_{j,k}\gamma_{j,l}\psi_{l,k} + \cdots .$$

Using the fact that each $\lambda$ is a factor between 0 and 1 and collecting terms with positive and negative coefficients prove Proposition 8.

## Comparison with Standard Sequential Decisions.

Here, in the sequential environment (15) in Section 4, 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 $\omega_1 = \left\{\theta_1, s_{1,2}^{\text{PR}} = \theta_2 + \epsilon_{1,2}^{\text{PR}}\right\}$, where $\epsilon_{1,2}^{\text{PR}} \sim \mathcal{N}\left(0, \sigma_{\text{PR}}^2\right)$. That is, she perfectly knows $\theta_1$, and receives a noisy signal about the future $\theta_2$, 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 $\omega_2 = \left\{\theta_1, \theta_2, s_{1,2}^{\text{PR}}, x_1\right\}$. 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 $\omega_2 = \left\{\theta_2, s_{2,1}^{\text{PR}} = \theta_1 + \epsilon_{2,1}^{\text{PR}}\right\}$ and $\omega_1 = \left\{\theta_1, \theta_2, s_{2,1}^{\text{PR}}, x_2\right\}$.

We then compare the narrow thinker's behavior $x_i^{\text{Narrow}}\left(\vec{\theta}\right)$ with the behavior of this decision maker with perfect recall (PR), $x_i^{\text{PR}}\left(\vec{\theta}\right) \equiv E\left[x_i^*\left(\omega_i\right)|\vec{\theta}\right]$, where, as above, $E\left[\cdot|\vec{\theta}\right]$ averages over not only the realization of noises in signals but also the potential order of decisions.

**Lemma 4** *There exists a $\bar{\sigma} > 0$ such that for any narrow thinker with the size of noise $\sigma > \bar{\sigma}$, her own- and cross- sensitivities deviate more from the frictionless counterparts than the counterparts of any decision maker with perfect recall (PR, with any $\sigma_{\text{PR}}^2$):*

$$\left|\frac{\partial x_i^{Narrow}}{\partial \theta_i}\right| \leq \left|\frac{\partial x_i^{PR}}{\partial \theta_i}\right| \leq \left|\frac{\partial x_i^{Standard}}{\partial \theta_i}\right| \quad and \quad \left|\frac{\partial x_i^{Narrow}}{\partial \theta_{-i}}\right| \leq \left|\frac{\partial x_i^{PR}}{\partial \theta_{-i}}\right| \leq \left|\frac{\partial x_i^{Stanford}}{\partial \theta_{-i}}\right|.$$

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 $x_1^*\left(\omega_1\right) = x_1^{\text{Standard}}\left(\theta_1, E_1\left[\theta_2\right]\right)$. 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. In this environment, as established in Corollary 1 below, the indirect effect works in the same direction as the direct effect in this environment. As a result, the additional frictional behavior under narrow thinking leads to more under-reaction, as in Lemma 4.

**Proof of Lemma 4.** First consider the case that the decision 1 is made before the decision 2. As self 2 has perfect recall, we have

$$x_2^*(\omega_2) = \theta_2 + \gamma x_1^*(\omega_1).$$

From self 1's perceptive, we then have

$$x_1^*(\omega_1) = \theta_1 + E_1[\gamma x_2^*(\omega_2)] = \theta_1 + E_1\left[\gamma\theta_2 + \gamma^2 x_1^*(\omega_1)\right].$$

As a result,

$$\begin{aligned} x_1^*(\omega_1) &= \frac{1}{1-\gamma^2}\theta_1 + \frac{\lambda^{\text{PR}}\gamma}{1-\gamma^2}\left(\theta_2 + \epsilon_{1,2}^{\text{PR}}\right), \\ x_2^*(\omega_2) &= \theta_2\left(1 + \frac{\lambda^{\text{PR}}\gamma^2}{1-\gamma^2}\right) + \frac{\gamma}{1-\gamma^2}\theta_1 + \frac{\lambda^{\text{PR}}\gamma^2}{1-\gamma^2}\epsilon_{1,2}^{\text{PR}} \end{aligned}$$

where $\lambda^{\text{PR}} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_{\text{PR}}^2}$.

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

$$\begin{aligned} x_2^*(\omega_2) &= \frac{1}{1-\gamma^2}\theta_2 + \frac{\lambda^{\text{PR}}\gamma}{1-\gamma^2}\left(\theta_1 + \epsilon_{2,1}^{\text{PR}}\right), \\ x_1^*(\omega_1) &= \theta_1\left(1 + \frac{\lambda^{\text{PR}}\gamma^2}{1-\gamma^2}\right) + \frac{\gamma}{1-\gamma^2}\theta_2 + \frac{\lambda^{\text{PR}}\gamma^2}{1-\gamma^2}\epsilon_{2,1}^{\text{PR}}. \end{aligned}$$

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

$$\frac{\partial x_i^{\text{PR}}}{\partial \theta_i} = 1 + \frac{1}{2}\frac{\gamma^2\left(1+\lambda^{\text{PR}}\right)}{1-\gamma^2} \quad \text{and} \quad \frac{\partial x_i^{\text{PR}}}{\partial \theta_{-i}} = \frac{1}{2}\frac{\gamma\left(1+\lambda^{\text{PR}}\right)}{1-\gamma^2}. \tag{46}$$

Compared to the formula about own- and cross- sensitivities under narrow thinking in (34), Lemma 4 follows with $\bar{\sigma} = \sqrt{1-\gamma^2}\sigma_\theta$.

# Appendix C: Consumer Theory under Narrow Thinking

## Identification of Demand Gradients.

Consider an environment with $K$ consumers. All consumers have the same utility as (18). 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 $t \in \{1, \cdots, T\}$, $p_{i,t}^k$, is drawn i.i.d (across time, consumers and goods) from consumer $k$'s price distribution, $\log\mathcal{N}\left(\log\bar{p}_i^k, \sigma_{p_i^k}^2\right)$. Same as Section 5.1, each self $i \in \{1, \cdots, N\}$ of the consumer $k$ at period $t$ perfectly knows the price of the good she buys $p_{i,t}^k$, but only receives a noisy signal about each of the other price faced by her other selves, $p_{j,t}^k$.

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. Specifically, for each consumer $k$, we can look at how each of her consumption $x_{i,k}^t$ moves with respect to $\vec{p}_t^k = \left(p_{1,t}^k, \cdots, p_{N,t}^k\right)$, 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.

# A General Case

Here I study a general non-quasilinear case with symmetry. Specifically, I let the consumer's utility be $v(x_1, \cdots, x_N) + h(y)$, where $v$ and $h$ are strictly increasing in each of her arguments, strictly concave and twice differentiable.[40] The consumer is subject to the budget constraint $\sum_{i=1}^N p_i x_i + y = w$. I consider the same information structure as in Section 5. Specifically, each self $i \in \{1, \cdots, N\}$ of the narrow thinker, who is in charge of purchasing good $i$, perfectly knows $p_i \sim \log \mathcal{N}\left(\log \bar{p}, \sigma_p^2\right)$, but receives a noisy signal about each of the other $p_k$. To summarize, for $i \in \{1, \cdots, N\}$, self $i$'s signal is given by $\omega_i = \left\{\{s_{i,k}\}_{k \in \{1, \cdots, N\}}\right\}$, where $s_{i,i} = p_i$ and, for all $i \neq k$, $s_{i,k} = p_k \epsilon_{i,k}$, with $\epsilon_{i,k} \sim \log \mathcal{N}\left(0, \sigma^2\right)$ and $\sigma^2 > 0$. All $\epsilon$s and $p$s are independent from each other. The last self, who is in charge of the consumption of $y$, has perfect knowledge of the fundamentals and other decisions. This makes sure that the budget constraint always holds. The problem is symmetric across $i \in \{1, \cdots, N\}$. It means that the utility function $v$ is symmetric across each good $i$ and each self $i$'s signal about other prices $p_k$ have the same signal-to-noise ratio, i.e. $\lambda_{i,k} = \lambda = \frac{\sigma_p^2}{\sigma_p^2 + \sigma^2}$ are the same for all $i \neq k$.

Similar to condition (23), the optimal consumption decision of each self $i \in \{1, \cdots, N\}$, $x_i^*(\omega_i)$, must satisfy

$$E_i\left[\frac{\partial v}{\partial x_i}\left(x_i^*(\omega_i), \vec{x}_{-i}^*\right)\right] = p_i E_i\left[h'(y^*)\right]. \tag{47}$$

That is, from each self $i$'s perfective, her expected marginal rate of substitution between the consumption of good $i$ and the consumption of $y$ should equal $p_i$. Log-linearizing the above condition and budget constraint around the point where each price $i$ is fixed at $\bar{p}_i$ and each decision is made with perfect knowledge of all

[40] I let $h(y)$ be well defined for all $y \in \mathbb{R}$. This allows the possibility that the "residual decision" $y$ is negative and guarantees that the budget constraint will always be satisfied.

prices, we have

$$\hat{x}_i^*(\omega_i) = -\psi^{'}\hat{p}_i + \sum_{j\neq i}\gamma^{'} E_i\left[\hat{x}_j^*(\omega_j)\right] + \kappa_h E_i\left[\hat{y}^*\right], \tag{48}$$

$$\sum_{i=1}^{N}\mu\left(\hat{x}_i^*(\omega_i) + \hat{p}_i\right) + \mu_y\hat{y}^* = 0, \tag{49}$$

where, with symmetry, $\psi^{'} = -\frac{\bar{p}_i\frac{\partial v(\bar{x}_1,\cdots,\bar{x}_N)}{\partial x_i}}{\frac{\partial^2 v(\bar{x}_1,\cdots,\bar{x}_N)}{\partial x_i^2}\bar{x}_i} > 0$, $\gamma^{'} = -\frac{\frac{\partial^2 v(\bar{x}_1,\cdots,\bar{x}_N)}{\partial x_i\partial x_j}\bar{x}_j}{\frac{\partial^2 v(\bar{x}_1,\cdots,\bar{x}_N)}{\partial x_i^2}\bar{x}_i} \in \left(-\frac{1}{N-1}, \frac{1}{N-1}\right)$ and $\kappa_h = \frac{h^{''}(\bar{y})\bar{y}\frac{\partial v(\bar{x}_1,\cdots,\bar{x}_N)}{\partial x_i}}{h^{'}(\bar{y})\frac{\partial^2 v(\bar{x}_1,\cdots,\bar{x}_N)}{\partial x_i^2}\bar{x}_i} > 0$. $\mu = \frac{\bar{p}_i\bar{x}_i}{w}$ and $\mu_y = \frac{\bar{y}}{w}$ are the spending share of each good $i$ and $y$ at the point of log-linearization.

Substituting the last self's consumption, $\hat{y}^*$ and using the budget constraint, we then have

$$\hat{x}_i^*(\omega_i) = -\psi\hat{p}_i - \sum_{j\neq i}\Psi E_i\left[\hat{p}_j\right] + \sum_{j\neq i}\gamma E_i\left[\hat{x}_j^*(\omega_j)\right], \tag{50}$$

where $\psi = \frac{\psi^{'}+\frac{\kappa_h\mu}{\mu_y}}{1+\frac{\kappa_h\mu}{\mu_y}} > 0$, $\Psi = \frac{\frac{\kappa_h\mu}{\mu_y}}{1+\frac{\kappa_h\mu}{\mu_y}} > 0$, and $\gamma = \frac{\gamma^{'}-\frac{\kappa_h\mu}{\mu_y}}{1+\frac{\kappa_h\mu}{\mu_y}} \in \left(-1, \frac{1}{N-1}\right)$. The current environment then falls into the symmetry case in Propositions 4 and 5. I can then establish results about own-price and cross-price demand elasticity under narrow thinking in this set-up.

Specifically, similar to Section 5.1, I define the narrow thinker's (log) demand function as $\hat{x}_i^L(\hat{p}_1,\cdots,\hat{p}_N) \equiv E\left[\hat{x}_i^*(\omega_i)\,|\hat{p}_1,\cdots,\hat{p}_N\right]$, averaging over the realization of noises in signals. Similar to the main text, I use superscript *Standard* to denote standard consumer theory's demand function when each decision is made with perfect knowledge of all prices.

