A Theory of Narrow Thinking*

Chen Lian†

December 10, 2018

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. The narrow thinker violates the fungibility principle, and can exhibit mental accounting-type behavior. Narrow thinking also reconciles other seemingly disparate phenomena in a unified framework, such as excess sensitivity to temporary income shocks, the small wage elasticity of daily labor supply, temptation and comfort zones. 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, behavioral economics, 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, Robert Gibbons, Itzhak Gilboa, Jonathon Hazell, Bengt Holmstrom, Jennifer La'O, Yueran Ma, Filip Matejka, Stephen Morris, Harry Pei, Collin Raymond, Karthik Sastry, Frank Schilbach, Ludwig Straub, Juuso Toikka, Laura Veldkamp, Olivier Wang, Ivan Werning, and seminar participants at Barcelona GSE, D-TEA and MIT for insightful comments. This paper supersedes the early working paper under the title "A Theory of Local Thinking."

†MIT; lianchen@mit.edu.

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 all these decisions. Consider a standard textbook consumer problem of demanding multiple goods. The classical demand function is derived imposing that, when the consumer purchases a particular 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 of her consumption decisions 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 researches in psychology and behavioral economics point out (Tversky and Kahneman, 1981; Rabin and Weizsacker, 2009), the decision maker often "narrowly brackets," and makes each decision in isolation.

In this paper I develop an approach, which I term *narrow thinking*, to break the decision-maker's ability to perfectly coordinate her multiple decisions. I then show how this approach 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 is motivated by the psychological observation that the decision maker may not incorporate all the relevant information when making each decision (Kahneman, 2011). 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 on top of her mind (which can arise from bounded recall or selective retrieval from memory when making a particular decision). When she purchases gasoline, she knows the gasoline price, but does not have the food price on 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 ω_i, which captures the *state of mind* when the decision maker decides on x_i. The decision-maker can then be thought 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.

Narrow thinking as a formalization of within-person coordination frictions. The decision problem under narrow thinking is then shown to be mathematically equivalent to multiple selves playing an *incomplete-information*, common interest, game. In the unique equilibrium of the game, each self's decision is made with an imperfect understanding of the others' decisions. In this sense, the narrow thinking approach formalizes within-person coordination frictions.

I then use a simple example to illustrate how such within-person coordination frictions influence the narrow thinker's behavior. In this example, the decision maker's utility depends on both how closely each of her decision x_i can track its "local fundamental" θ_i, and how her different decisions interact. 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 decision maker's decisions 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.

It is worth clarifying the relationship between narrow thinking and rational inattention (Sims, 2003; Matejka and McKay, 2015; Koszegi and Matejka, 2018). The narrow thinking approach builds upon the rational inattention literature 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 to coordinate 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] 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

[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 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.

fundamental.[2]

The classical consumer problem and the case without income effects. I first study a classical consumer problem of demanding multiple goods under narrow thinking. In this context, it is natural to consider the following narrow thinker: when each self i of the narrow thinker decides on the consumption of good i, she perfectly knows its price p_i, but only receives noisy signals about other prices.

To tease out the main mechanism, I start from the case with quasi-linear utility and 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 response to shocks to prices, narrow thinking effectively attenuates such interaction. Such a friction and the accompanying dampening of indirect effects can then translate into the attenuation of both cross-price and own-price demand elasticities. The narrow thinker's Slutsky matrix can also be asymmetric.

The case with income effects and the violation of the fungibility principle. I then turn to the case with income effects. The interaction across different decisions now comes from the budget constraint. In this context, different selves 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 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. Under standard consumer theory, the decision maker can decrease other consumption to smooth out the drop in the 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. It is worth noting that 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, can be frictionless.

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). It also generates new

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

testable predictions. For example, excess sensitivity to price changes under narrow thinking only happens for goods with respect to which the decision maker's demand is not very elastic.

Under-reaction and over-reaction with narrow thinking. The previous analysis focuses on how the narrow thinker's demand responds to price changes. Narrow thinking can generate interesting economic implications beyond this scope. In a general multiple-decision problem, the key lessons from the previous analysis remain valid: in response to shocks to the fundamental, narrow thinking leads to an effective attenuation of interaction across decisions and a dampening of indirect effects.

A point worth emphasizing is that, depending on the environment, narrow thinking can translate into either over- or under-reaction. A rule of thumb is the following: 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 of the direct effect, a dampening of the indirect effect under narrow thinking often 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.

Narrow thinking can then explain various empirical examples of under- and over-reaction. For example, excess smoothness to taste shocks (Heath and Soll, 1996), the small wage elasticity of daily labor supply (Camerer et al., 1997), excess sensitivity to income shocks (Thaler, 1999) and the label effect (Abeler and Marklein, 2016). Narrow thinking also offers a novel theory of comfort zones and temptation.

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 the context of the simple illustrative example, it is optimal for each self to receive a more precise signal about her local fundamental than other selves. In this sense, narrow thinking arises endogenously.

Layout. The remainder of the paper is organized as follows. Section 2 sets up a general multiple-decision problem, defines the notion of narrow thinking, and provides a game theoretic representation. Section 3 uses a simple example to show how narrow thinking leads to an effective attenuation of interaction across decisions and a dampening of indirect effects. Section 4 studies consumer theory under narrow thinking and shows how the narrow thinker violates the fungibility principle and generates mental-accounting type behavior. Section 5 turns to other applications. Section 6 studies the costly contemplation problem. Section 7 concludes. The Appendix contains proofs and additional results.

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)[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.[4] 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

[3]Other notable contributions in the rational inattention and endogenous information acquisition literature include, but are not limited to: 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 and Caplin and Dean (2015), Caplin, Dean and Leahy (2017) and Oliveira et al. (2017) for axiomatic foundations. In other contributions, Ofek, Yildiz and Haruvy (2007) and Gabaix and Laibson (2017) focus on the decision maker's imperfect perception of her own utility, Ilut and Valchev (2017) focus on the decision maker's difficulty of deriving optimal decision rules, and Khaw, Li and Woodford (2018) focus on the decision maker's noisy mental representations of the decision situation.

[4]Technically, compared to these works, the studied game among multiple selves has a particular feature: common interest. This feature facilitates a sharper characterization of the narrow thinker's behavior, especially in the costly contemplation problem.

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) mostly focuses on motivated beliefs and reasoning, and explores reasons why the decision maker's beliefs and behavior can be systematically biased. The focus of the current paper is about frictional behavior in response to shocks. Narrow thinking does not necessarily leads to systematical bias on average. The team theory literature (Marschak and Radner, 1972; Dessein and Santos, 2006; 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. Angeletos and Pavan (2009) 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 and lead to narrow thinking – for example, Gennaioli and Shleifer (2010), Kahana (2012), Bordalo, Gennaioli and Shleifer (2017) and Jehiel and Steiner (2018) on bounded recall, and Tversky and Kahneman (1973), Anderson (2009), Kahneman (2011) on heuristics, biases, and selective retrieval from memory. Gennaioli and Shleifer (2010) use the term "local thinking" to capture the representativeness heuristic and study its implications 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 mental accounting (Thaler, 1985, 1999; Heath and Soll, 1996; Gilboa and Gilboa-Schechtman, 2003; Hastings and Shapiro, 2013; Abeler and Marklein, 2016), excessive sensitivity to temporary income shock (Johnson, Parker and Souleles, 2006; Parker et al., 2013; Kueng, 2018), the small cross-price demand elasticity (Gabaix and Laibson, 2006; Abaluck and Gruber, 2011, 2016; Allcott and Wozny, 2014; Allcott and Taubinsky, 2015), the small wage elasticity of daily labor supply (Camerer et al., 1997; Crawford and Meng, 2011; Farber, 2015; Thakral and To, 2017) and temptation (Laibson, 1997; O'Donoghue and Rabin, 1999; Gul and Pesendorfer, 2001; Fudenberg and Levine, 2006).

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 under narrow thinking 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 function 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. First, 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.[5]

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),

[5]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.

under which decision i is made, where Ω_i denotes the set of possible signal realizations for decision. As further discussed in Section 2.3, one should interpret ω_i as the *state of mind* when the decision maker decides on x_i. Here, each ω_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 ω_i.[6]

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

It is worth noticing that, as u is strictly concave over $\vec{x}$, the optimum of (2), if it exists, is unique.[7,8]

Lemma 1 *If the optimum of (2) exists, 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

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

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

[8] 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.

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.[9]

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 on 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 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.[10] To further understand the mathematical nature of the decision problem under narrow thinking in (2)

[9]Specifically, I use $\mathcal{I}_i$ to denote the partition of the state space S generated by decision i's signal ω_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.