**Corollary 5** *In terms of own-demand elasticity:*

1. *When the indirect effect works in the same direction as the direct effect, that is, when* $\frac{(N-1)\gamma(\Psi+\gamma\psi)}{1-\gamma^2(N-1)-\gamma(N-2)} > 0$, *narrow thinking leads to under-reaction:*

$$\frac{\partial x_i^{\text{Standard}}}{\partial p_i} < \frac{\partial x_i^{\text{Narrow}}}{\partial p_i} < 0.$$

2. *When the indirect effect works in the opposite direction to the direct effect, that is, when* $\frac{(N-1)\gamma(\Psi+\gamma\psi)}{1-\gamma^2(N-1)-\gamma(N-2)} < 0$, *narrow thinking leads to over-reaction:*

$$\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} < \frac{\partial x_i^{\text{Standard}}}{\partial p_i} < 0.$$

**Corollary 6** *The cross-price demand elasticities are attenuated under narrow thinking:*

$$\left|\frac{\partial\hat{x}_i^{\text{Narrow}}}{\partial\hat{p}_k}\right| \leq \left|\frac{\partial\hat{x}_i^{\text{Standard}}}{\partial\hat{p}_k}\right| \quad \forall i\neq k.$$

**Proof of Corollary 5 and 6.** Because of the symmetry in the environment, the result follows directly from Propositions 4 and 5. One can also prove the corollaries directly. In this environment, we have, for $i \neq k$,

$$\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} = -\psi + \gamma (N-1) \frac{-\lambda \Psi - \lambda \gamma \psi}{1 - \lambda \gamma^2 (N-1) - \lambda \gamma (N-2)}, \tag{51}$$

$$\frac{\partial x_i^{\text{Narrow}}}{\partial p_k} = \frac{-\lambda \Psi - \lambda \gamma \psi}{1 - \lambda \gamma^2 (N-1) - \lambda \gamma (N-2)}. \tag{52}$$

Using the fact that $\lambda \in [0,1)$, $\psi > \Psi > 0$ and $\gamma \in \left(-1, \frac{1}{N-1}\right)$,[41] the corollaries follow.

# Appendix D: Other Applications

## Comfort Zones

The decision maker's utility is given by,

$$\sum_{i=1}^{N} \varphi_i v_i (x_i) - c \left( \sum_{i=1}^{N} x_i \right), \tag{53}$$

where $x_i$ is the time the decision maker assigns to activity $i$, $\varphi_i v_i (x_i) = \varphi_i \frac{x_i^{1-\kappa_i}}{1-\kappa_i}$ with $\kappa_i > 0$ is her utility from activity $i$, $\varphi_i$ parametrizes the attractiveness of activity $i$, $c\left(\sum_{i=1}^{N} x_i\right)$ is the opportunity cost of time, and $c(x) = \frac{x^{1+\kappa_c}}{1+\kappa_c}$ with $\kappa_c > 0$ is a strictly convex function.

For consistency, I consider a information structure for the narrow thinker similar to the one used in the main text. As I will work with log-linearization later, I let fundamentals and signals be log-normally distributed. Specifically, each self $i \in \{1, \cdots, N\}$ of the narrow thinker, who is in charge of activity $i$, perfectly knows $\varphi_i \sim \log \mathcal{N}\left(\log \bar{\varphi}_i, \sigma_{\varphi_i}^2\right)$, but receives a noisy signal about each of the other $\varphi_k$. This makes sure that the budget constraint always holds. Specifically, for $i \in \{1, \cdots, N\}$, self $i$'s signal is given by $\omega_i = \left\{ \{s_{i,k}\}_{k \in \{1, \cdots, N\}} \right\}$, where $s_{i,i} = \varphi_i$ and, for $i \neq j$, $s_{i,k} = \varphi_k \epsilon_{i,k}$, with $\epsilon_{i,k} \sim \log \mathcal{N}\left(0, \sigma_{i,k}^2\right)$ and $\sigma_{i,k}^2 > 0$. All $\epsilon$s and $\varphi$s are independent from each other.

I use a hat over a variable to denote its log-deviation from the point of log-linearization.[42] Then, similar to condition (20), for all $i$, I define the narrow thinker's (log) decision function as a function of fundamentals:

$$\hat{x}_i^{\text{Narrow}} (\hat{\varphi}_1, \cdots, \hat{\varphi}_N) \equiv E\left[\hat{x}_i^* (\omega_i) | \hat{\varphi}_1, \cdots, \hat{\varphi}_N\right], \tag{54}$$

[41] This means $1 - \lambda \gamma^2 (N-1) - \lambda \gamma (N-2) > 0$.

[42] Specifically, I log linearize around the point where each $\varphi_i$ is fixed at $\bar{\varphi}_i$ and each decision is made with perfect knowledge of all fundamentals.

where $E\left[\cdot|\hat{\varphi}_1, \cdots, \hat{\varphi}_N\right]$ averages over the realization of noises in signals. 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 then have:

**Corollary 7 (A narrow thinking theory of comfort zones)** *For each $i$, the narrow thinker increases (decreases) her time allocated for activity $i$ less in response to positive (negative) taste shocks to $\varphi_i$ :*

$$\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{\varphi}_i} > \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\varphi}_i} > 0.$$

To understand the intuition behind the Corollary, first consider the case that each decision is made with perfect knowledge of all the fundamentals. An increase in $\varphi_i$ will increase $x_i$ as activity $i$ becomes more attractive (a positive direct effect), but decrease other $x_j$ for $j \neq i$, as the cost function is convex over the sum of efforts. The decrease of other $x_j$ then further increases $x_i$ (a positive indirect effect). Under narrow thinking, such a coordinated response to $\varphi_i$ is hindered: other selves' $x_j$ will not decrease as much in response to the increase in $\varphi_i$. As a result, $x_i$ will not increase as much. A dampening of the indirect effect under narrow thinking then leads to under-reaction. The narrow thinker stays within her comfort zones, despite the fact that activity $i$ becomes more attractive.

For a concrete example of the above comfort zones behavior, consider an engineering student who decides how much time she will spend on the economics class. She realizes the economics class has a great professor, that is, the $\varphi_i$ of spending time on the economics class is high. However, she is concerned that she will not decrease the amount of time spent on engineering classes. In the end, as the student has limited time, it is hard for the engineering student to go outside of her comfort zone and engage in the economics class, even with an excellent professor.

**Proof of Corollary 7.** In the proof, for notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. The optimality condition for each decision $i$ becomes

$$\varphi_i - \kappa_i x_i^*\left(\omega_i\right) = \kappa_c E_i\left[\sum_{j=1}^{N} \mu_j x_j^*\left(\omega_j\right)\right] \qquad \forall i,$$

where $\mu_i = \frac{\bar{x}_i}{\sum_{l=1}^{N} \bar{x}_l}$ denotes the share of time spent on activity $i$ in the steady state. Collecting terms, we have

$$x_i^*\left(\omega_i\right) = \frac{\frac{1}{\kappa_i}}{1 + \frac{\kappa_h}{\kappa_i}\frac{\mu_i}{\mu_y}}\varphi_i - E_i\left[\sum_{j\neq i} \frac{\frac{\kappa_c}{\kappa_i}\mu_j}{1 + \frac{\kappa_c}{\kappa_i}\mu_i} x_j^*\left(\omega_j\right)\right].$$

The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\frac{\kappa_c}{\kappa_i}}{1+\frac{\kappa_c}{\kappa_i}\mu_i}$, $\Gamma_i = \mu_i$ and $\Delta_i = 0$. The Corollary then follows from Proposition 4.

## 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

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

where $l_i$ is the labor supply on day $i$, $v\left(l_i\right) = \frac{l_i^{1+\kappa}}{1+\kappa}$ captures the disutility of labor on day $i$, and $h\left(y\right) = \frac{y^{1-\kappa_h}}{1-\kappa_h}$ is her utility from consumption, with $\kappa > 0$ and $\kappa_h > 0$. 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 I am focusing on response to daily wage fluctuations, I let different $w_i$s be independent.

**Information.** In this environment, each self $i \in \{1, \cdots, N\}$ 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' $w_j$. The last self, who is in charge of the consumption of $y$, has perfect knowledge of all fundamentals and other decisions. This makes sure that the budget constraint always holds. As I would work with log-linearization later, I let prices and signals be log-normally distributed. Specifically, for $i \in \{1, \cdots, N\}$, self $i$'s information (signals) is given by $\omega_i = \left\{\{s_{i,j}\}_{j\in\{1,\cdots,N\}}\right\}$, where $s_{i,i} = w_i \sim \log\mathcal{N}\left(\log\bar{w}_i, \sigma^2_{w_i}\right)$ and, for $i \neq j$, $s_{i,j} = w_j\epsilon_{i,j}$ with $\epsilon_{i,j} \sim \log\mathcal{N}\left(0, \sigma^2_{i,j}\right)$ and $\sigma^2_{i,j} > 0$. $\epsilon$s are independent from each other and all $w$s.

**Narrow thinker's behavior.** Parallel with (23), each optimal labor supply $l_i^*\left(\omega_i\right)$ must satisfy $v'\left(l_i^*\left(\omega_i\right)\right) = w_i E_i\left[h'\left(y^*\right)\right].$ Similar to the main text, I then use a hat over a variable to denote its log-deviation from the point of log-linearization.[43] The optimal labor supply condition for each $i$ and the budget constraint then become

$$\kappa\hat{l}_i^*\left(\omega_i\right) = \hat{w}_i - \kappa_h E_i\left[\hat{y}^*\right], \tag{55}$$

$$\sum_{i=1}^{N}\mu_i\left(\hat{l}_i^*\left(\omega_i\right) + \hat{w}_i\right) = \hat{y}^*, \tag{56}$$

where $\mu_i = \frac{\bar{w}_i\bar{l}_i}{\bar{y}}$ 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 (20), for each $i$, I define the

[43] 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.

narrow thinker's (log) labor supply function as $\hat{l}_i^{\text{Narrow}}(\hat{w}_1, \cdots, \hat{w}_N) \equiv E\left[\hat{l}_i^*(\omega_i)\,|\hat{w}_1, \cdots, \hat{w}_N\right]$. 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.

**Corollary 8** *For each $i$, the narrow thinker's labor supply $l_i$ is smaller (larger) in response to positive (negative) shocks to $w_i$ :*

$$\frac{\partial \hat{l}_i^{\text{Narrow}}}{\partial \hat{w}_i} \leq \frac{\partial \hat{l}_i^{\text{Standard}}}{\partial \hat{w}_i}.$$

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 is dampened, and the narrow thinker's $l_i$ is smaller in response to the increase in $w_i$.

**Economic implications and testable predictions.** First, there are some potential testable differences between the narrow thinking theory of small wage elasticity of daily labor supply and the existing daily income targeting model, potentially micro-founded by loss aversion around the target in Farber (2015). As Farber (2015) points out, such model tends to predict negative wage elasticity of daily labor supply, which is inconsistent with the empirical evidence. The prediction under narrow thinking, however, can be consistent with the empirically documented positive, but small, wage elasticity of daily labor supply.

Second, in line with Proposition 6, 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.

Third, the within-person coordination friction driven by narrow thinking can decrease with the experience. More experience is akin to an increase in cognitive capacity $\tau$ in Section 6, facilitating the decision maker to coordinate her multiple selves. This is indeed consistent with the finding in Camerer et al. (1997) and Section 8 in Farber (2015) that wage elasticity of daily labor supply increases with the taxi driver's experience.

**Proof of Corollary 8.** In the proof, for notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. From (55) and (56), we have

$$l_i^*(\omega_i) = \frac{1 - \kappa_h \mu_i}{\kappa + \kappa_h \mu_i} w_i - \frac{\kappa_h}{\kappa + \kappa_h \mu_i} E_i \left[ \sum_{j \neq i} \mu_j \left( l_j^*(\omega_j) + w_j \right) \right].$$

The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\kappa_h}{\kappa+\kappa_h\mu_i}$, $\Gamma_i = \mu_i$ and $\Delta_i = \mu_i$. The result then follows from Proposition 4.

## Excess Sensitivity to Temporary Income Shocks.

One often mentioned form of over-reaction is excess sensitivity to temporary income shocks. As Stephens Jr and Unayama (2011), Parker (2017) and Kueng (2018) document, such excess sensitivity cannot be fully explained by the existence of liquidity constraints. As an example of such behavior, Thaler (1999) mentioned his own experience: he spent most of his speaking fee for a conference in Switzerland on fancy hotels and meals there. He said he would not spend so much without the speaking fee. Such behavior is inconsistent with the standard consumption smoothing behavior: the decision maker should, instead, increase her consumption by a small amount at different points in response to a temporary income shock. Under narrow thinking, however, consumption at other points in time may not be as responsive to the shock. As a result, the narrow thinker' consumption at the time of the income shock increases more.