[10]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.

and provide an alternative interpretation, it is useful to provide an equivalent, game-theoretic, representation of (2).[11] 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 ω_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 underlying probability (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 ω_i.

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

Narrow thinking 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

[11] 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 an important behavioral feature: the decision maker often neglects other decisions' impact when making a particular decision. As further discussed in Section 2.3, the only departure from the standard individual decision problem is that ω_i, which captures the state of mind when the decision maker decides on x_i, may not summarize all the relevant information.

coordinating her multiple 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 the decision maker "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

Comparison with standard sequential decisions. Though the main focus of the paper is multiple-decision problems that are traditionally treated as static (e.g. demand of multiple goods), it is also worth clarifying the difference between narrow thinking and the standard practice for the sequential-decision problem. The environment set up in Section 2.1 can also be interpreted as a sequential-decision problem in which decision i is made before decision j for all $i < j$. In this case, the standard practice imposes the information of the earlier decision i is perfectly nested in 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.

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 psychological study of memory is that people have bounded recall (Kahana, 2012). For example, the recency effect in psychology documents that a person often only has perfect recall of the last a 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 heuristics and biases in decision making (Tversky and Kahneman, 1973; Gennaioli and Shleifer (2010); Kahneman, 2011), which is closely connected to the notion of selective retrieval from memory 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. Kahneman (2011) and Enke (2018) use the

"What You See Is All There Is" principle to summarize this type of cognitive friction. For example, when the decision maker purchases food (gasoline), she only sees the food (gasoline) price, and does not have the gasoline (food) price on 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, ω_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.

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 conflicted 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.[12] 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 example studied in Section 4, I let $\mathcal{X}_{N+1} = \mathbb{R}$, hence one can always find an $x_{N+1} \in \mathcal{X}_{N+1}$ (possibly negative) such that (4) is satisfied. In this case, the key friction of interest is the difficulty of coordinating the first N decisions. In fact, when the

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

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 An Illustrative Example

I first use a simple example to illustrate a few key abstract insights about the optimal behavior under narrow thinking. As each self of the narrow thinker has an imperfect perception of the others' decisions, each self's beliefs about the others' decisions are anchored in response to shocks to fundamentals. I then show how such belief anchoring 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. These insights will be useful later in more concrete applications.

3.1 Environment

Utility. 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:

$$u\left(\vec{x}, \vec{\theta}\right) = -\frac{1}{2}\sum_{i=1}^{N}(x_i - \theta_i)^2 + \sum_{1\leq i<j\leq N}\frac{\gamma_{i,j}+\gamma_{j,i}}{2}x_i x_j, \tag{5}$$

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 $\vec{x}$. For notation simplicity, I also set $\gamma_{i,i} = 0$ for all i.

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," $\theta_i \sim \mathcal{N}\left(\bar{\theta}_i, \sigma^2_{\theta_i}\right)$. 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. Finally, different θ_is are independent from each other.

Optimal decision rules. The optimal decision rule for each decision i can be characterized by the decision specific optimality condition in (3). Taking the first order condition of (3), we have

$$E\left[\frac{\partial u}{\partial x_i}\left(x^*_i(\omega_i), \vec{x}^*_{-i}, \vec{\theta}\right)|\omega_i\right] = 0 \quad \forall i, \omega_i \in \Omega_i. \tag{6}$$

Together with the utility in (5), the optimal decision rule for each i is then given by:

$$x_i^*(\omega_i) = E_i\left[BR_i\left(\vec{x}_{-i}^*;\vec{\theta}\right)\right] \equiv \underbrace{E_i[\theta_i]}_{\text{direct effect}} + \underbrace{E_i\left[\sum_{j\neq i}\gamma_{i,j}x_j^*(\omega_j)\right]}_{\text{indirect effect}} \quad \forall i, \omega_i \in \Omega_i, \tag{7}$$

where $E_i[\cdot] = E[\cdot|\omega_i]$ and the signals $\{\omega_i\}_{i=1}^N$ will be specified shortly.

For each i, the local fundamental θ_i summarizes the fundamental $\vec{\theta}$'s direct influence on the optimal decision i, holding other decisions fixed. In the equivalent game among multiple selves, θ_i captures the fundamental's influence on self i's best response function, i.e. the "intercept" of $BR_i\left(\vec{x}_{-i}^*;\vec{\theta}\right)$.

On the other hand, $\gamma_{i,j}$ summarizes how decision i is influenced by decision j. A positive (negative) $\gamma_{i,j}$ means that self i's decision increases (decreases) with self j's decision. 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 (7) 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.

In fact, the optimal decision rules here are essentially the same as those in the quasi-linear consumer theory context in Section 4.2 (baring an approximation there). There, each decision i is the consumption of the good i, and its "local fundamental" is (a negative multiple of) the price of the good i, p_i. 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.

Information. To isolate the friction of interest, I study the following narrow thinker: each self $i \in \{1, \cdots N\}$, who is in charge of decision i, perfectly knows its local fundamental θ_i, but only receives a noisy signal about each of the other θ_j: $s_{i,j} = \theta_j + \epsilon_{i,j}$, $\epsilon_{i,j} \sim \mathcal{N}\left(0, \sigma_{i,j}^2\right)$. Noises $\epsilon_{i,j}$ are independent from fundamentals and each other. To summarize, $\omega_i = \left\{\{s_{i,j}\}_{j\in\{1,\cdots,N\}}\right\}$, where for notation simplicity, $s_{i,i} = \theta_i$ and $\sigma_{i,i}^2 = 0$, i.e. each self i has a perfect signal about θ_i. As each self i perfectly knows its local fundamental θ_i, the direct effect in (7) is the same as in the frictionless case, and the friction comes solely from each self's imperfect perception of other decisions.[13]

The above information structure captures the idea that each self has more precise information

[13]This case then offers a sharp contrast from the one considered in rational inattention and sparsity. There, the key friction is the decision maker's uncertainty about the fundamental, that is, the frictional direct effects. On the other hand, the decision maker perfectly knows her other decisions when making a particular decision.

about the fundamental that directly influences her decision, the form of narrow thinking that is of interest throughout the paper. A similar information structure arises endogenously in the costly contemplation problem studied in Section 6, when the decision maker chooses optimally the information upon which each decision is made: the decision maker has incentives to let each self i receive a more precise signal about her own local fundamental θ_i than other selves. In the context of quasi-linear consumer theory (Section 4.2), as mentioned above, the local fundamental θ_i corresponds to (a negative multiple of) the price of the good i, p_i. The information structure then means that each self i perfectly knows the price of the good she buys, p_i, and receives a noisy signal of each of the other prices p_j.

3.2 Belief Anchoring

Question of interest. I now turn to the question of interest: how the narrow thinker's decisions respond to shocks to each of the fundamental θ_k.

Belief anchoring. From (7), we know these responses depend on both the direct effects (frictionless) and the indirect effects (frictional). To understand these indirect effects, we need to understand how each self i's beliefs about how other decisions, i.e. $E_i\left[x_j^*\right]$, respond to shocks to θ_k. Under narrow thinking, as each self has an imperfect perception of other selves' decisions, this type of beliefs will be anchored.

Specifically, I use $E\left[\cdot|\theta_k\right]$ to denote the conditional expectation with respect to θ_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 respect to shocks to θ_k and $E\left[x_j^*|\theta_k\right]$ captures how much decision j itself moves with respect to shocks to θ_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 θ_k, and is anchored towards the prior $E\left[x_j^*\right]$ (the unconditional mean of decision j).

Proposition 2 *In response to shocks to θ_k, each self i's belief about decision $j \neq i$, $E_i\left[x_j^*\right]$, is anchored:*

$$E\left[E_i\left[x_j^*\right]|\theta_k\right] = \lambda_{i,k} E\left[x_j^*|\theta_k\right] + \left(1-\lambda_{i,k}\right) E\left[x_j^*\right] \qquad \forall k,\ \forall i \neq j, \tag{8}$$

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

The degree of such anchoring is parametrized by $\lambda_{i,k} = \frac{\sigma_{\theta k}^2}{\sigma_{\theta_k}^2 + \sigma_{i,k}^2} \in (0,1]$, which is a function of the signal-to-noise ratio of self i's signal about θ_k, $s_{i,k}$. For self i, the signal $s_{i,k}$ can have two roles. First, it helps self i predict θ_k. Second, it helps self i predict how other decisions respond to θ_k. From (7), the first role does not matter per se for the optimal x_i for any $i \neq k$. On the other hand, the second role is crucial in determining each optimal x_i and the noise in such prediction leads to the belief anchoring in condition (8).[14]

3.3 Effective Attenuation of Interaction

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, $x_i^{\text{Narrow}}\left(\vec{\theta}\right)$, as

$$x_i^{\text{Narrow}}\left(\vec{\theta}\right) \equiv E\left[x_i^*\left(\omega_i\right)|\vec{\theta}\right] \quad \forall i, \tag{9}$$

where $E\left[\cdot|\vec{\theta}\right]$ denotes the conditional expectation with respect to $\vec{\theta}$.[15] 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

$$\left\{x_i^{\text{Standard}}\left(\vec{\theta}\right)\right\}_{i=1}^N = \arg\max_{\{x_i\}_{i=1}^N} u\left(x_1, \cdots, x_N, \vec{\theta}\right), \tag{10}$$

the standard, frictionless, decision function in which each decision is made with perfect knowledge of all the fundamentals. From the perspective of an econometrician who has data on fundamentals and decisions (but not the signal of each narrow thinker's self), $x_i^{\text{Narrow}}\left(\vec{\theta}\right)$ defined in (9) 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 θ_k, i.e. $\left\{\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k}\right\}_{i=1}^N$.[16] For each decision i, the belief anchoring in (8) 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.

---

[14] As each self i perfectly knows her own θ_i, we have $\lambda_{i,i} = \frac{\sigma_{i,i}^2}{\sigma_{\theta_i}^2 + \sigma_{i,i}^2} = 1$. That is, self i can perfectly predict how other x_js respond to her own local fundamental θ_i, i.e., $E\left[E_i\left[x_j^*\right]|\theta_i\right] = E\left[x_j^*|\theta_i\right]$. This is a corner case as self i perfectly knows her own θ_i. Even so, x_i's response to θ_i will still deviate from the standard benchmark, as other x_js respond differently to θ_i. In the consumer theory context studied in the next section, this point will be explained further when I study the own-price demand elasticities in Propositions 7 and 8.

[15] Without taking conditional expectation with respect to $\vec{\theta}$, narrow thinking then provides a model of stochastic choice. This is a direction worth further exploring. Different from the logit model commonly studied in the stochastic choice literature, the narrow thinker's problem studied here is a continuous, multiple-decision, problem.

[16] For each i, the $x_i^{\text{Narrow}}\left(\vec{\theta}\right)$ defined in (9) is linear in its arguments. This is because fundamentals and signals are Normally distributed and the optimal decision rule in (7) is linear. This permits me to characterize decision functions' partial derivatives in Proposition 3.

Proposition 3 *In response to shocks to θ_k, the effective interaction across decisions is attenuated:*

$$\begin{pmatrix} \frac{\partial x_1^{\text{Narrow}}}{\partial \theta_k} \\ \frac{\partial x_2^{\text{Narrow}}}{\partial \theta_k} \\ \cdots \\ \\ \frac{\partial x_N^{\text{Narrow}}}{\partial \theta_k} \end{pmatrix} = \left(\mathbb{I}_N - \tilde{\Gamma}_k\right)^{-1} \begin{pmatrix} 0 \\ \cdots \\ 1 \\ \cdots \\ 0 \end{pmatrix}, \tag{11}$$

where $\tilde{\Gamma}_k$ captures the narrow thinker's effective interaction matrix in response to θ_k with

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

and $\circ$ is the element by element product.

Proposition 3 is the main result of this Section. From the above condition, one can see that, in response to shocks to θ_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 θ_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.

One interesting feature about Proposition 3 is that the degree of such effective attenuation of interaction is shock specific (the matrix $\tilde{\Gamma}_k$ depends on k). As a result, the analysis here allows the following possibility: if a fundamental θ_k is volatile, each self may pay "more attention" to θ_k and all selves' signal about θ_k will be more precise. Different selves then better coordinate in response to θ_k, that is, there is less attenuation of interaction in response to θ_k driven by narrow thinking. 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 25 in Appendix E.

3.4 Dampening of Indirect Effects

Dampening of indirect effects. As defined in condition (7), each decision i's response to shocks to fundamentals can be decomposed into a direct and an indirect effect. As each self i knows her local fundamental θ_i, the direct effect is maintained under narrow thinking. On the other hand, as each self has an imperfect understanding about 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.

Formally, in the current environment, the absolute value of all eigenvalues of the interaction matrix Γ is less than one,[17] and the game among multiple selves is solvable by iterating the best response:

$$\begin{aligned} x_i^*(\omega_i) &= \theta_i + \sum_{j\neq i} \gamma_{i,j} E_i\left[x_j^*\right] = \theta_i + \sum_{j\neq i} \gamma_{i,j} E_i\left[\theta_j\right] + \sum_{j\neq i} \gamma_{i,j}\left(\sum_{l\neq j} \gamma_{j,l} E_i\left[E_j\left[x_l^*\right]\right]\right) \\ &= \underbrace{\theta_i}_{\text{Direct}} + \underbrace{\sum_{j\neq i} \gamma_{i,j} E_i\left[\theta_j\right] + \sum_{j\neq i} \gamma_{i,j}\left(\sum_{l\neq j} \gamma_{j,l} E_i\left[E_j\left[\theta_l\right]\right]\right) + \cdots}_{\text{Indirect}}. \end{aligned} \tag{12}$$

The above condition means that, as the indirect effect on x_i comes from self i's belief about other decisions, it in turn depends on self i's belief about other selves' local fundamentals, self i's belief about other selves' beliefs about other selves' local fundamentals, 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.

I then compare the size of each component of the narrow thinker's indirect effect with its frictionless counterpart (indexed by the superscript *Standard*, as above). To facilitate the comparison, similar to (9) above about the narrow thinker's decision function, I average the realization of noises in signals, and define $x_i^{\text{Ind},+,\text{Narrow}}\left(\vec{\theta}\right) \equiv E\left[x_i^{\text{Ind},+}(\omega_i)\,|\vec{\theta}\right]$ and $x_i^{\text{Ind},-,\text{Narrow}}\left(\vec{\theta}\right) \equiv E\left[x_i^{\text{Ind},-}(\omega_i)\,|\vec{\theta}\right]$.

[17] This is because, for all i, $\sum_{j\neq i} |\gamma_{i,j}| < 1$.

Proposition 4 *For each decision x_i and in response to shocks to each θ_k, each component of the indirect effect is dampened under narrow thinking:*

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

The result comes from the general property that beliefs of higher-order are anchored more in response to shocks (e.g. Samet, 1998 and Angeletos and Lian, 2018). In environments in which one component of the indirect effect dominates the other,[18] Proposition 4 then means the *net* total of the indirect effect is dampened. In fact, this will be the case for most applications studied in the rest of paper. Depending on whether the net total of the indirect effect works in the same direction as the direct effect, such a dampening of the indirect effect can then translate into either under- or over-reaction with narrow thinking. This is a main theme throughout the rest of the paper.

3.5 Additional Discussion

Correlated fundamentals and rational confusion. In the analysis above, I let different θ_is be uncorrelated. When different θ_is are correlated, an additional channel, "rational confusion," emerges: as self i perfectly knows her own local fundamental θ_i but only receives a noisy signal about other θ_j, to the extent that θ_i and θ_j are correlated, self i can use θ_i to forecast θ_j. 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 rational confusion considerations. In fact, when different θ_is 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 θ_j is based on her signal about θ_j, $s_{i,j}$, solely. In this case, $E_i[\theta_j] = E[\theta_j | s_{i,j}]$. In later analysis, I also establish results about the behavior under narrow thinking allowing such rational confusion considerations, e.g. Proposition 22 in Appendix D.[19]

Frictional response to shocks and unbiasedness on average. The above frictional

[18]For example, in the context here, if all $\gamma_{i,j} > 0$, all indirect effects are in the same direction. For all i, all indirect effects are captured by $x_i^{\text{Ind},+}(\omega_i)$, and $x_i^{\text{Ind},-}(\omega_i)$ is zero.

[19]Alternatively, one can derive robust predictions that are independent from the correlation structure of the fundamental and the information structure, along the line of Bergemann and Morris (2013) and Bergemann, Heumann and Morris (2017). In the narrow thinking context, predictions that can be delivered along this line seem to be quite limited. Still, in Appendix F, I am able to establish a version of the dampening of indirect effects result in Proposition 4 under general information structure. A difference compared to Bergemann and Morris (2013) and Bergemann, Heumann and Morris (2017) is that I focus on conditional response to shocks, while they mainly focus on volatility.

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 5 *On average, each of the narrow thinker's decisions coincide with the frictionless one:*

$$E\left[x_i^{\text{Narrow}}\right] = E\left[x_i^{\text{Standard}}\right] \quad \forall i, \tag{13}$$

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

Comparison with rational inattention. It is worth clarifying the difference between narrow thinking and rational inattention in the context of the illustrative example here. 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 ω is her imperfect, but common, signal. Narrow thinking, on the other hand, breaks such certainty equivalence: $x_i^*(\omega_i) \neq x_i^{\text{Standard}}\left(E\left[\vec{\theta}|\omega_i\right]\right)$.