To formalize, consider a decision maker whose utility is

$$\sum_{i=1}^{N} v_i\left(x_i\right) + h\left(y\right), \tag{57}$$

where $v_i\left(x\right)$ and $h\left(y\right)$ are defined similar to those in Section 5.2. The decision maker is subject to the budget constraint: $\sum_{i=1}^{N} x_i + y \le w + \sum_{i=1}^{N} w_i$, where $w$ is the decision maker's initial wealth (treat as a constant) and $w_i$ is the income earned by self $i$.

In this environment, each self $i \in \{1, \cdots, N\}$ should be interpreted as in charge of the consumption decision for a period of time.[44] Self $i$ perfectly knows $w_i \sim \log\mathcal{N}\left(\log \bar{w}_i, \sigma^2_{w_i}\right)$, the income she earns during that period. She receives a noisy signal about each of the other selves' $w_j$. Specifically, for $i \in \{1, \cdots, N\}$, self $i$'s information (signals) is given by $\omega_i = \left\{\{s_{i,j}\}_{j\in\{1,\cdots,N\}}\right\}$, where $s_{i,i} = w_i$ and, for $i \neq j$, $s_{i,j} = w_j \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \log\mathcal{N}\left(0, \sigma^2_{i,j}\right)$ and $\sigma^2_{i,j} > 0$. All $\epsilon$s and $w$s are independent from each other. As above, I always let the last self's decision $y$ be made with perfect knowledge about all the fundamentals and other decisions, which guarantees that the budget constraint is satisfied.

I use a hat over a variable to denote its log-deviation from the point of log-linearization.[45] Similar to conditions (11) and (20), for all $i$, I define the narrow thinker's (log) consumption function:

$$\hat{x}_i^{\text{Narrow}}\left(\hat{w}_1, \cdots, \hat{w}_N\right) \equiv E\left[\hat{x}_i^*\left(\omega_i\right) | \hat{w}_1, \cdots, \hat{w}_N\right] \quad \forall i, \tag{58}$$

[44] To determine the length of such a period here, one can also apply the cognitive inertia principle about the boundary of a self discussed above. For example, when a decision maker decides on her consumption in Switzerland, the income she earns in Switzerland, but not other incomes, is at the top of her mind.

[45] Specifically, I log linearize around the point where each $w_i$ is fixed at $\bar{w}_i$ and each decision is made with perfect knowledge of all fundamentals.

where $E[\cdot|\hat{w}_1, \cdots, \hat{w}_N]$ averages over the realization of noises in signals. 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.

**Corollary 9** *For each $i$, the narrow thinker's consumption $x_i$ increases (decreases) more in response to positive (negative) shocks to $w_i$ :*

$$\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{w}_i} > \frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{w}_i} > 0.$$

To understand the intuition behind the excess sensitivity, note that in standard consumer theory, an increase in $w_i$ will increase the consumption of both $x_i$ (the positive direct effect) and other consumption $x_j$. The increase in other consumption $x_j$ then decreases $x_i$ (the negative indirect effect). This scenario falls into the case that the indirect effect works in the opposite direction to the direct effect, and the dampening of the indirect effect under narrow thinking leads to over-reaction. Specially, under narrow thinking, other consumption $x_j$ increases less, $x_i$ increases more.

Finally, it is worth noting that, in Corollary 9, each self of the benchmark frictionless consumer (indexed by the superscript *Standard*) makes her consumption decision with perfect knowledge of all $w_i$s, including those earned by future selves. Corollary 9 is then designed to explain the empirical evidence that consumers exhibit excess sensitivity to *anticipated* temporary income shocks. One can also establish results about how narrow thinking generates excess sensitivity to *unanticipated* temporary income shocks (Hall and Mishkin, 1982; Jappelli and Pistaferri, 2014). There, the benchmark frictionless consumer has perfect recall, but does not have perfect knowledge about her future incomes.

**Proof of Corollary 9.** In the proof, for notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. We first derive the log-linearized optimal decision rule for each consumption $x_i^*(\omega_i)$ and the budget constraint:

$$-\kappa_i x_i^*(\omega_i) = -\kappa_h E_i[y^*],$$

$$\sum_{i=1}^{N} \mu_i^x x_i^*(\omega_i) + \mu_y y^* = \sum_{i=1}^{N} \mu_i^w w_i,$$

where $\kappa_h = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$, $\mu_i^x = \frac{\bar{x}_i}{w}$ and $\mu_i^w = \frac{\bar{w}_i}{w}$ are the spending and income share of self $i$, and $\mu_y = \frac{\bar{y}}{w}$ is the spending share of $y$. Substituting $y^*$ in the first expression based on the second expression, we have

$$x_i^*(\omega_i) = \frac{\frac{\kappa_h}{\kappa_i \mu_y}}{1 + \frac{\kappa_h \mu_i}{\kappa_i \mu_y}} E_i\left[\sum_{j=1}^{N} \mu_j^w w_j - \sum_{j \neq i} \mu_j x_j^*(\omega_j)\right].$$

The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\frac{\kappa_h}{\kappa_i \mu_y}}{1+\frac{\kappa_h \mu_i}{\kappa_i \mu_y}}$, $\Gamma_i = \mu_i$ and $\Delta_i = \mu_i^w$. The result then follows from Proposition 4.

## The Label Effect.

One often mentioned form of over-reaction is "the label effect:" consumption decisions can be sensitive to the label attached to the consumer's budget (Beatty et al., 2014, Benhassine et al., 2015, Abeler and Marklein, 2016, Hastings and Shapiro, 2017). For example, Beatty et al. (2014) study the UK Winter Fuel Payment program. Despite its label, the program 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 much more after receiving the Winter Fuel Payment than after receiving cash. Such behavior is inconsistent with standard consumer theory and violates the fungibility principle.

Here, I will show how narrow thinking can generate such excess sensitivity. Similar to Section 5.2 (consumer theory with income effects), consider a decision maker whose utility is

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

where $v_i(x)$ and $h(y)$ are defined as in Section 5.2. 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 initial wealth (treat as a constant) and $w_i$ is the money labelled for the consumption of good $i$. Note that under standard consumer theory (indexed by the superscript *Standard*, as above), consumption decisions only depend on the total wealth level, $w + \sum_{i=1}^{N} w_i$, independent from the labels. As above, the last self, who is in charge of the consumption of $y$, has perfect knowledge of all the fundamentals and other decisions. This makes sure that the budget constraint always holds.

Let me turn to the narrow thinker. Each self $i \in \{1, \cdots, N\}$ of the narrow thinker, who is in charge of purchasing good $i$, perfectly knows $w_i \sim \mathcal{N}\left(\bar{w}_i, \sigma_{w_i}^2\right)$, but receives a noisy signal about each of the other $w_j$: $s_{i,j} = w_j + \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \mathcal{N}\left(0, \sigma_{i,j}^2\right)$ and $\sigma_{i,j}^2 > 0$. All $\epsilon$s and $w$s are independent from each other. The information structure captures the idea that decision maker has the winter Fuel Payment at the top of her mind when purchasing fuel, but not necessarily when making other purchases.

Similar to conditions (11) and (20), for all $i$, I define the narrow thinker's demand function:[46] $x_i^{\text{Narrow}}(w_1, \cdots, w_N) \equiv E\left[x_i^*(\omega_i) | w_1, \cdots, w_N\right] \quad \forall i$. 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 the label effect under narrow thinking.

[46] For this application, I work with linearization instead of log-linearization, as the empirical evidence cited above focuses on the marginal propensity to spend instead of elasticities.

**Corollary 10** *1. The label effect: it is possible to find a pair $(i, j)$ such that*

$$\frac{\partial x_i^{\text{Narrow}}}{\partial w_i} \neq \frac{\partial x_i^{\text{Narrow}}}{\partial w_j}.$$

*2. Excess sensitivity: the narrow thinker's consumption $x_i$ increases (decreases) more in response to positive (negative) shocks to $w_i$ :*

$$\frac{\partial x_i^{\text{Narrow}}}{\partial w_i} > \frac{\partial x_i^{\text{Standard}}}{\partial w_i} > 0 \quad \forall i.$$

Part 1 of Corollary 10 shows how narrow thinking can generate the label effect. Different from the standard consumer theory, the label attached to each component of the total wealth level, $w + \sum_{i=1}^{N} w_i$, is relevant for consumption decisions. As different selves of the narrow thinker have different beliefs about shocks to each $w_i$ and hence different beliefs about the marginal value of money, the fungibility principle is violated and the label effect emerges.

Part 2 of Corollary 10 further establishes the excess sensitivity result. To understand the intuition behind the excess sensitivity, note that in standard consumer theory, an increase in $w_i$ will increase the consumption of both $x_i$ (the positive direct effect) and other consumption $x_j$. The increase in other consumption $x_j$ then decreases $x_i$ (the negative indirect effect). This scenario falls into the case that the indirect effect works in the opposite direction to the direct effect in Proposition 4, and narrow thinking leads to over-reaction.

**Proof of Corollary 10.** For this application, I work with linearization instead of log-linearization, as the empirical evidence cited in the main text focuses on the marginal propensity to spend instead of elasticities. In the proof, for notation simplicity, each variable denotes its deviation from the point in which each $w_i$ is fixed at its mean $\bar{w}_i$ . We first derive the linearized optimal decision rule for each consumption $x_i^*(\omega_i)$ and the budget constraint:

$$-\kappa_i \frac{x_i^*(\omega_i)}{\bar{x}_i} = -\kappa_h E_i\left[\frac{y^*}{\bar{y}}\right],$$

$$\sum_{i=1}^{N} p_i x_i^*(\omega_i) + y^* = \sum_{i=1}^{N} w_i,$$

where $\kappa_h = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$. Substituting $y^*$ in the first expression based on the second expression, we have

$$x_i^*(\omega_i) = \frac{\frac{\kappa_h \bar{x}_i}{\kappa_i \bar{y}}}{1 + \frac{\kappa_h \bar{x}_i}{\kappa_i \bar{y}} p_i} E_i\left[\sum_{j=1}^{N} w_j - \sum_{j \neq i} p_j x_j^*(\omega_j)\right].$$

The environment then falls into case 3) of Assumption 1, with $\rho_i = \frac{\frac{\kappa_h \bar{x}_i}{\kappa_i \bar{y}}}{1+\frac{\kappa_h \bar{x}_i}{\kappa_i \bar{y}} p_i}$, $\Gamma_i = p_i$ and $\Delta_i = 1$. Part 2 of the Corollary then follows from Proposition 4. To prove part 1, we can have explicit formula for own-price and cross-price sensitivities:

$$\frac{\partial x_i^{\text{Narrow}}}{\partial w_k} = \frac{\lambda_{i,k}\kappa_h \frac{\bar{x}_i}{\bar{y}}}{\kappa_i + \kappa_h \frac{\bar{x}_i}{\bar{y}} \left(1-\lambda_{i,k}\right) p_i} \frac{1}{\frac{\kappa_h p_k \bar{x}_k}{\kappa_k \bar{y}} + \sum_{j\neq k} \frac{\lambda_{j,k}\kappa_h \frac{p_j \bar{x}_j}{\bar{y}}}{\kappa_j + \kappa_h \left(1-\lambda_{j,k}\right) \frac{p_j \bar{x}_j}{\bar{y}}} + 1} \quad \forall i \neq k,$$

$$\frac{\partial x_k^{\text{Narrow}}}{\partial w_k} = \frac{\kappa_h \bar{x}_k}{\kappa_k \bar{y}} \frac{1}{\frac{\kappa_h p_k \bar{x}_k}{\kappa_k \bar{y}} + \sum_{j\neq k} \frac{\lambda_{j,k}\kappa_h \frac{p_j \bar{x}_j}{\bar{y}}}{\kappa_j + \kappa_h \left(1-\lambda_{j,k}\right) \frac{p_j \bar{x}_j}{\bar{y}}} + 1} > 0 \quad \forall k.$$

where $\lambda_{i,k} = \frac{\sigma_{w_k}^2}{\sigma_{w_k}^2 + \sigma_{i,k}^2}$. To prove part 1, note that, if and only if when all $\lambda$s are 1 (the standard counterpart), $\frac{\partial x_i^{\text{Narrow}}}{\partial w_k} = \frac{\partial x_i^{\text{Narrow}}}{\partial w_i}$ for all $i \neq k$.

## Temptation.

The environment is similar to the time management problem used for the study of comfort zones behavior. There, I study the response to shocks to an individual activity's attractiveness $\varphi_i$. Here I am interested in the impact of a common shock influencing the attractiveness of all activities. Importantly, here I allow different $\theta_k$s be correlated and the rational confusion channel discussed at the end of Section 3.

Specifically, the decision maker's utility is given by

$$\sum_{i=1}^{N} \varphi_i v\left(x_i\right) - c\left(\sum_{i=1}^{N} x_i\right), \tag{59}$$

where $x_i$ is the time the decision maker assigns to activity $i$, $\varphi_i v\left(x_i\right) = \varphi_i \frac{x_i^{1-\kappa}}{1-\kappa}$ with $\kappa > 0$ is her utility from activity $i$, $\varphi_i$ parametrizes the attractiveness of activity $i$, $c\left(\sum_{i=1}^{N} x_i\right)$ is the opportunity cost of time, and $c\left(x\right) = \frac{x^{1+\kappa_c}}{1+\kappa_c}$ with $\kappa_c > 0$ is a strictly convex function.

As I am interested in the impact of the common shock, I let the stochastic property of shocks and the information structure be symmetric across each $i$. Specifically, I let the attractiveness of each activity $i$ have an idiosyncratic and a common component: $\hat{\varphi}_i = \hat{\varphi} + \delta_i$, where $\hat{\varphi} \sim \mathcal{N}\left(0, \sigma_\varphi^2\right)$, $\delta_i \sim \mathcal{N}\left(0, \sigma_\delta^2\right)$ and they are independent from each other.[47] Similar to the information structure considered throughout, each self $i$ perfectly knows her own $\hat{\varphi}_i$, and receives a noisy signal about each of the other $\hat{\varphi}_j : s_{i,j} = \hat{\varphi}_j + \epsilon_{i,j} \;\; \forall j \neq i$. Noise $\epsilon_{i,j} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$ is independent from the fundamental and each other.[48]

Let me put the environment in a concrete setting. Consider that each self $i$ is in charge of how long she will play computer games on day $i$, $x_i$. Her utility from playing on day $i$ is captured by $\varphi_i v\left(x_i\right)$, where $\varphi_i$ parametrizes the attractiveness of playing computer games on day $i$, and $c\left(\sum_{i=1}^{N} x_i\right)$ captures the cost of playing computer games. Each self $i$ perfectly knows how attractive it is to play on day $i$, but only receives noisy signals about how attractive it is to play on other days. Now, a new computer game is introduced,

[47] As above, a hat over a variable to denote its log-deviation from the point of log-linearization.

[48] As different $\varphi_i$s are correlated, there is room for rational confusion. For example, self $i$ can use the attractiveness of her own activity, $\varphi_i$, to predict the attractiveness of other activities other $\varphi_j$s, which she does not perfectly know. Such motif is taken into consideration in the proof of Proposition 9.

and it generates a common shock increasing all $\varphi_i$. I am then interested in how the narrow thinker responds to the introduction of the new game.

Similar to (54), I define the narrow thinker's (log) decision as a function of the fundamentals as $\hat{x}_i^{\text{Narrow}}(\hat{\varphi}_1, \cdots, \hat{\varphi}_N) \equiv E[\hat{x}_i^*(\omega_i)|\hat{\varphi}_1, \cdots, \hat{\varphi}_N]$. I then study $\frac{d\hat{x}_i^{\text{Narrow}}}{d\hat{\varphi}} = lim_{\hat{\varphi}\to 0}\frac{\hat{x}_i^{\text{Narrow}}(\hat{\varphi},\cdots,\hat{\varphi}) - \hat{x}_i^{\text{Narrow}}(0,\cdots,0)}{\hat{\varphi}}$, which summarizes each decision $i$'s response to the common shock. 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 then have:

**Proposition 9** *For each $i$, the narrow thinker increases (decreases) her time allocated for activity $i$ more in response to positive (negative) common taste shocks $\varphi$ :*

$$\frac{d\hat{x}_i^{\text{Narrow}}}{d\hat{\varphi}} > \frac{d\hat{x}_i^{\text{Standard}}}{d\hat{\varphi}} > 0 \quad \forall i.$$

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, her belief about how other $x_j$s respond to the common shock is anchored. 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.[49]

To further illustrate Proposition 9, consider the above example where a new computer game is introduced. When the narrow thinker decides how long she will play on a particular day, her belief about playing time on other days is anchored. As a result, she will play longer on that particular day. In this sense, the decision maker is tempted by the new computer game. Such a prediction also connects to the neglect of the "adding-up effects" in Read, Loewenstein and Rabin (1999). In this setting, the cost from playing the computer game on a single day is low. However, the cumulative costs can be large (e.g. opportunities costs and eye damage), and increase faster than the cumulative benefits (due to the convex cost function). The narrow thinker, who underestimates how long she will play on other days, then also underestimates the "adding-up" costs.

The temptation motive predicted by narrow thinking is particularly pronounced in response to a new stimuli. This differs from to the prediction based on self-control (Laibson, 1997, O'Donoghue and Rabin,

[49] In the case of the common shock, if different selves' decisions are strategic substitutes (as here), the direct effect and the indirect effect of the common shock on each $x_i$ are in opposite directions. Narrow thinking generates over-reaction. If different selves' decisions are strategic complements, the direct effect and the indirect effect of the common shock on each $x_i$ are in the same direction. Narrow thinking generates under-reaction. However, such relationship between strategic complementarity/substitutability and under-/over- reaction under narrow thinking only hold in response to the common shock. If the shock is idiosyncratic, as the case in Proposition 7, the indirect effect of $\varphi_i$ on $x_i$ is from a second order interaction ($\varphi_i$ to $x_j$ then to $x_i$). As a result, under-reaction can arise in the case of strategic substitutability.

1999, Gul and Pesendorfer, 2001, Fudenberg and Levine, 2006). Moreover, such prediction can also explain the supply side of the temptation good production. As the decision maker is particularly tempted to new attractions, the computer game company always has incentives to develop new versions of their products.

**Proof of Proposition 9.** 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

$$x_i^*(\omega_i) = E_i\left[\psi\varphi_i - \gamma\sum_{j\neq i}\frac{x_j^*(\omega_j)}{N}\right], \tag{60}$$

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

Given the information structure, we have

$$E_i[\varphi_i] = \varphi_i \quad \forall i, \tag{61}$$

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

$$\begin{aligned} E_i[\varphi_j] &= E_i[\varphi] + \frac{\sigma_\epsilon^{-2}}{\sigma_\delta^{-2}+\sigma_\epsilon^{-2}}\left(s_{i,j} - E_i[\varphi]\right) \\ &\equiv \lambda s_{i,j} + \mu\varphi_i + \omega\sum_{l\neq i,j}s_{i,l} \quad \forall i\neq j, \end{aligned} \tag{62}$$

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

Similar to the proof of Proposition 3, as the optimal decision rule (60) is linear and all variables are distributed Normally, for all $i$, $x_i^*(\omega_i)$ is linear in its signal and $x_i^L(\varphi_1,\cdots,\varphi_N)$ is linear in all $\varphi$s. From condition (60) and the fact that the noise in each self's private signal is not predictable, we have

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

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

Using (61) and (62), averaging across noises in the realizations of signals, and taking partial derivatives with respect to each $\theta_j$, we have

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} = \psi - \gamma \sum_{j \neq i} \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_i} - \underbrace{\mu\gamma \sum_{j \neq i} \sum_{l \neq i} \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_l}}_{\text{rational confusion}} \quad \forall i,$$

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} = -\lambda\gamma \sum_{j \neq i} \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_k} - \underbrace{\omega\gamma \sum_{j \neq i} \sum_{l \neq k,i} \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_l}}_{\text{rational confusion}} \quad \forall i, k.$$

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,

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

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

Collecting terms, we have

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} = \frac{\psi}{1 + \mu\gamma (N-1) - \frac{\gamma^2 (N-1)(1+\mu(N-2))(\lambda+\omega(N-2))}{\left(1+\lambda\gamma(N-2)+\omega\gamma(N-2)^2\right)}}$$

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} = -\frac{\gamma (\lambda + \omega (N-2))}{(1 + \gamma (N-2)(\lambda + \omega (N-2)))} \frac{\partial x_i^L}{\partial \varphi_i}.$$

Based on the definition of $\frac{dx_i^{\text{Narrow}}}{d\varphi}$, we then have

$$\begin{aligned}
\frac{dx_i^{\text{Narrow}}}{d\varphi} &= \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} + (N-1) \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} = \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_i} \left( 1 - \frac{\gamma (\lambda + \omega (N-2))(N-1)}{(1 + \gamma (N-2)(\lambda + \omega (N-2)))} \right) \\
&= \frac{\psi}{1 + \mu\gamma (N-1) - \frac{\gamma^2 (N-1)(1+\mu(N-2))(\lambda+\omega(N-2))}{\left(1+\lambda\gamma(N-2)+\omega\gamma(N-2)^2\right)}} \frac{1 + \gamma (N-2)(\lambda + \omega (N-2)) - \gamma (\lambda + \omega (N-2))(N-1)}{(1 + \gamma (N-2)(\lambda + \omega (N-2)))} \\
&= \frac{\psi (1 - \gamma (\lambda + \omega (N-2)))}{(1 + \mu\gamma (N-1))(1 + \gamma (N-2)(\lambda + \omega (N-2))) - \gamma^2 (N-1)(1 + \mu (N-2))(\lambda + \omega (N-2))} \\
&= \frac{\psi (1 - \gamma (\lambda + \omega (N-2)))}{1 + \mu\gamma (N-1) + \gamma (N - 2 - \gamma (N-1))(\lambda + \omega (N-2))}.
\end{aligned}$$

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

$$\begin{aligned}
\frac{dx_i^{\text{Narrow}}}{d\varphi} &> \frac{\psi (1 - \gamma t)}{1 + \gamma (N-1)(1-t) + \gamma (N - 2 - \gamma (N-1)) t} \\
&= \frac{\psi}{1 + \gamma (N-1)} = \frac{dx_i^S}{d\varphi}.
\end{aligned}$$

# Appendix E: Endogenous Narrow Thinking: Costly Contemplation

## Costly Contemplation for the Symmetric N-decisions Problem.

*Environment.* Consider the simple abstract example in (18). Let the decision maker's utility be symmetric, given by $u\left(x_1,\ldots,x_N,\vec{\theta}\right)=-\frac{1}{2}\sum_{i=1}^{N}\left(x_i-\theta_i\right)^2+\sum_{i\neq j}\gamma x_i x_j$, where $\gamma\equiv\gamma_{i,j}\in\left(-\frac{1}{N-1},\frac{1}{N-1}\right)$ for all $i\neq j$. At the information side, I do not directly impose that each self $i$ has perfect knowledge of $\boldsymbol{\theta_i}$ and receives a noisy signal of each of the other $\boldsymbol{\theta_j}$ as in Section 3. Instead, I let the decision maker choose endogenously the precision of each self's signal. Specifically, each potential signal $\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}$ for self $i$ consists of $N$ noisy signals, one for each $\boldsymbol{\theta_j}$ : $\boldsymbol{s_{i,j}}=\boldsymbol{\theta_j}+\boldsymbol{\epsilon_{i,j}}$. All $\boldsymbol{\epsilon}$s and $\boldsymbol{\theta}$s are independent from each other, but the exact variances of noises in these signals are free to choose, subject to the cognitive constraint in (29).