4 Consumer Theory under Narrow Thinking

In this section, I study a classical consumer problem of demanding multiple goods under narrow thinking. To tease out the mechanism, I first solve two polar cases. In the first, the decision maker's utility is quasi-linear. The interaction across decisions comes from the complementarity/substitutability embedded in the utility function, and there are no income effects. In the second, the decision maker's utility is separable but there are income effects. The interaction across decisions then comes from the budget constraint. In this case, 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 the money allocated to another. The narrow thinking approach then generates a model of mental accounting-type behavior. It provides an alternative to the explicit mental budgeting model in Heath and Soll (1996) and offers new testable predictions. I finally study the general case with both non-separable utility and income effects.

4.1 Set up

Set up. 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 consumption in the future 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 shocks to prices here).[20] Along the line of 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 the budget constraint to substitute y:

$$u(x_1, \cdots, x_N, \vec{p}) = \tilde{u}\left(x_1, \cdots, x_N, w - \sum_{i=1}^{N} p_i x_i\right). \tag{14}$$

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

Information. In this context, it is natural to consider the following case of narrow thinking: when self i decides on the consumption x_i, she perfectly knows its price p_i, but only receives noisy signals about other prices. Similar to the information structure considered in Section 3, it captures the idea that each self i has more precise knowledge about her "local fundamental." It is also consistent with the "What You See is All There Is" principle raised in Kahneman (2011).

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 $p_i \sim \log\mathcal{N}\left(\log \bar{p}_i, \sigma_{p_i}^2\right)$, but receives a noisy signal about each of the other p_j: $s_{i,j} = p_j \epsilon_{i,j}$, with $\epsilon_{i,j} \sim \log\mathcal{N}\left(0, \sigma_{i,j}^2\right)$ and $\sigma_{i,j}^2 > 0$. To summarize, for $i \in \{1, \cdots, N\}$, self i's signal is given by $\omega_i = \left\{\{s_{i,j}\}_{j\in\{1,\cdots,N\}}\right\}$, where for notation simplicity, $s_{i,i} = p_i$ and $\sigma_{i,i}^2 = 0$, i.e. each self i has a perfect signal about p_i. Finally, different ps and ϵs are independently distributed and there is no "rational confusion."

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

[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.

to analytically characterize the narrow thinker's behavior. 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). It also allows me to state the prediction about the narrow thinker's behavior in terms of elasticities, better connecting to the empirical literature.[21]

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: $\{\bar{x}_i\}_{i=1}^N = \arg\max_{\{x_i\}_{i=1}^N} u(x_1, \cdots, x_N, \bar{p}_1, \cdots \bar{p}_N)$. I then use a hat over a variable to denote its log-deviation from this point, e.g. $\hat{x}_i = \log \frac{x_i}{\bar{x}_i}$.

4.2 The Case Without Income Effects

Utility. In this subsection, I study a commonly used utility function, namely CES with quasi-linear utility. Specifically, the decision maker's utility in (14) is given by

$$u(x_1, \cdots, x_N, \vec{p}) = \frac{\left(\sum_{i=1}^N \frac{1}{N} x_i^{\frac{\varrho-1}{\varrho}}\right)^{\frac{\varrho(1-\kappa)}{\varrho-1}}}{1-\kappa} + w - \sum_{i=1}^N p_i x_i, \tag{15}$$

where $\rho > 0$ captures the elasticity of substitution between each pair of goods (in standard consumer theory), 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 (15) is strictly concave over $\vec{x}$. As discussed above, I study the following narrow thinker: when her self i decides on the consumption x_i, she perfectly knows its price p_i, but only receives noisy signals about other prices. The distributions of prices and signals are as specified above.[22]

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 (15) and taking first order condition of (3), we have, for each i,

$$E_i\left[\frac{\partial u}{\partial x_i}\left(x_i^*(\omega_i), \vec{x}_{-i}^*\right)\right] = E_i\left[\frac{1}{N} x_i^*(\omega_i)^{-\frac{1}{\varrho}} X^{\frac{1}{\varrho}-\kappa}\right] = p_i. \tag{16}$$

I then log-linearize the above condition and use a hat over a variable to denote its log-deviation

[21] One can also work with linearized optimal decision rules. Then, the prediction about the narrow thinker's behavior would be stated in terms of gradients instead of elasticities.

[22] 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 (17). This scenario arises when $\bar{p}_i = \bar{p}_j$ for each $i \neq j$.

from the point of log-linearization, e.g. $\hat{x}_i = \log \frac{x_i}{\bar{x}_i}$. Collecting terms, we have

$$\hat{x}_i^* (\omega_i) = E_i \left[BR_i \left(\vec{x}_{-i}^*; \vec{p} \right) \right] \equiv \underbrace{-\psi \hat{p}_i}_{\text{direct effect}} + \underbrace{\sum_{j \neq i} \gamma E_i \left[\hat{x}_j^* (\omega_j) \right]}_{\text{indirect effect}}, \tag{17}$$

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 (17) captures the direct effect of price changes on x_i, that is, the effect of $\vec{p}$ holding other decisions fixed. ψ 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 (17) captures the indirect effect on x_i, that is, other consumption decisions' impact on x_i. A positive (negative) γ means that each pair of goods are complements (substitutes),[23,24] and that the optimal consumption of good i increases (decreases) with self i's belief about each of the other consumption x_j.

In fact, the optimal decision rules in (17) are exactly the same as those in the illustrative quadratic example, (7). The lessons from Section 3 then remain valid: in response to shocks to the fundamental, each self's beliefs about the others' decisions are anchored; and narrow thinking leads to an effective attenuation of interaction across decisions and a dampening of indirect effects.

Narrow thinker's demand. Now I translate those abstract insights (effective attenuation of interaction and dampening of indirect effects) into predictions about the narrow thinker's behavior. Specifically, the main question of interest is how the narrow thinker's consumption responds to price changes. Similar to (9), for each i, I define the narrow thinker's (log) demand as a function of (log) prices:

$$\hat{x}_i^{\text{Narrow}} (\hat{p}_1, \cdots, \hat{p}_N) \equiv E \left[\hat{x}_i^* (\omega_i) \,|\hat{p}_1, \cdots, \hat{p}_N \right] \quad \forall i, \tag{18}$$

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.

For $i \neq j$, $\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i}$ and $\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_j}$ then capture consumption i's own- and the cross-price elasticities for the narrow thinker, while $\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i}$ and $\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_j}$ capture consumption i's own- and the cross-price elasticities in standard consumer theory.

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

[24] From the expression of γ, we know that if the elasticity of substitution is large, i.e. $\varrho > \frac{1}{\kappa}$, $\gamma < 0$.

Cross-price demand elasticities. From condition (17), we know that cross-price demand elasticities are driven solely by changes of each self's beliefs about the others' decisions, i.e. indirect effects. Not surprisingly, from Proposition 2 (belief anchoring) and Proposition 4 (dampening of indirect effects), cross-price demand elasticities are typically attenuated under narrow thinking.

Proposition 6 *All of the narrow thinker's cross-price demand elasticities are attenuated:*

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

when at least one of the following three conditions hold: 1) Complements, i.e. $\gamma > 0$*. 2) Two goods, i.e.* $N = 2$*. 3) Symmetric information, i.e.* $\lambda_{i,j} \equiv \frac{\sigma_{p_j}^2}{\sigma_{p_j}^2 + \sigma_{i,j}^2} = \lambda$ *for all* $i \neq j$*.*

In the case of substitutes without symmetric information, an intriguing possibility arises: it is not always the case that all of the cross-price elasticities are attenuated. This possibility contrasts with the prediction under rational inattention and arises because of the coexistence of opposing indirect effects discussed in Proposition 4. Such possibility is discussed further in the proof of Proposition 6.[25]

Own-price demand elasticities. We now turn to own-price demand elasticities. Interestingly, even though each self perfectly knows the price of the good she buys, own-price elasticities are attenuated under narrow thinking. That is, the narrow thinker's consumption of each good is excessively smooth in response to changes to the price of that good.

Proposition 7 *For each* i*, the narrow thinker's consumption* x_i *decreases (increases) less in response to positive (negative) shocks to* p_i *:*

$$\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} \leq \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} < 0 \quad \forall i. \tag{19}$$

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

[25] As an example, fix $i \neq j$. Let both self i and self j perfectly know p_i, but let other selves do not have knowledge about the shock to p_i. That is, $\lambda_{j,i} = 1$ and $\lambda_{l,i} = 0$ for all $l \neq i, j$. Consider a positive shock to p_i. An increase in p_i has a negative direct effect on x_i. As different goods are substitutes, the consumption of other goods should increase. However, for all $l \neq i, j$, x_l does not respond to the shock. As a result, self j, who does know the shock, increases x_j more. That is, her cross-price elasticity is larger under narrow thinking. Such possibility arises because of the coexistence of opposing indirect effects. The increase in p_i has two opposing indirect effects on x_j. First, an increase in p_i will decrease x_i, which will in turn increase x_j. Second, an increase in p_i will decrease x_i, which will in turn increase other x_l and then in turn decrease x_j. Under the information structure considered, the first type of indirect effect is maintained while the second type of indirect effect is attenuated. Together, it leads to an increase in the cross-price elasticity is under narrow thinking.

To understand (19), let us first consider the complements case with $\gamma > 0$. From (17), 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 an decrease will further decrease x_i, generating a negative indirect effect on x_i. Under narrow thinking, as discussed above, other consumption x_j decreases less in response to an increase in p_i. The indirect effect on x_i is then 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, other consumption x_j increases less in response to an increase in p_i.[26] The indirect effect on x_i is dampened, and x_i decreases less in response to an increase in p_i.

In sum, as the indirect effect of p_i on x_i comes from a second degree interaction, the indirect effect 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 Proposition 7.

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 the first self's signal about p_2 differs from the signal-to-noise ratio of the second self's signal about p_1, the Slutsky matrix becomes asymmetric.[27]

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 5 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 B provides further discussion along this line.

4.3 Income Effects and the Violation of the Fungibility Principle

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,

[26] As discussed above, there is a possibility that some x_j may increase more in response to an increase in p_i under narrow thinking. However, in this context, the total indirect effect on x_i, summing over the influence of all other consumption, is always dampened under narrow thinking.

[27] 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, j, $\frac{\partial x_i^{\text{Narrow}}}{\partial p_j} = \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} \frac{\bar{x}_i}{\bar{p}_j}$.

different selves 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 the money allocated to another. Such a violation then generates mental accounting-type behavior which has "no agreed-upon model" (Farhi and Gabaix, 2015),[28] e.g. excess sensitivity to own-price changes (Hastings and Shapiro, 2013).[29] 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 Section 4.1, I let the consumer's utility be given by

$$\tilde{u}\left(x_1, \cdots, x_N, y\right) = \sum_{i=1}^{N} v_i\left(x_i\right) + h\left(y\right). \tag{20}$$

In (20), $v_i\left(x_i\right) = \frac{x_i^{1-\kappa_i}}{1-\kappa_i}$ 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 κ_i means the demand for good i is less elastic.[30] $h\left(y\right)$, a strictly concave function on $\mathbb{R}$,[31] captures the consumer's utility from the last decision, which can be interpreted as utility from the consumption in the future 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 (treated as a constant now). 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. Same as the previous section, I consider the following narrow thinker: each self $i \in \{1, \cdots, N\}$ of the narrow thinker decides on the consumption x_i; she perfectly

[28] 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 Proposition 8 will elaborate further.

[29] 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.

[30] Holding the spending share for good i fixed, a higher κ_i means a less elastic Marshallian demand for good i. In fact, in the frictionless consumer theory, $-1/\kappa_i$ captures the "Frisch" elasticity of demand for good i, that is, the elasticity holding the marginal utility of money fixed.

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

knows its price p_i, but only receives noisy signals about other prices. The distributions of prices and signals are as specified above.

Optimal consumption decisions. The optimal decision for each self $i \in \{1, \cdots, N\}$, $x_i^* (\omega_i)$, must satisfy

$$\underbrace{v_i^{'} (x_i^* (\omega_i))}_{\text{marginal utility of consuming good i}} = p_i E_i \left[\underbrace{h' (y^*)}_{\text{marginal value of monery}} \right]. \tag{21}$$

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. This condition holds because the last decision is made based on perfect knowledge, and the standard perturbation argument holds between the consumption x_i and y.

I then log-linearize the above condition and the budget constraint, and use a hat over a variable to denote its log-deviation from the point of log-linearization, e.g. $\hat{x}_i = \log \frac{x_i}{\bar{x}_i}$. The optimal decision rule (21) for each i then becomes

$$-\kappa_i \hat{x}_i^* (\omega_i) = \hat{p}_i - \kappa_h E_i [\hat{y}^*] , \tag{22}$$

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

$$\sum_{i=1}^{N} \mu_i (\hat{x}_i^* (\omega_i) + \hat{p}_i) + \mu_y \hat{y}^* = 0, \tag{23}$$

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 $\hat{p}_j$ directly influences the budget constraint (23) 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 influence the optimal consumption for each $\hat{x}_i$. This channel is muted in the quasi-linear case above, as the marginal value of money is a constant there.

Violation of the fungibility principle. From (22), one can also 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, $-\kappa_h E_i [\hat{y}^*]$. From (22), 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.

Lemma 3 *In standard consumer theory (as above, indexed by superscript Standard), the marginal value of spending an additional unit of money on each good is the same:*

$$-\kappa_i \hat{x}_i^{\text{Standard}}(\hat{p}_1, \cdots, \hat{p}_N) - \hat{p}_i = -\kappa_j \hat{x}_j^{\text{Standard}}(\hat{p}_1, \cdots, \hat{p}_N) - \hat{p}_j \quad \forall i \neq j. \tag{24}$$

Under narrow thinking, it is possible to find a pair (i, j) of decisions such that the marginal value of spending an additional unit of money on them differs:

$$-\kappa_i \hat{x}_i^*(\omega_i) - \hat{p}_i \neq -\kappa_j \hat{x}_j^*(\omega_j) - \hat{p}_j. \tag{25}$$

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. Similar to condition (18), 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.

Proposition 8 *For each good i such that $\kappa_i > 1$, the narrow thinker's consumption x_i decreases (increases) more in response to positive (negative) shocks to p_i:*[32]

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

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). Under narrow thinking, such smoothing out from the decrease of other consumption is limited. This dampens the indirect effect and generates a larger drop of x_i in response to an increase in p_i.

One may wonder why we see over-reaction (excess sensitivity) in the own-price demand elasticity, here, in Proposition 8 but under-reaction (excess smoothness) in the own-price demand elasticity in Proposition 7. Here, when $\kappa_i > 1$, the indirect effect of an increase in p_i on x_i, through the decrease of other consumption, positively influences x_i. The indirect effect works in the opposite direction of the direct effect. A dampening of the indirect effect under narrow thinking then leads to over-reaction. This contrasts with the case in Proposition 7, where the indirect

[32] Similar to Proposition 6, in the current environment, one can establish that all of the narrow thinker's cross-price demand elasticities are attenuated under narrow thinking when either $N = 2$ or $\lambda_{i,j} = \lambda$ for all $i \neq j$.

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.

Additional economic implications. First, the excess sensitivity in Proposition 8 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.[33] 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. $\left\{\lambda_{i,j} \equiv \frac{\sigma_{p_j}^2}{\sigma_{p_j}^2+\sigma_{i,j}^2}\right\}_{i,j=1}^{N}$.

Second, when $\kappa_i < 1$, instead, the narrow thinker's consumption x_i drops less in response to an increase in p_i. That is, for all i, $\frac{\partial \hat{x}_i^{\text{Standard}}}{\partial \hat{p}_i} < \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i}$.[34] 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 works in the same direction as the direct effect. 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 possible that the consumer's demand with respect to gasoline is less elastic (has a higher κ_i). As a result, the empirical finding is line with the prediction in Proposition 8.

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 5 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 in Read, Loewenstein and Rabin (1999): 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.

[33] See the proof of Proposition 8 for a proof of this statement.

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

4.4 Additional Results and Discussion

Consumer theory with general non-quasi linear utility. In Section 4.2, I studied a case in which the utility is quasi-linear and the interaction across decisions comes solely from the complementarity/substitutability embedded in the utility function. In Section 4.3, I studied a case in which the utility function is separable across i and the interaction across different decisions comes solely from the budget constraint. These cases help me tease out the key mechanism behind the narrow thinker's frictional behavior. Now, I study the general case introduced in Section 4.1, with both non-separable utility and income effects. In this general case, the lessons from the illustrative example in Section 3 remain valid: in response to shocks to the fundamental, narrow thinking leads to an effective attenuation of interaction across decisions and a dampening of indirect effects. I refer the reader to Appendix C for detail.

In terms of how each of the narrow thinker's consumption x_i responds to its own price p_i, we know that there are competing channels from Propositions 7 and 8: the attenuation of interaction that comes from the complementarity/substitutability directly embedded in the utility function leads to excess smoothness, while the attenuation of interaction that comes from the budget constraint can lead to excess sensitivity. Whether one channel dominates the other then depends on the environment.[35] On the other hand, one can establish that each of the narrow thinker's consumption x_i responds more sluggishly to shocks to other prices p_j, i.e. the attenuation of cross-price demand elasticities, in a general non-quasilinear case with symmetry. This is Proposition 14 in Appendix C.