**Proposition 10** *In the optimum of the costly contemplation problem in (28), self i's signal about $\boldsymbol{\theta_i}$ is more precise than other selves' signal about $\boldsymbol{\theta_i}$ :*

$$\left(\sigma_{i,i}^*\right)^2<\left(\sigma_{j,i}^*\right)^2\quad\forall i\neq j,$$

*where $\sigma_{j,i}^*$ is the variance of the noise of self j's signal about $\boldsymbol{\theta_i}$ in the optimum.*

*Proof.* As discussed in the main text, the costly contemplation problem in (28) can be divided into two subproblems, the optimal information choice subject to the cognitive constraint in (29), and the optimal decisions *given* the chosen information. From Lemma 2, given any chosen information $\{\boldsymbol{\omega_i}\}_{i=1}^N$, the optimal decision rule $\{x_i^*(\cdot)\}_{i=1}^N$ can be characterized by

$$E\left[\frac{\partial u}{\partial x_i}\left(x_i^*\left(\omega_i\right),x_{-i}^*\left(\omega_{-i}\right),\vec{\theta}\right)|\omega_i\right]=0\quad\forall i,\omega_i. \tag{63}$$

Using law of iterated expectations, we have

$$\frac{1}{2}E\left[x_i^*\left(\omega_i\right)\frac{\partial u}{\partial x_i}\left(x_i^*\left(\omega_i\right),x_{-i}^*\left(\omega_{-i}\right),\vec{\theta}\right)\right]=0\quad\forall i.$$

Substituting into the decision maker's utility function, the optimal information choice in (28) is then equivalent to

$$\max_{\{\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}\}_{i=1}^{N}} \frac{1}{2}E\left[\sum_{i=1}^{N}\theta_i x_i^*\left(\omega_i\right)\right] \tag{64}$$
$$s.t.\ \ \forall i\ \ x_i^*\left(\omega_i\right)\ \text{satisfy (63)}$$
$$\sum_{i=1}^{N} I\left(\boldsymbol{\omega_i};\vec{\boldsymbol{\theta}}\right)\leq\tau.$$

Now, note that any $\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}$ takes the form of $\boldsymbol{\omega_i}=\{\boldsymbol{s_{i,1}},\cdots,\boldsymbol{s_{i,N}}\}$ where $\boldsymbol{s_{i,j}}=\boldsymbol{\theta_j}+\boldsymbol{\epsilon_{i,j}}$ with $\boldsymbol{\epsilon_{i,j}}\sim N\left(0,\sigma_{i,j}^2\right)$ and all $\boldsymbol{\epsilon}$s and $\boldsymbol{\theta}$s are independent from each other. Similar to the proof of Proposition 3, we have

$$\begin{pmatrix}\frac{\partial x_1^{\text{Narrow}}}{\partial\theta_j}\\ \frac{\partial x_2^{\text{Narrow}}}{\partial\theta_j}\\ \cdots\\ \\ \frac{\partial x_N^{\text{Narrow}}}{\partial\theta_j}\end{pmatrix}=\left(\mathbb{I}_N-\begin{pmatrix}1 & \lambda_{1,j} & \cdots & \lambda_{1,j} & \lambda_{1,j}\\ \lambda_{2,j} & 1 & \cdots & \lambda_{2,j} & \lambda_{2,j}\\ & & \cdots & & \\ \\ \lambda_{N,j} & \lambda_{N,j} & \cdots & \lambda_{N,j} & 1\end{pmatrix}\circ\Gamma\right)^{-1}\begin{pmatrix}0\\ \cdots\\ \lambda_{j,j}\\ \cdots\\ 0\end{pmatrix},$$

where $\lambda_{i,j}=\frac{\sigma_{\theta_j}^2}{\sigma_{\theta_j}^2+\sigma_{i,j}^2}\in(0,1]$ and $x_i^{\text{Narrow}}\left(\vec{\theta}\right)=E_i\left[x_i^*\left(\omega_i\right)|\vec{\theta}\right]$. I then use the Sherman–Morrison formula[50] (for matrix inversion), the fact that all $\epsilon$s and $\theta$s are independent from each other, and $I\left(\boldsymbol{\omega_i^*};\vec{\boldsymbol{\theta}}\right)=$

---

[50] In fact, $\begin{pmatrix}\frac{\partial x_1^{\text{Narrow}}}{\partial\theta_j}\\ \frac{\partial x_2^{\text{Narrow}}}{\partial\theta_j}\\ \cdots\\ \\ \frac{\partial x_N^{\text{Narrow}}}{\partial\theta_j}\end{pmatrix}=\left(\begin{pmatrix}\lambda_{1,j}^{-1} & 1 & \cdots & 1 & 1\\ 1 & \lambda_{2,j}^{-1} & \cdots & 1 & 1\\ & & \cdots & & \\ \\ 1 & 1 & \cdots & 1 & \lambda_{N,j}^{-1}\end{pmatrix}\circ\left(\mathbb{I}_N-\Gamma\right)\right)^{-1}\begin{pmatrix}0\\ \cdots\\ 1\\ \cdots\\ 0\end{pmatrix}$. Then note $\begin{pmatrix}\lambda_{1,j}^{-1} & 1 & \cdots & 1 & 1\\ 1 & \lambda_{2,j}^{-1} & \cdots & 1 & 1\\ & & \cdots & & \\ \\ 1 & 1 & \cdots & 1 & \lambda_{N,j}^{-1}\end{pmatrix}\circ\left(\mathbb{I}_N-\Gamma\right)=-\gamma J+diag\left\{\lambda_{1,j}^{-1}+\gamma,\lambda_{2,j}^{-1}+\gamma,\cdots,\lambda_{N,j}^{-1}+\gamma\right\}$, where $J=u'u$ and $u=\begin{pmatrix}1\\ \\ \cdots\\ \\ 1\end{pmatrix}$ is a $N\times 1$ vector. One can then use the Sherman–Morrison formula to calculate the matrix inverse.

$\frac{1}{2}\sum_{1\leq j\leq N} log_2\left(\frac{1}{1-\lambda_{i,j}}\right)$. The problem in (64) becomes

$$\max_{\{0\leq\lambda_{i,j}\leq 1\}_{1\leq i,j\leq N}} g\left(\{\lambda_{i,j}\}_{1\leq i,j\leq N}\right) \equiv \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\lambda_{i,i}}{1+\gamma\lambda_{i,i}} + \frac{\gamma\left(\frac{\lambda_{i,i}}{1+\gamma\lambda_{i,i}}\right)^2}{1-\gamma\sum_{1\leq j\leq N}\left(\frac{\lambda_{j,i}}{1+\gamma\lambda_{j,i}}\right)}\right)\sigma^2_{\theta_i} \tag{65}$$

$$s.t.\ h\left(\{\lambda_{i,j}\}_{1\leq i,j\leq N}\right) \equiv \sum_{1\leq i,j\leq N}\frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,j}}\right) \leq \tau.$$

Now we prove Proposition *10*. If in the optimum of the costly contemplation problem, we have $\left(\sigma^*_{x,x}\right)^2 \geq \left(\sigma^*_{y,x}\right)^2$ for a pair $x \neq y$. This means $\lambda^*_{x,x} \leq \lambda^*_{y,x}$, where $\lambda^*_{i,j} = \frac{\sigma^2_{\theta_j}}{\sigma^2_{\theta_j}+\left(\sigma^*_{i,j}\right)^2} \in (0,1]\quad \forall i,j$. If $\lambda^*_{x,x} < \lambda^*_{y,x}$, then consider $\left\{\lambda'_{i,j}\right\}_{1\leq i,j\leq N}$ where $\lambda'_{x,x} = \lambda^*_{y,x}$, $\lambda'_{y,x} = \lambda^*_{x,x}$ and, for other $(i,j)$, $\lambda'_{i,j} = \lambda_{i,j}$. $\left\{\lambda'_{i,j}\right\}_{1\leq i,j\leq N}$ increases the objective in (65) without changing the constraint and leads to a contradiction. If $\lambda^*_{x,x} = \lambda^*_{y,x}$, we have $\frac{\partial g\left(\{\lambda^*_{i,j}\}_{1\leq i,j\leq N}\right)}{\partial\lambda^*_{x,x}} > \frac{\partial g\left(\{\lambda^*_{i,j}\}_{1\leq i,j\leq N}\right)}{\partial\lambda^*_{y,x}}$ and $\frac{\partial h\left(\{\lambda^*_{i,j}\}_{1\leq i,j\leq N}\right)}{\partial\lambda^*_{x,x}} = \frac{\partial h\left(\{\lambda^*_{i,j}\}_{1\leq i,j\leq N}\right)}{\partial\lambda^*_{y,x}}$. This is inconsistent the first order condition in (65):

$$\frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq N}\right)}{\partial\lambda^*_{x,x}} / \frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq N}\right)}{\partial\lambda^*_{x,x}} = \frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq N}\right)}{\partial\lambda^*_{y,x}} / \frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq N}\right)}{\partial\lambda^*_{y,x}}.$$

As a result, $\left(\sigma^*_{i,i}\right)^2 < \left(\sigma^*_{j,i}\right)^2 \quad \forall i \neq j$.

## Arbitrary N-decisions Case, a Limit Result.

*Environment.* Consider the same utility as in the simple abstract example in (18) and the above subsection in Appendix. The only difference from the above subsection is that I allow $u$ to be asymmetric. At the information side, I do not directly impose that each self $i$ has perfect knowledge of $\boldsymbol{\theta_i}$ and receives a noisy signal of each of the other $\boldsymbol{\theta_j}$ as in Section 3. Instead, I let the decision maker choose endogenously the precision of each self's signal. Specifically, each potential signal $\boldsymbol{\omega_i} \in \boldsymbol{\Omega_i}$ for self $i$ consists of $N$ noisy signals, one for each $\boldsymbol{\theta_j}$ : $\boldsymbol{s_{i,j}} = \boldsymbol{\theta_j} + \boldsymbol{\epsilon_{i,j}}$. All $\boldsymbol{\epsilon}$s and $\boldsymbol{\theta}$s are independent from each other, but the exact variances of noises in these signals are free to choose, subject to the cognitive constraint in (29).

**Proposition 11** *There exists $\bar{\tau} > 0$, such that when the cognitive capacity $\tau \leq \bar{\tau}$, it is optimal for each decision $i$ to be only based on its local fundamental about $\boldsymbol{\theta_i}$. That is, when cognitive capacity is small enough, in the optimum of the costly contemplation problem in (28),*

$$\sigma^*_{i,i} < \infty \quad and \quad \sigma^*_{i,j} = \infty \quad \forall i \neq j,$$

*where $\sigma^*_{i,j}$ is the variance of the noise of self $i$'s signal about $\boldsymbol{\theta_j}$ in the optimum.*

*Proof.* Similar to the proof of Proposition 10, the costly contemplation problem in (28) is equivalent to

$$\max_{\{\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}\}_{i=1}^N} \frac{1}{2}E\left[\sum_{i=1}^N \theta_i x_i^*(\omega_i)\right] \tag{66}$$

$$s.t.\ \forall i\ \ x_i^*(\omega_i)\ \text{ satisfy (63)}$$

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

Now, note that any $\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}$ takes the form of $\boldsymbol{\omega_i}=\{\boldsymbol{s_{i,1}},\cdots,\boldsymbol{s_{i,N}}\}$ where $\boldsymbol{s_{i,j}}=\boldsymbol{\theta_j}+\boldsymbol{\epsilon_{i,j}}$ with $\boldsymbol{\epsilon_{i,j}}\sim N\left(0,\sigma_{i,j}^2\right)$ and all $\boldsymbol{\epsilon}$s and $\boldsymbol{\theta}$s are independent from each other. Similar to the proof of Proposition 3, we have

$$\begin{pmatrix}\frac{\partial x_1^{\text{Narrow}}}{\partial\theta_j}\\ \frac{\partial x_2^{\text{Narrow}}}{\partial\theta_j}\\ \cdots\\ \\ \frac{\partial x_N^{\text{Narrow}}}{\partial\theta_j}\end{pmatrix}=\left(\mathbb{I}_N-\begin{pmatrix}1&\lambda_{1,j}&\cdots&\lambda_{1,j}&\lambda_{1,j}\\ \lambda_{2,j}&1&\cdots&\lambda_{2,j}&\lambda_{2,j}\\ &&\cdots&&\\ \\ \lambda_{N,j}&\lambda_{N,j}&\cdots&\lambda_{N,j}&1\end{pmatrix}\circ\Gamma\right)^{-1}\begin{pmatrix}0\\ \cdots\\ \lambda_{j,j}\\ \cdots\\ 0\end{pmatrix}, \tag{67}$$

where $\lambda_{i,j}=\frac{\sigma_{\theta_j}^2}{\sigma_{\theta_j}^2+\sigma_{i,j}^2}\in(0,1]$, and $x_i^{\text{Narrow}}\left(\vec{\theta}\right)=E_i\left[x_i^*(\omega_i)\,|\vec{\theta}\right]$. As different $\boldsymbol{\theta}$s are independent, the problem in (66) is then equivalent to

$$\max_{\{0\leq\lambda_{i,j}\leq1\}_{1\leq i,j\leq N}} g\left(\{\lambda_{i,j}\}_{1\leq i,j\leq N}\right)\equiv\frac{1}{2}\sum_{i=1}^N\left(\frac{\partial x_i^{\text{Narrow}}}{\partial\theta_i}\sigma_{\theta_i}^2\right) \tag{68}$$

$$s.t.\ \ \frac{\partial x_i^{\text{Narrow}}}{\partial\theta_i}\ \text{ is from (67)}$$

$$h\left(\{\lambda_{i,j}\}_{1\leq i,j\leq N}\right)\equiv\sum_{1\leq i,j\leq N}\frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,j}}\right)\leq\tau. \tag{69}$$

Now, from the Cramer's rule, we know, for all $i$, $\frac{\partial x_i^{\text{Narrow}}}{\partial\theta_i}=\lambda_{i,i}\frac{1+P_{i,1}\left(\{\lambda_{j,i}\}_{j=1}^N\right)}{1+P_{i,2}\left(\{\lambda_{j,i}\}_{j=1}^N\right)}$, where $P_{i,1}$ and $P_{i,2}$ are polynomials (without constant) capturing first and higher order terms of $\{\lambda_{j,i}\}_{j=1}^N$. Further note that from the constraint (69), for all $i,j$, $\lim_{\tau\to0}\lambda_{i,j}=0$. We then have, for all $i$ and $j\neq i$,

$$\lim_{\tau\to0}\frac{\partial g}{\partial\lambda_{i,i}}=\frac{1}{2}\sigma_{\theta_i}^2\quad\text{and}\quad\lim_{\tau\to0}\frac{\partial g}{\partial\lambda_{i,j}}=0$$

$$\lim_{\tau\to0}\frac{\partial h}{\partial\lambda_{i,i}}=\frac{1}{2\ln2}\quad\text{and}\quad\lim_{\tau\to0}\frac{\partial h}{\partial\lambda_{i,j}}=\frac{1}{2\ln2}.$$

This then proves Proposition 11 and means that, when the cognitive capacity is small enough, it is optimal for each decision $i$ to be only based on information about $\boldsymbol{\theta_i}$.

## Endogenous Shock-Specific Coordination Friction.

As discussed in Section 3, the degree of effective attenuation of interaction under narrow thinking is shock specific (e.g. the interaction matrix $\Gamma_k^{'}$ in Proposition 3 depends on $k$). In this part of the appendix, I show how such shock-specific effective attenuation of interactions can arise endogenously in the costly contemplation problem. I further find conditions characterizing in response to which shocks the decision maker chooses to better coordinate on.

*Environment.* I use the 2-decisions case in Subsection *6.2* as an example. Similar results can be established for symmetric $N$-goods cases.

**Proposition 12** *In the optimum of the costly contemplation problem in (28), when $\sigma_{\theta_1}^2 > \sigma_{\theta_2}^2$, we have*

$$\lambda_{1,1}^* > \lambda_{2,2}^* \quad \textit{and} \quad \lambda_{2,1}^* > \lambda_{1,2}^*,$$

*where $\lambda_{i,j}^* = \frac{\sigma_{\theta_j}^2}{\sigma_{\theta_j}^2 + \left(\sigma_{i,j}^*\right)^2}$, and $\sigma_{i,j}^*$ is the variance of the noise of self $i$'s signal about $\boldsymbol{\theta_j}$ in the optimum. Similarly, when $\sigma_{\theta_1}^2 < \sigma_{\theta_2}^2$, we have*

$$\lambda_{1,1}^* < \lambda_{2,2}^* \quad \textit{and} \quad \lambda_{2,1}^* < \lambda_{1,2}^*.$$

To illustrate, consider the case that $\sigma_{\theta_1}^2 > \sigma_{\theta_2}^2$, which means that the first decision's local fundamental is more volatile (high $\sigma_{\theta_1}^2$) than the second decision's local fundamental. In this case, Proposition 12 means that: first, self 1's signal about her own local fundamental $\theta_1$ is more precise than self 2's signal about her own local fundamental $\theta_2$; second, self 2's signal about $\theta_1$ is more precise than self 1's signal about $\theta_2$.[51] As a result, the decision maker chooses to better coordinate on shocks to $\theta_1$.

*Proof.* Following the proof of Proposition 7, $\left\{\lambda_{i,j}^*\right\}_{1\le i,j\le 2}$ must solve:

$$\max_{\{0\le\lambda_{i,j}\le 1\}_{1\le i,j\le 2}} g\left(\{\lambda_{i,j}\}_{1\le i,j\le 2}\right) \equiv \frac{1}{2}\frac{\lambda_{1,1}}{1-\lambda_{1,1}\lambda_{2,1}\gamma^2}\sigma_{\theta_1}^2 + \frac{1}{2}\frac{\lambda_{2,2}}{1-\lambda_{2,2}\lambda_{1,2}\gamma^2}\sigma_{\theta_2}^2 \tag{70}$$

$$s.t.\ h\left(\{\lambda_{i,j}\}_{1\le i,j\le 2}\right) \equiv \sum_{1\le i,j\le 2}\frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,j}}\right) \le \tau. \tag{71}$$

When $\sigma_{\theta_1}^2 > \sigma_{\theta_2}^2$, we must have

$$\frac{\lambda_{1,1}^*}{1-\lambda_{1,1}^*\lambda_{2,1}^*\gamma^2} \ge \frac{\lambda_{2,2}^*}{1-\lambda_{2,2}^*\lambda_{1,2}^*\gamma^2}. \tag{72}$$

Otherwise one could let $\left\{\lambda_{i,j}^{'} = \lambda_{-i,-j}^*\right\}_{1\le i,j\le 2}$, which improve the objective in (70) while maintaining the constraint in (71).

[51] Note the difference from Proposition 7, which compares $\lambda_{1,1}^*$ to $\lambda_{2,1}^*$ (and $\lambda_{1,2}^*$ to $\lambda_{2,2}^*$), i.e. the precision of self 1's signal about $\theta_1$ and the precision of self 2's signal about $\theta_1$.

One can then prove $\lambda^*_{1,1} > \lambda^*_{2,2}$. If $\lambda^*_{1,1} \le \lambda^*_{2,2}$, from (72), it must be that $1/\left[1-\lambda^*_{1,1}\lambda^*_{2,1}\gamma^2\right]^2 \ge 1/\left[1-\lambda^*_{1,2}\lambda^*_{2,2}\gamma^2\right]^2$. We then have

$$\begin{aligned}\frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\le i,j\le 2}\right)}{\partial \lambda^*_{1,1}} &= \frac{1}{\left[1-\lambda^*_{1,1}\lambda^*_{2,1}\gamma^2\right]^2}\sigma^2_{\theta_1} \\ &> \frac{1}{\left[1-\lambda^*_{1,2}\lambda^*_{2,2}\gamma^2\right]^2}\sigma^2_{\theta_2} = \frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\le i,j\le 2}\right)}{\partial \lambda^*_{2,2}}.\end{aligned}$$

However, $\frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\le i,j\le 2}\right)}{\partial \lambda^*_{1,1}} \le \frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\le i,j\le 2}\right)}{\partial \lambda^*_{2,2}}$. This is contradictory to the of FOC the problem in (70).

One can further prove $\lambda^*_{2,1} > \lambda^*_{1,2}$. This comes from the fact that

$$\begin{aligned}\frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\le i,j\le 2}\right)}{\partial \lambda^*_{2,1}} &= \left[\frac{\lambda^*_{1,1}}{1-\lambda^*_{1,1}\lambda^*_{2,1}\gamma^2}\right]^2\sigma^2_{\theta_1} \\ &> \left[\frac{\lambda^*_{2,2}}{1-\lambda^*_{1,2}\lambda^*_{2,2}\gamma^2}\right]^2\sigma^2_{\theta_2} = \frac{\partial g\left(\left\{\lambda^*_{i,j}\right\}_{1\le i,j\le 2}\right)}{\partial \lambda^*_{1,2}},\end{aligned}$$

and $h$ is convex in each $\lambda$. This finishes the proof for the case of $\sigma^2_{\theta_1} > \sigma^2_{\theta_2}$. The proof for the case of $\sigma^2_{\theta_2} > \sigma^2_{\theta_1}$ is similar.

## The General Case: Flexible Information Acquisition

**Environment.** In Section 6.2, I restrict each potential $\boldsymbol{\omega_i} \in \boldsymbol{\Omega_i}$ to have a particular form: each $\boldsymbol{\omega_i}$ consists of $N$ noisy signals, one for each $\boldsymbol{\theta_k}$, $k \in \{1, \cdots, N\}$. 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 $\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \cdots, \boldsymbol{\omega_N}$ are always conditionally independent given $\vec{\boldsymbol{\theta}}$. That is, the noise in each decision $i$'s signal about $\vec{\boldsymbol{\theta}}$ is idiosyncratic.[52]

With such flexible form of information acquisition, I can achieve a sharp characterization about the optimum of the costly contemplation problem (28) in the general multiple-decision setting.

Specifically, in this subsection, I allow arbitrary concave and quadratic utility functions $u$, and arbitrarily Normally correlated fundamentals $\vec{\boldsymbol{\theta}}$. For notation simplicity, I normalize the mean of $\vec{\boldsymbol{\theta}}$ to be $\vec{0}$. Without

[52]This follows the literature on information acquisition in games (e.g. Yang, 2015, Morris and Yang, 2016). 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's signals will always be different because of the idiosyncratic noise. Nevertheless, this section focuses on how different decisions' signals take different *forms*.

loss of generality, I also restrict that $u$ does not have terms which are linear functions of $\vec{x}$. Such terms will only add a constant to each optimal decision rule.

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

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

**Lemma 5** *With unrestricted $\boldsymbol{\Omega_i}$, in the optimum of the costly contemplation problem in (28), each decision $i$ is based on an one-dimensional signal $\boldsymbol{s_i^*}$ :*[53]

$$\boldsymbol{\omega_i^*} = \{\boldsymbol{s_i^*}\} \quad \textit{and} \quad \boldsymbol{s_i^*} = \boldsymbol{\vartheta_i} + \mathbb{E}\left[\sum_{j\neq i} \gamma_{i,j} x_j^*\left(\boldsymbol{\omega_j^*}\right) | \vec{\boldsymbol{\theta}}\right] + \boldsymbol{\epsilon_i} \equiv \boldsymbol{t_i} + \boldsymbol{\epsilon_i}. \tag{73}$$

In (73), $\boldsymbol{\epsilon_i} \sim N\left(0, \sigma_i^2\right)$ is the idiosyncratic noise in the signal, $\sigma_i^2$ is pinned down by $\frac{1}{2}\log_2\left(\frac{\sigma_i^2 + \sigma_{t_i}^2}{\sigma_i^2}\right) = \tau_i^*$, $\sigma_{t_i}^2$ is the variance of $\boldsymbol{t_i}$ defined in (73), and $\boldsymbol{\vartheta_i}$ is a linear function of $\vec{\boldsymbol{\theta}}$ that summarizes how the fundamental directly influences optimal decision $i$, holding other decisions fixed.[54]

Without the cognitive constraint, the optimal decision $i$ is given by $\boldsymbol{\vartheta_i} + \sum_{j\neq i}\gamma_{i,j}x_j^*\left(\boldsymbol{\omega_j^*}\right)$. Now, with limited cognitive capacity, Lemma 5 shows that the optimal information for decision $i$ will be given by a signal about the fundamental $\vec{\boldsymbol{\theta}}$ that is closest to $\boldsymbol{\vartheta_i} + \sum_{j\neq i}\gamma_{i,j}x_j^*\left(\boldsymbol{\omega_j^*}\right)$. The variance of the noise in this signal is pinned down by decision $i'$s allocated cognitive capacity $\tau_i^*$.

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 $\boldsymbol{s_i^*}$ depends on the fundamental $\vec{\boldsymbol{\theta}}$.

[53]For each $i$, the optimal signal $\boldsymbol{s_i^*}$ is unique up to a linear transformation. That is, from an informational perspective, $\boldsymbol{s_i^*}$ is equivalent to $\alpha\boldsymbol{s_i^*} + \beta$, where $\alpha \neq 0$ and $\beta$ are scalars.