Cognitive inertia and the boundary of a self. In the above analysis, each self is in charge of purchasing one good. She perfectly knows its price, and receives noisy signals about other prices. More generally, one should define the boundary of a self as a group of decisions made based on the same information. Such definition also connects naturally 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

[35] In a general non-quasilinear case with symmetry, Proposition 15 in Appendix C gives conditions about when each channel dominates.

made with perfect knowledge of other decisions.[36] In sum, as Kahneman (2011) points out, "we tend to make decisions as problems arise."

5 Under- and Over-reaction with Narrow Thinking

In this section, I move beyond the response to price changes, and study the narrow thinker's behavior in other contexts. The key lessons from the previous sections remain valid: in response to shocks to the fundamental, narrow thinking leads to an effective attenuation of interaction across decisions and a dampening of indirect effects. In the general multiple-decision problem introduced in Section 2, I establish those results in Propositions 16 and 17 in Appendix D.

A point worth emphasizing is that, depending on the environment, narrow thinking can translate into either over- or under-reaction. A rule of thumb is the following: 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 of the direct effect, a dampening of the indirect effect under narrow thinking often leads to over-reaction.[37] The analysis also contrasts with the often-held belief that noises in the decision maker's mental representation of the world typically lead to under-reaction. In the rest of this Section, I then explain how narrow thinking provides a unified framework to explain various empirical examples of over- and under-reaction.

5.1 Under-reaction with Narrow Thinking

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,

[36] As an example of the second case, consider a decision maker who plans to buy fuji and gala apples at the supermarket. The decision maker can observe the fuji apple's price and the gala apple's price at the same time, and make the consumption decisions of both goods together. The analysis in this Section in fact can easily nest this case. If decisions $(x_{i_1}, \cdots, x_{i_K})$ are made together and based on the same information, I can simply let $\lambda_{i_a,i_b} = 1$ and $\lambda_{i_a,j} = \lambda_{i_b,j}$ for all $a, b \in \{1, \cdots, K\}$ and $j \in \{1, \cdots, N\}$.

[37] This is called a rule of thumb instead of a proposition, because, as discussed in Proposition 4, there can be opposing indirect effects that work in different directions. A dampening of each part of the indirect effects may not mean the dampening of the net total. Reassuringly, this rule of thumb holds in all application in this Section. Moreover, in a symmetric consumer theory context with income effects, I am able to formalize this rule of thumb in Proposition 15 in Appendix C.

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 the food consumption, in response to the positive taste shock. Under narrow thinking, however, such a coordinated response is limited, and food consumption will increase less.

To formalize, consider an environment similar to the one in Section 4.3 (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 4.3. Here I introduce taste shocks, and $\varphi_i \sim \log \mathcal{N}\left(\log \bar{\varphi}_i, \sigma_{\varphi_i}^2\right)$ 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 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. As I am interested in response to taste shocks, I treat w and all ps 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 φ_i, but receives a noisy signal about each of the other φ_j. Specifically, for $i \in \{1, \cdots, N\}$, self i's signal is given by $\omega_i = \left\{\{s_{i,j}\}_{j\in\{1,\cdots,N\}}\right\}$, where $s_{i,i} = \varphi_i$ and, for $i \neq j$, $s_{i,j} = \varphi_j \epsilon_{i,j}$, with $\epsilon_{i,j} \sim \log \mathcal{N}\left(0, \sigma_{i,j}^2\right)$ and $\sigma_{i,j}^2 > 0$. All ϵs and φs are independent from each other.

I use a hat over a variable to denote its log-deviation from the point of log-linearization.[38] Similar to conditions (9) and (18), for all i, I define the narrow thinker's (log) demand function as $\hat{x}_i^{\text{Narrow}}\left(\hat{\varphi}_1, \cdots, \hat{\varphi}_N\right) \equiv E\left[\hat{x}_i^*\left(\omega_i\right) | \hat{\varphi}_1, \cdots, \hat{\varphi}_N\right]$, 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 can then establish excess smoothness to taste shocks under narrow thinking.

Proposition 9 *For each good i, the narrow thinker's consumption x_i increases (decreases) less in response to positive (negative) taste shocks to φ_i :*

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

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

To see the mechanism behind the excess smoothness, note an increase in taste φ_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). This scenario falls into the case that the indirect effect works in the same direction as the direct effect, and the dampening of the indirect effect under narrow thinking leads to under-reaction. Specifically, under narrow thinking, other x_j's coordinated decrease is limited, and the consumer increases x_i less. As a result, the narrow thinker's consumption exhibits excess smoothness to taste shocks.[39]

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.

Specifically, the decision maker's utility is given by:

$$\sum_{i=1}^{N} \varphi_i v_i \left(x_i\right) - c\left(\sum_{i=1}^{N} x_i\right), \tag{28}$$

where x_i is the time the decision maker assigns to activity i, $\varphi_i v_i\left(x_i\right) = \varphi_i \frac{x_i^{1-\kappa_i}}{1-\kappa_i}$ with $\kappa_i > 0$ is her utility from activity i, φ_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. Similar to the case above, I consider the following narrow thinker: when self i of the narrow thinker decides on x_i, she perfectly knows the attractiveness φ_i, but only receives noisy signals about other $\varphi_j s$.

In fact, the time management problem here is formally similar to the previous model of excess smoothness to taste shocks. To see this, note that, after using the budget constraint, the decision maker's utility in the previous model can be written as $\sum_{i=1}^{N} \varphi_i v_i\left(x_i\right) + h\left(w - \sum_{i=1}^{N} p_i x_i\right)$, where h is concave. Her utility in the previous model is then convex in $\sum_{i=1}^{N} x_i$, similar to the utility in (28) here. A result similar to Proposition 9 can then also be established here (Proposition 18 in Appendix D). In this context, it means, under narrow thinking, the amount of time allocated for each activity i is more sluggish in response to shocks to its attractiveness. The narrow thinker will stay within her "comfort zones," even if activity i becomes more attractive.

For a concrete example of the above comfort zones behavior, consider an engineering student who decides on how much time she will spend on the economics class. She realizes the economics class has a great professor, that is, the φ_i of spending time on the economics class is high. However,

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

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 zones and engage in the economics class, even with an excellent professor.

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 and decreasing her labor supply on other days. Such a coordinated response then generates a large elasticity of daily labor supply. Under narrow thinking, however, labor supply on other days will not be as responsive, and such frictions will prevent a large increase in labor supply on the day of wage increase. Proposition 19 in Appendix D formalizes how narrow thinking generates the 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 income targeting and reference dependence model often predicts negative wage elasticity of daily labor supply, inconsistent with the empirical evidence. Second, in line with Proposition 5, 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 frequency, 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 τ in Section 6, facilitating the decision maker to coordinate her multiple selves.

5.2 Over-reaction with Narrow Thinking

An information-based theory of 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 4.3 (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 4.3. 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 $w_i \sim \mathcal{N}\left(\bar{w}_i, \sigma^2_{w_i}\right)$, but receives a noisy signal about each of the other 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 $\epsilon_{i,j} \sim \mathcal{N}\left(0, \sigma^2_{i,j}\right)$ and $\sigma^2_{i,j} > 0$. All ϵs and ws are independent from each other. The information structure captures the idea that decision maker has the winter Fuel Payment on top of her mind when purchasing fuel, but not necessarily when making other purchases.

Similar to conditions (9) and (18), for all i, I define the narrow thinker's demand function:[40]

$$x_i^{\text{Narrow}}(w_1, \cdots, w_N) \equiv E\left[x_i^*(\omega_i) | w_1, \cdots, w_N\right] \quad \forall i, \tag{29}$$

where $E[\cdot | w_1, \cdots, w_N]$ 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 the label effect under narrow thinking.

[40]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.

Proposition 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 Proposition 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 Proposition 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 of the direct effect, and narrow thinking leads to over-reaction.

Excess sensitivity to temporary income shocks. Similar mechanism can also explain the 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: the decision maker should increase her consumption by a small amount at all points in time. Under local thinking, however, consumption at other points in time may not be as responsive to current temporary income shocks. As a result, the local thinker increases her current consumption more. This is Proposition 20 in Appendix D.

Temptation. Now I offer another example of over-reaction: the narrow thinker is particularly tempted by new attractions. To illustrate, consider that each self i of the narrow thinker is in charge of how long she will play computer games on day i. Now a new computer game is introduced, and the attractiveness of playing on all days increases.[41] When the narrow thinker decides how long she will play on each day i, her belief about playing time on other days is anchored. The indirect effect from longer playing time on other days is then dampened. As a result, she will play longer on day i. In this sense, the narrow thinker over-reacts: she is "tempted" by the new computer game.[42] This is Proposition 22 in Appendix D.

The narrow thinking theory of temptation also connects to the neglect of the "adding-up effects" in Read, Loewenstein and Rabin (1999). In this setting, the cost of playing more computer games on a single day is low. However, the cumulative costs can be large (e.g. opportunities costs and eye damage), and can increase faster than the cumulative benefits. The narrow thinker, who underestimates how long she will play on other days after the introduction of the new game, 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 the prediction based on self-control and habit formation (Laibson, 1997; O'Donoghue and Rabin, 1999; Gul and Pesendorfer, 2001; Fudenberg and Levine, 2006). Moreover, this 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.

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.[43] 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, it is optimal for different decisions' signals to take different forms. The analysis also provides a framework to study the optimal information choice problem

[41] In the model of comfort zones, I study the impact of shocks to the attractiveness of *one* activity. Here, the shock of interest is a common shock influencing the attractiveness of *all* activities.

[42] Here the common increase in the attractiveness will have a positive direct effect on the playing time on each day. As the cost of playing computer games (e.g. opportunities costs and eye damage) is convex in the total playing time, the increase in playing time on other days will then have a negative indirect effect on the playing time on each day. As a result, the direct effect and the indirect effect work in opposite directions, and narrow thinking leads to over-reaction.

[43] 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.

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 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 $\vec{\boldsymbol{\theta}} : S \to \Theta$, where $\Theta \subseteq \mathbb{R}^M$ denotes the set of possible fundamental.

I then use ω_i to denote the signal (potentially multi-dimensional) under which each decision i is made. ω_i is the realization of a random vector $\boldsymbol{\omega_i} : S \to \Omega_i$, where Ω_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:

$$\max_{\{\boldsymbol{\omega_i} \in \boldsymbol{\Omega_i}, x_i(\cdot)\}_{i=1}^N} E\left[u\left(x_1(\omega_1), \cdots, x_N(\omega_N), \vec{\theta}\right)\right] \tag{30}$$

$$s.t. \quad \sum_{i=1}^N I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right) \le \tau. \tag{31}$$

In the cognitive constraint (31), $I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right)$ denotes the mutual information between decision i's signal $\boldsymbol{\omega_i}$ and the fundamental $\vec{\boldsymbol{\theta}}$, which equals to the entropy reduction $H\left(\vec{\boldsymbol{\theta}}\right) - H\left(\vec{\boldsymbol{\theta}}|\boldsymbol{\omega_i}\right)$. It captures the cognitive cost for decision i. (31) then means the sum of cognitive costs used by all decisions i cannot surpass the decision maker's total cognitive capacity, τ. Finally, I let $\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \cdots, \boldsymbol{\omega_N}$ be conditionally independent given $\vec{\boldsymbol{\theta}}$. That is, the noise in each decision i's signal about $\vec{\boldsymbol{\theta}}$ is idiosyncratic.[44]

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

[44] 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 these idiosyncratic noises. Nevertheless, this section focuses on how different decisions' signals take different *forms.*

same as the one studied in Section 2, and the optimal decision rule can be characterized by (7). 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 (31). One can henceforth interpret the costly contemplation problem in (30) 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 (31). 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. Moreover, there is also an additional layer to the costly contemplation problem here: the optimal allocation of which decision uses more cognitive capacities, i.e. the allocation of $\left\{\tau_i \equiv I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right)\right\}_{i=1}^{N}$. It is also worth noting that, to be parallel with the rational inattention literature, I let the 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 the cognitive cost for decision i is an arbitrary continuously differentiable convex function of $I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right)$.

Notation-wise, for the rest of the section, 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. Finally, I use $\tau_i^* = I\left(\boldsymbol{\omega_i^*}; \vec{\boldsymbol{\theta}}\right)$ to denote the cognitive capacity allocated for decision i in the optimum.

6.2 Revisiting the Illustrative Example

I first study the costly contemplation problem in the context of the quadratic example in Section 3. To illustrate, let me first consider the $N = 2$ case. Same as (5), 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 $i \in \{1, 2\}$, $\boldsymbol{\theta_i} \sim \mathcal{N}\left(\bar{\theta}_i, \sigma_{\theta_i}^2\right)$ 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 variance of the noise in these signals is free to choose, subject to the cognitive constraint in (31).

Proposition 11 *In the optimum of the costly contemplation problem in (30):*

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

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

Proposition 11 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, in which self i has a more precise signal about her local fundamental $\boldsymbol{\theta_i}$ than other selves.[45]

In the appendix, I also work out the analogue of Proposition 11 in the N-decisions case. A result similar to Proposition 11 can be established in the case that u is symmetric across i. 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 τ is small enough, it is optimal for each self i to only receive signal about her local fundamental $\boldsymbol{\theta_i}$. That is, in the optimum, self i's signals about the others' local fundamental are completely uninformative. This is Proposition 24 in the Appendix.

6.3 An Alternative: Flexible Information Acquisition

Environment. In the previous subsection, I restrict each potential signal $\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}$ to have a particular form: each $\boldsymbol{\omega_i}$ consists of N noisy signals, one for each $\boldsymbol{\theta_j}$, $j\in\{1,\cdots,N\}$. This is consistent with the information structure studied in the rest of the paper. An alternative is to let the set of potential signals, $\boldsymbol{\Omega_i}$, be unrestrictive. That is, it is possible for the signal to depend

[45] As the optimal decision rules in this illustrative example are the same as those in the quasi-linear consumer theory context in Section 4.2, Proposition 11 also applies to the quasi-linear consumer theory context (barring an approximation). In that context, it means that self 1's signal about her $\boldsymbol{p_1}$ is more precise than self 2's signal about $\boldsymbol{p_1}$, and vice versa.

on the fundamental arbitrarily. With such flexible form of information acquisition, I can achieve a sharp characterization about the optimum of the costly contemplation problem (30) 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 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.

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, τ_i^*.

Lemma 4 *With unrestricted $\boldsymbol{\Omega_i}$, in the optimum of the costly contemplation problem in (30), each decision i is based on an one-dimensional signal $\boldsymbol{s_i^*}$:*[46]

$$\boldsymbol{\omega_i^*} = \{\boldsymbol{s_i^*}\} \quad \textit{and} \quad \boldsymbol{s_i^*} = \boldsymbol{\vartheta_i} + \mathbb{E}\left[\sum_{j\neq i} \gamma_{i,j} x_j^* \left(\boldsymbol{\omega_j^*}\right) |\vec{\boldsymbol{\theta}}\right] + \boldsymbol{\epsilon_i} \equiv \boldsymbol{t_i} + \boldsymbol{\epsilon_i}. \tag{32}$$

In (32), $\boldsymbol{\epsilon_i} \sim N\left(0, \sigma_i^2\right)$ is the idiosyncratic noise in the signal, σ_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 (32), 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.[47]

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 4 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 τ_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 (32), one can then solve optimal decision rules $\{x_i^*(\cdot)\}_{i=1}^N$. From (32), we know each self's optimal signal in turn depends on other selves' optimal decision rules.

[46]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 β are scalars.

[47]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.

Solving this fixed-point problem, one can then characterize how each optimal signal $\boldsymbol{s_i^*}$ depends on the fundamental $\vec{\boldsymbol{\theta}}$.

Proposition 12 *The optimal signals depend on the fundamental* $\vec{\boldsymbol{\theta}}$ *as follows:*

$$\begin{pmatrix} E\left[\boldsymbol{s_1^*}|\vec{\boldsymbol{\theta}}\right] \\ \\ \cdots \\ \\ E\left[\boldsymbol{s_N^*}|\vec{\boldsymbol{\theta}}\right] \end{pmatrix} = \begin{pmatrix} \boldsymbol{t_1} \\ \\ \cdots \\ \\ \boldsymbol{t_N} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1 & \lambda_1 \\ \lambda_2 & 1 & \cdots & \lambda_2 & \lambda_2 \\ & & \cdots & & \\ \lambda_{N-1} & \lambda_{N-1} & \cdots & 1 & \lambda_{N-1} \\ \lambda_N & \lambda_N & \cdots & \lambda_N & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} \boldsymbol{\vartheta_1} \\ \cdots \\ \boldsymbol{\vartheta_k} \\ \cdots \\ \boldsymbol{\vartheta_N} \end{pmatrix}, \tag{33}$$

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* τ_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 λ_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 (33), 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 λ 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 (τ is small so λ 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 24 discussed above.[48]

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

[48]Proposition 24 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.