[54]Remember that in the general set up here, $\vec{\boldsymbol{\theta}} = (\boldsymbol{\theta_1}, \cdots, \boldsymbol{\theta_M})$ is an $M$-dimensional fundamental. Taking the first order condition of the decision-specific optimality condition in (3) and collecting terms, the optimal decision rule for each self $i$ is then given by $x_i^*(\boldsymbol{\omega_i}) = E_i\left[\sum_{1\leq m\leq M}\psi_{i,m}\boldsymbol{\theta_m} + \sum_{j\neq i}\gamma_{i,j}x_j^*(\boldsymbol{\omega_j})\right]$, where $\psi_{i,m} = -\frac{\partial^2 u}{\partial x_i \partial \theta_m}\left(\frac{\partial^2 u}{\partial x_i^2}\right)^{-1}$ and $\gamma_{i,j} = -\frac{\partial^2 u}{\partial x_i \partial x_j}\left(\frac{\partial^2 u}{\partial x_i^2}\right)^{-1}$. $\boldsymbol{\vartheta_i} \equiv \sum_{1\leq m\leq M}\psi_{i,m}\boldsymbol{\theta}_m$ then summarizes how the fundamental directly influences decision $i$.

**Proposition 13** *The optimal signals depend on the fundamental $\vec{\boldsymbol{\theta}}$ as follows:*

$$\begin{pmatrix} E\left[\boldsymbol{s_1^*}|\vec{\boldsymbol{\theta}}\right] \\ \\ \cdots \\ \\ E\left[\boldsymbol{s_N^*}|\vec{\boldsymbol{\theta}}\right] \end{pmatrix} = \begin{pmatrix} \boldsymbol{t_1} \\ \\ \cdots \\ \\ \boldsymbol{t_N} \end{pmatrix} = \left( \mathbb{I}_N - \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1 & \lambda_1 \\ \lambda_2 & 1 & \cdots & \lambda_2 & \lambda_2 \\ & & \cdots & & \\ \lambda_{N-1} & \lambda_{N-1} & \cdots & 1 & \lambda_{N-1} \\ \lambda_N & \lambda_N & \cdots & \lambda_N & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} \boldsymbol{\vartheta_1} \\ \cdots \\ \boldsymbol{\vartheta_k} \\ \cdots \\ \boldsymbol{\vartheta_N} \end{pmatrix}, \tag{74}$$

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

Similar to Proposition 3, 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 $\boldsymbol{\vartheta_i}$, summarizing the fundamental's direct influence. This in turn lets the *optimal signal* for self $i$ depend more on her own $\boldsymbol{\vartheta_i}$. To further illustrate the last point, consider a symmetric optimum for the costly contemplation problem with two decisions. From (74), we have

$$\begin{pmatrix} \boldsymbol{t_1} \\ \boldsymbol{t_2} \end{pmatrix} = \left( \mathbb{I}_2 - \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} \boldsymbol{\vartheta_1} \\ \boldsymbol{\vartheta_2} \end{pmatrix} = \begin{pmatrix} \frac{\boldsymbol{\vartheta_1}}{1-\gamma^2\lambda^2} + \lambda\gamma\frac{\boldsymbol{\vartheta_2}}{1-\gamma^2\lambda^2} \\ \frac{\boldsymbol{\vartheta_2}}{1-\gamma^2\lambda^2} + \lambda\gamma\frac{\boldsymbol{\vartheta_1}}{1-\gamma^2\lambda^2} \end{pmatrix},$$

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 $\boldsymbol{\vartheta_{-i}}$ compared to her own $\boldsymbol{\vartheta_i}$ 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 $\boldsymbol{\vartheta_i}$. 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 $\boldsymbol{\vartheta_i}$. The decision maker becomes a "completely" narrow thinker: each decision $i$ is only based on her own $\boldsymbol{\vartheta_i}$, i.e. the fundamental's direct influence. This also echoes the limit result in Proposition 11 discussed above.[55]

**Allocation of cognitive capacities across different decisions.** We finally turn to the optimal allocation of cognitive capacities, $\tau_i^*$, across different decisions.

**Proposition 14** *In the optimum of the costly contemplation problem in (28),*

$$\tau_i^* > \tau_j^* \iff \left|\frac{\partial^2 u}{\partial x_i^2}\right| Var\left(\boldsymbol{t_i}\right) > \left|\frac{\partial^2 u}{\partial x_j^2}\right| Var\left(\boldsymbol{t_j}\right).$$

Proposition 14 shows that more volatile decisions (with high $Var\left(\boldsymbol{t_i}\right)$) and decisions with respect to

[55]Proposition 11 discussed above means that, when the cognitive constraint is severe, it is optimal for each self $i$ to only receive signal about $\boldsymbol{\theta_i}$. In the context there, $\boldsymbol{\theta_i}$ captures the fundamental's influence on self $i$'s best response function and corresponds to $\boldsymbol{\vartheta_i}$ defined in the general multiple-decision problem here.

which the marginal utility is more sensitive (with high $\left|\frac{\partial^2 u}{\partial x_i^2}\right|$) 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 $\left|\frac{\partial^2 u}{\partial x_i^2}\right|$) than to the self who is in charge of purchasing apples (low $\left|\frac{\partial^2 u}{\partial x_i^2}\right|$). Similarly, the decision maker may allocate more cognitive capacities to the self who invests bitcoins (volatile $\boldsymbol{t_i}$) than to the self who invests ETFs (stable $\boldsymbol{t_i}$).[56]

**Proof of Lemma 5.** A necessary condition for $\{\boldsymbol{\omega_i^*}, x_i^*(\cdot)\}_{i=1}^N$ to be an optimum of the costly contemplation problem in (28) is that, for each $i$, $(\boldsymbol{\omega_i^*}, x_i^*(\cdot))$ is optimally chosen, taking other $\left\{\boldsymbol{\omega_j^*}, x_j^*(\cdot)\right\}_{j\neq i}$ as given. That is, $(\boldsymbol{\omega_i^*}, x_i^*(\cdot))$ solves

$$\max_{\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}, x_i(\cdot)} E\left[u\left(x_1(\omega_1), \cdots, x_N^*(\omega_N^*), \vec{\theta}\right)\right] \tag{75}$$

$$s.t. \quad I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right) \leq \tau - \sum_{j\neq i} I\left(\boldsymbol{\omega_j^*}; \vec{\boldsymbol{\theta}}\right), \tag{76}$$

As $u$ is quadratic, maximizes the objective in (75) is equivalent to maximizing

$$E\left[\frac{u_{i,i}}{2}\left(x_i(\omega_i) - \vartheta_i - \sum_{j\neq i}\gamma_{i,j}E\left[x_j^*(\omega_j^*)|\vec{\theta}\right]\right)^2 + \frac{u_{i,i}}{2}\left(\sum_{j\neq i}\gamma_{i,j}\left(x_j^*(\omega_j^*) - E\left[x_j^*(\omega_j^*)|\vec{\theta}\right]\right)\right)^2 + f\left(\left\{x_j^*(\omega_j^*)\right\}_{j\neq i}, \vec{\theta}\right)\right] \tag{77}$$

where $u_{i,i} = \frac{\partial^2 u}{\partial x_i^2}$ and I use the fact that $\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \cdots, \boldsymbol{\omega_N}$ is conditionally independent given $\vec{\boldsymbol{\theta}}$. 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 $\boldsymbol{\omega_i}$ takes the form in Lemma 5.

**Proof of Proposition 13.** Given the chosen information $\{\boldsymbol{\omega_i^*}\}_{i=1}^N$, the optimal decision rule $\{x_i^*(\cdot)\}_{i=1}^N$ can be characterized by (7). We then have $x_i^*(\boldsymbol{\omega_i^*}) = E\left[\boldsymbol{\vartheta_i} + \sum_{j\neq i}\gamma_{i,j}x_j^*\left(\boldsymbol{\omega_j^*}\right)|\boldsymbol{s_i^*}\right] = \lambda_i \boldsymbol{s_i^*}$, where $\lambda_i = \frac{\sigma_{t_i}^2}{\sigma_i^2+\sigma_{t_i}^2}$. Together with (73), we have

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

This leads to (74).

[56] One may wonder whether it is the case that $\tau_i^* > \tau_j^* \iff \left|\frac{\partial^2 u}{\partial x_i^2}\right| Var(\boldsymbol{v_i}) > \left|\frac{\partial^2 u}{\partial x_j^2}\right| Var(\boldsymbol{v_j})$, where $\boldsymbol{\vartheta_i}$ is defined above, summarizing how the fundamental directly influences optimal decision $i$. This is not necessarily the case. Even if $\boldsymbol{\vartheta_i}$ is volatile, if the $\boldsymbol{\vartheta_j}$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 $i$.

**Proof of Proposition 14.** For $\left\{0 \leq \Lambda_i \leq 1, \varsigma_i^2\right\}_{i=1}^N$, define $g\left(\left\{\Lambda_i, \varsigma_i^2\right\}_{i=1}^N\right) \equiv E\left[u\left(x_1, \cdots, x_N, \vec{\theta}\right)\right]$, where for all $i$, $\boldsymbol{x_i} = \Lambda_i\left(\boldsymbol{t_i} + \boldsymbol{\epsilon_i}\right)$, $\boldsymbol{\epsilon_i} \sim N\left(0, \varsigma_i^2\right)$, and $\left\{\boldsymbol{t_i}\right\}_{i=1}^N$ are given by

$$\begin{pmatrix} \boldsymbol{t_1} \\ \\ \cdots \\ \\ \boldsymbol{t_N} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \Lambda_1 & \cdots & \Lambda_1 & \Lambda_1 \\ \Lambda_2 & 1 & \cdots & \Lambda_2 & \Lambda_2 \\ & & \cdots & & \\ \Lambda_{N-1} & \Lambda_{N-1} & \cdots & 1 & \Lambda_{N-1} \\ \Lambda_N & \Lambda_N & \cdots & \Lambda_N & 1 \end{pmatrix} \circ \Gamma\right)^{-1} \begin{pmatrix} \boldsymbol{\vartheta_1} \\ \cdots \\ \boldsymbol{\vartheta_k} \\ \cdots \\ \boldsymbol{\vartheta_N} \end{pmatrix}. \tag{78}$$

Based on Lemma 5 and Proposition 13, in the optimum of the costly contemplation problem in (28), $\left\{\lambda_i, \sigma_i^2\right\}_{i=1}^N$ defined in Lemma 5 and Proposition 13 must solve

$$\max_{\left\{\Lambda_i, \varsigma_i^2\right\}_{i=1}^N} g\left(\left\{\Lambda_i, \varsigma_i^2\right\}_{i=1}^N\right) \tag{79}$$
$$h\left(\left\{\varsigma_i^2\right\}_{i=1}^N\right) = \frac{1}{2}\sum_i \log_2 \frac{\varsigma_i^2 + \sigma_{t_i}^2}{\varsigma_i^2} \leq \tau,$$

where $\sigma_{t_i}^2$ is the variance of $\boldsymbol{t}_i$ defined based on (78). Rewrite $g$ in a way similar to (77), we have

$$\frac{\partial g\left(\left\{\Lambda_i, \varsigma_i^2\right\}_{i=1}^N\right)}{\partial\left(\varsigma_i^2\right)} = \frac{u_{i,i}}{2}\Lambda_i^2 \quad \text{and} \quad \frac{\partial h\left(\left\{\varsigma_i^2\right\}_{i=1}^N\right)}{\partial\left(\varsigma_i^2\right)} = -\frac{1}{2\log 2}\frac{\sigma_{t_i}^2}{\varsigma_i^2\left(\sigma_{t_i}^2 + \varsigma_i^2\right)}$$

where $u_{i,i} = \frac{\partial^2 u}{\partial x_i^2}$. As $\left\{\lambda_i = \frac{\sigma_{t_i}^2}{\sigma_i^2 + \sigma_{t_i}^2}, \sigma_i^2\right\}_{i=1}^N$ must solve (42), in the optimum of the costly contemplation problem in (28), we must have

$$u_{i,i}\Lambda_i^2 / \left(\frac{\sigma_{t_i}^2}{\sigma_i^2\left(\sigma_{t_i}^2 + \sigma_i^2\right)}\right) = u_{j,j}\Lambda_j^2 / \left(\frac{\sigma_{t_j}^2}{\sigma_j^2\left(\sigma_{t_j}^2 + \sigma_j^2\right)}\right) \quad \forall i, j,$$

and

$$u_{i,i}\sigma_{t_i}^2\left(1 - \lambda_i\right) = u_{j,j}\sigma_{t_j}^2\left(1 - \lambda_j\right).$$

As $\lambda_i = 1 - 2^{-2\tau_i^*}$, Proposition 14 follows.

# References

**Abaluck, Jason, and Jonathan Gruber.** (2011) "Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program." *American Economic Review*, 101(4): 1180–1210.

**Abaluck, Jason, and Jonathan Gruber.** (2016) "Evolving choice inconsistencies in choice of prescription drug insurance." *American Economic Review*, 106(8): 2145–84.

**Abeler, Johannes, and Felix Marklein.** (2016) "Fungibility, labels, and consumption." *Journal of the European Economic Association*, 15(1): 99–127.

**Allcott, Hunt, and Dmitry Taubinsky.** (2015) "Evaluating behaviorally motivated policy: experimental evidence from the lightbulb market." *American Economic Review*, 105(8): 2501–38.

**Anderson, John.** (2009) *Cognitive Psychology and its Implications.* Macmillan.

**Angeletos, George-Marios, and Alessandro Pavan.** (2007) "Efficient Use of Information and Social Value of Information." *Econometrica*, 75(4): 1103–1142.

**Angeletos, George-Marios, and Alessandro Pavan.** (2009) "Policy with dispersed information." *Journal of the European Economic Association*, 7(1): 11–60.

**Angeletos, George-Marios, and Chen Lian.** (2016) "Dampening General Equilibrium: From Micro to Macro." *NBER Working Paper No. 23379.*

**Angeletos, George-Marios, and Chen Lian.** (2018) "Forward Guidance without Common Knowledge." *American Economic Review (forthcoming).*

**Angeletos, George-Marios, and Jennifer La'O.** (2018) "Optimal Monetary Policy with Informational Frictions." *MIT mimeo.*

**Angrist, Joshua D, Sydnee Caldwell, and Jonathan V Hall.** (2017) "Uber vs. Taxi: A Driver's Eye View." *NBER Working Paper No. 23891.*

**Beatty, Timothy KM, Laura Blow, Thomas F Crossley, and Cormac O'Dea.** (2014) "Cash by Any Other Name? Evidence on Labeling from the UK Winter Fuel Payment." *Journal of Public Economics*, 118: 86–96.

**Benabou, Roland, and Jean Tirole.** (2002) "Self-confidence and personal motivation." *Quarterly Journal of Economics*, 117(3): 871–915.

**Benabou, Roland, and Jean Tirole.** (2003) "Intrinsic and extrinsic motivation." *Review of Economic Studies*, 70(3): 489–520.

**Benabou, Roland, and Jean Tirole.** (2004) "Willpower and personal rules." *Journal of Political Economy*, 112(4): 848–886.

**Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen.** (2015) "Turning a Shove Into a Nudge? A "Labeled Cash Transfer" for Education." *American Economic Journal: Economic Policy*, 7(3): 86–125.

**Bergemann, Dirk, and Stephen Morris.** (2013) "Robust Predictions in Games with Incomplete Information." *Econometrica*, 81(4): 1251–1308.

**Bergemann, Dirk, Tibor Heumann, and Stephen Morris.** (2017) "Information and interaction." *mimeo.*

**Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer.** (2017) "Memory, Attention, and Choice." *NBER Working Paper No. 23256.*

**Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler.** (1997) "Labor supply of New York City cabdrivers: One day at a time." *Quarterly Journal of Economics*, 112(2): 407–441.

**Colombo, Luca, Gianluca Femminis, and Alessandro Pavan.** (2014) "Information acquisition and welfare." *Review of Economic Studies*, 81(4): 1438–1483.

**Crawford, Vincent P, and Juanjuan Meng.** (2011) "New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income." *American Economic Review*, 101(5): 1912–32.

**Denti, Tommaso.** (2017*a*) "Network Effects in Information Acquisition." *Cornell University mimeo.*

**Denti, Tommaso.** (2017*b*) "Unrestricted information acquisition." *Cornell University mimeo.*

**Dessein, Wouter, Andrea Galeotti, and Tano Santos.** (2016) "Rational Inattention and Organizational Focus." *American Economic Review*, 106(6): 1522–36.

**Enke, Benjamin.** (2018) "What You See Is All There Is." *Harvard University mimeo.*

**Farber, Henry S.** (2015) "Why you can't find a taxi in the rain and other labor supply lessons from cab drivers." *Quarterly Journal of Economics*, 130(4): 1975–2026.

**Farhi, Emmanuel, and Xavier Gabaix.** (2015) "Optimal taxation with behavioral agents." *NBER Working Paper.*

**Fehr, Ernst, and Lorenz Goette.** (2007) "Do workers work more if wages are high? Evidence from a randomized field experiment." *American Economic Review*, 97(1): 298–317.

**Fudenberg, Drew, and David K Levine.** (2006) "A Dual-self Model of Impulse Control." *American Economic Review*, 96(5): 1449–1476.

**Gabaix, Xavier.** (2014) "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics*, 129(4).

**Gabaix, Xavier.** (2017) "Behavioral inattention." *NBER Working Paper, No. 24096.*

**Gabaix, Xavier, and David Laibson.** (2006) "Shrouded attributes, consumer myopia, and information suppression in competitive markets." *Quarterly Journal of Economics*, 121(2): 505–540.

**Gennaioli, Nicola, and Andrei Shleifer.** (2010) "What Comes to Mind." *Quarterly Journal of Economics*, 125(4): 1399–1433.

**Golub, Benjamin, and Stephen Morris.** (2017*a*) "Expectations, Networks, and Conventions." *mimeo.*

**Golub, Benjamin, and Stephen Morris.** (2017*b*) "Higher-Order Expectations." *mimeo.*

**Gottlieb, Daniel.** (2014) "Imperfect Memory and Choice under Risk." *Games and Economic Behavior*, 85: 127–158.

**Gottlieb, Daniel.** (2017) "Will You Never Learn? Self Deception and Biases in Information Processing." *mimeo.*

**Gul, Faruk, and Wolfgang Pesendorfer.** (2001) "Temptation and Self-control." *Econometrica*, 69(6): 1403–1435.

**Hall, Robert, and Frederic Mishkin.** (1982) "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." *Econometrica*, 461–481.

**Hastings, Justine, and Jesse Shapiro.** (2017) "How are SNAP Benefits Spent? Evidence from a Retail Panel." *NBER Working Paper No. 23112*.

**Hastings, Justine S, and Jesse M Shapiro.** (2013) "Fungibility and consumer choice: Evidence from commodity price shocks." *Quarterly Journal of Economics*, 128(4): 1449–1498.

**Heath, Chip, and Jack Soll.** (1996) "Mental Budgeting and Consumer Decisions." *Journal of Consumer Research*, 23(1): 40–52.

**Hébert, Benjamin, and Michael Woodford.** (2017) "Rational Inattention and Sequential Information Sampling." *NBER Working Paper No. 23379*.

**Hellwig, Christian, and Laura Veldkamp.** (2009) "Knowing What Others Know: Coordination Motives in Information Acquisition." *The Review of Economic Studies*, 76(1): 223–251.

**Hirshman, Samuel, Devin Pope, and Jihong Song.** (2018) "Mental Budgeting vs. Relative Thinking." *University of Chicago mimeo.*

**Jappelli, Tullio, and Luigi Pistaferri.** (2014) "Fiscal policy and MPC heterogeneity." *American Economic Journal: Macroeconomics*, 6(4): 107–36.

**Jehiel, Philippe, and Jakub Steiner.** (2018) "Selective Sampling with Information-Storage Constraints." *University of Edinburgh mimeo.*

**Kahana, Michael Jacob.** (2012) *Foundations of Human Memory.* OUP USA.

**Kahneman, Daniel.** (2011) *Thinking, Fast and Slow.* Macmillan.

**Koszegi, Botond, and Filip Matejka.** (2018) "An Attention-Based Theory of Mental Accounting." *CERGE-EI mimeo.*

**Kueng, Lorenz.** (2018) "Excess Sensitivity of High-Income Consumers." *The Quarterly Journal of Economics.*

**Laibson, David.** (1997) "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2): 443–478.

**Lorenzoni, Guido.** (2010) "Optimal Monetary Policy with Uncertain Fundamentals and Dispersed Information." *The Review of Economic Studies*, 77(1): 305–338.

**Mackowiak, Bartosz, and Mirko Wiederholt.** (2009) "Optimal Sticky Prices under Rational Inattention." *American Economic Review*, 99(3): 769–803.

**Mackowiak, Bartosz, Filip Matejka, and Mirko Wiederholt.** (2018) "Rational Inattention: A Disciplined Behavioral Model." *Goethe University Frankfurt mimeo.*

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

**Mas-Colell, Andreu, Michael Dennis Whinston, and Jerry Green.** (1995) *Microeconomic theory.* Vol. 1, Oxford university press New York.

**Matejka, Filip.** (2015) "Rigid pricing and rationally inattentive consumer." *Journal of Economic Theory*, 158: 656–678.

**Matejka, Filip.** (2016) "Rationally inattentive seller: Sales and discrete pricing." *Review of Economic Studies*, 83(3): 1125–1155.

**Matejka, Filip, and Alisdair McKay.** (2015) "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105(1): 272–98.

**Morris, Stephen, and Hyun Song Shin.** (2002) "Social Value of Public Information." *American Economic Review*, 92(5): 1521–1534.

**Morris, Stephen, and Ming Yang.** (2016) "Coordination and Continuous Choice." *mimeo.*

**Myatt, David P, and Chris Wallace.** (2012) "Endogenous information acquisition in coordination games." *Review of Economic Studies*, 79(1): 340–374.

**Myatt, David P, and Chris Wallace.** (2017) "Information acquisition and use by networked players." *mimeo.*

**O'Donoghue, Ted, and Matthew Rabin.** (1999) "Doing it Now or Later." *American Economic Review*, 89(1): 103–124.

**Parker, Jonathan A.** (2017) "Why Don't Households Smooth Consumption? Evidence from a 25 Million Dollar Experiment." *American Economic Journal: Macroeconomics*, 9(4): 153–83.

**Parker, Jonathan A, Nicholas S Souleles, David S Johnson, and Robert McClelland.** (2013) "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review*, 103(6): 2530–53.

**Piccione, Michele, and Ariel Rubinstein.** (1997) "On the interpretation of decision problems with imperfect recall." *Games and Economic Behavior*, 20(1): 3–24.

**Rabin, Matthew, and Georg Weizsacker.** (2009) "Narrow Bracketing and Dominated Choices." *American Economic Review*, 99(4): 1508–43.

**Read, Daniel, George Loewenstein, and Matthew Rabin.** (1999) "Choice bracketing." In *Elicitation of preferences*. 171–202. Springer.

**Sims, Christopher A.** (2003) "Implications of Rational Inattention." *Journal of Monetary Economics*, 50(3): 665–690.

**Sims, Christopher A.** (2010) "Rational Inattention and Monetary Economics." *Handbook of Monetary Economics*, 3: 155–181.

**Steiner, Jakub, Colin Stewart, and Filip Matejka.** (2017) "Rational Inattention Dynamics: Inertia and Delay in Decision-Making." *Econometrica*, 85(2): 521–553.

**Stephens Jr, Melvin, and Takashi Unayama.** (2011) "The Consumption Response to Seasonal Income: Evidence from Japanese Public Pension Benefits." *American Economic Journal: Applied Economics*, 86–118.

**Stevens, Luminita.** (2015) "Coarse pricing policies." *University of Maryland mimeo.*

**Thakral, Neil, and Linh T To.** (2017) "Daily Labor Supply and Adaptive Reference Points." *mimeo.*

**Thaler, Richard.** (1985) "Mental accounting and consumer choice." *Marketing science*, 4(3): 199–214.

**Thaler, Richard H.** (1999) "Mental accounting matters." *Journal of Behavioral decision making*, 12(3): 183.

**Tversky, Amos, and Daniel Kahneman.** (1973) "Availability: A heuristic for judging frequency and probability." *Cognitive psychology*, 5(2): 207–232.

**Tversky, Amos, and Daniel Kahneman.** (1981) "The framing of Decisions and the Psychology of Choice." *Science*, 211(4481): 453–458.

**Veldkamp, Laura.** (2011) *Information Choice in Macroeconomics and Finance.* Princeton University Press.

**Wilson, Andrea.** (2014) "Bounded Memory and Biases in Information Processing." *Econometrica*, 82(6): 2257–2294.

**Yang, Ming.** (2015) "Coordination with flexible information acquisition." *Journal of Economic Theory*, 158: 721–738.