Proposition 13 *In the optimum of the costly contemplation problem in (30),*

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

Proposition 13 shows that more volatile decisions (with high $Var\left(\boldsymbol{t_i}\right)$) and decisions with respect to 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 $\boldsymbol{t_i}$).[49]

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 I use throughout the paper 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, such as the attenuation of cross-demand elasticity, mental accounting, over- and under- reaction.

[49] One may wonder whether it is the case that $\tau_i^* > \tau_j^* \iff \left|\frac{\partial^2 u}{\partial x_i^2}\right| Var\left(\boldsymbol{v_i}\right) > \left|\frac{\partial^2 u}{\partial x_j^2}\right| Var\left(\boldsymbol{v_j}\right)$, 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 js 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.

Appendix A: Proofs

Proof of Lemma 1. If the optimum of (2) is not unique,[50] 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

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

proposition is essentially Theorem 1 in Chapter 5 of Marschak and Radner (1972).

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 θ_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,

$$x_i^*(\omega_i) = \sum_{k=1}^{N} \alpha_{i,k} s_{i,k}, \tag{34}$$

with a reminder that $s_{i,i} = \theta_i$, i.e. each self i has a perfect signal about θ_i.

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

$$E_i[s_{j,k}] = E_i[\theta_k] = E[\theta_k | s_{i,k}] = \lambda_{i,k} s_{i,k}.$$

We then have

$$E_i\left[x_j^*\right] = \sum_{k=1}^{N} \lambda_{i,k} \alpha_{j,k} s_{i,k}. \tag{35}$$

Together with the optimal decision rule in (7) and the guess in (34), we have, for all i,

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

For the guess in (34) to be valid, we then need to have, for all i, k,

$$\alpha_{i,k} = 1_{i=k} + \sum_{j \neq i} \lambda_{i,k} \gamma_{i,j} \alpha_{j,k}. \tag{36}$$

(36) are satisfied when

$$\begin{pmatrix} \alpha_{1,k} \\ \alpha_{2,k} \\ \cdots \\ \\ \alpha_{N,k} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,k} & \cdots & \lambda_{1,k} & \lambda_{1,k} \\ \lambda_{2,k} & 1 & \cdots & \lambda_{2,k} & \lambda_{2,k} \\ & & \cdots & & \\ \\ \lambda_{N,k} & \lambda_{N,k} & \cdots & \lambda_{N,k} & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} 0 \\ \cdots \\ 1 \\ \cdots \\ 0 \end{pmatrix}.$$

This verifies that the guess in (34) indeed characterizes the narrow thinker's optimal decision rules. Because different θ_is are independent, Proposition 2 then follows immediately from (34) and (35). To prove Proposition 3, note that based on the definition in (9), 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 θ_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 condition (36) can 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,

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

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. For all $i_1, \cdots, i_k \in \{1, \cdots, N\}$,where $i_l \neq i_{l+1}$ for $1 \leq l \leq k-1$, we have

$$E_{i_1}\left[E_{i_2}\left[\cdots E_{i_{k-1}}\left[\theta_{i_k}\right]\right]\right] = \lambda_{i_1,i_k} \cdots \lambda_{i_{k-1},i_k} s_{i_1,i_k}, \tag{37}$$

and

$$E\left[E_{i_1}\left[E_{i_2}\left[\cdots E_{i_{k-1}}\left[\theta_{i_k}\right]\right]\right] |\vec{\theta}\right] = \lambda_{i_1,i_k} \cdots \lambda_{i_{k-1},i_k} \theta_{i_k}. \tag{38}$$

From conditions (12) and (38), we have

$$\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k}\left(\theta_1, \cdots, \theta_N\right) = 1_{i=k} + \sum_{j \neq i} \lambda_{i,k} \gamma_{i,j} 1_{j=k} + \sum_{j \neq i} \lambda_{i,k} \gamma_{i,j} \sum_{l \neq j} \lambda_{j,k} \gamma_{j,l} 1_{l=k} + \cdots .$$

Using the fact that each λ is a factor between 0 and 1 and collecting terms with positive and negative coefficients prove Proposition 4.

Proof of Proposition 5. Taking an unconditional expectation, averaging over the realization of all fundamentals and signals, of condition (7), we have

$$E\left[x_i^{\text{Narrow}}\right] = E\left[\theta_i\right] + \sum_{j\neq i} \gamma_{i,j} E\left[x_j^{\text{Narrow}}\right] \quad \forall i,$$

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

$$E\left[x_i^{\text{Standard}}\right] = E\left[\theta_i\right] + \sum_{j\neq i} \gamma_{i,j} E\left[x_j^{\text{Standard}}\right] \quad \forall i.$$

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

Proof of Propositions 6 and 7. In the proof, for notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. From the optimal consumption decisions (17) and similar to the proof of Proposition 3, we have

$$\begin{aligned}
\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} &= -\psi + \gamma \sum_{l\neq i} \frac{\partial x_l^{\text{Narrow}}}{\partial p_i} \quad \forall i,\\
\frac{\partial x_i^{\text{Narrow}}}{\partial p_j} &= \lambda_{i,j}\gamma \sum_{l\neq i} \frac{\partial x_l^{\text{Narrow}}}{\partial p_j} \quad \forall i \neq j,
\end{aligned}$$

Solving the above two equations, we have

$$\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} = -\frac{\psi}{1+\gamma} - \frac{\gamma}{1+\gamma} \frac{\frac{\psi}{1+\gamma}}{1 - \frac{\gamma}{1+\gamma} - \sum_{l\neq i} \frac{\lambda_{l,i}\gamma}{1+\lambda_{l,i}\gamma}} \quad \forall i, \tag{39}$$

$$\frac{\partial x_j^{\text{Narrow}}}{\partial p_i} = \frac{\lambda_{j,i}\gamma}{1+\lambda_{j,i}\gamma} \frac{-\frac{\psi}{1+\gamma}}{1 - \frac{\gamma}{1+\gamma} - \sum_{l\neq i} \frac{\lambda_{l,i}\gamma}{1+\lambda_{l,i}\gamma}} \quad \forall i \neq j. \tag{40}$$

Case 1: $\gamma > 0$. Using the fact that $\gamma < \frac{1}{N-1}$ and that $1 - \frac{\gamma}{1+\gamma} - \sum_{l\neq i} \frac{\lambda_{l,i}\gamma}{1+\lambda_{l,i}\gamma}$ is larger than its standard counterpart (all λs are 1), Propositions 6 and 7 follow directly from (39) and (40) when $\gamma > 0$.

Case 2: $\gamma < 0$. Let $\gamma' = -\gamma \in (0, 1)$

$$\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} = -\frac{\psi}{1-\gamma'} + \frac{\gamma'}{1-\gamma'} \frac{\frac{\psi}{1-\gamma'}}{\frac{1}{1-\gamma'} + \sum_{l\neq i} \frac{\lambda_{l,i}\gamma'}{1-\lambda_{l,i}\gamma'}} \quad \forall i.$$

$$\frac{\partial x_j^{\text{Narrow}}}{\partial p_i} = \frac{\lambda_{j,i}\gamma'}{1-\lambda_{j,i}\gamma'} \frac{\frac{\psi}{1-\gamma'}}{\frac{1}{1-\gamma'} + \sum_{l\neq i} \frac{\lambda_{l,i}\gamma'}{1-\lambda_{l,i}\gamma'}} \quad \forall i \neq j,$$

Using the fact that $\gamma' < 1$ and that $\frac{1}{1-\gamma'} + \sum_{l\neq i} \frac{\lambda_{l,i}\gamma'}{1-\lambda_{l,i}\gamma'}$ is smaller than its standard counterpart (all λs are 1), we have $\frac{\partial x_i^{\text{Standard}}}{\partial p_i} < \frac{\partial x_i^{\text{Narrow}}}{\partial p_i}$ for all i. Together with the fact that $\frac{\gamma'}{1-\gamma'} \frac{\frac{\psi}{1-\gamma'}}{\frac{1}{1-\gamma'}+\sum_{l\neq i} \frac{\lambda_{l,i}\gamma'}{1-\lambda_{l,i}\gamma'}} \leq \frac{\psi\gamma'}{1-\gamma'}$, we know $\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} < 0$ for all i. This proves Proposition 7 for $\gamma < 0$.

For cross-price elasticities (Proposition 6), if, for all $i \neq j$, $\lambda_{j,i} = \lambda$, we have $\frac{\partial x_j^{\text{Standard}}}{\partial p_i} > \frac{\partial x_j^{\text{Narrow}}}{\partial p_i} = \frac{\lambda\gamma'}{1-\lambda\gamma'} \frac{\frac{\psi}{1-\gamma'}}{\frac{1}{1-\gamma'}+\frac{(N-1)\lambda\gamma'}{1-\lambda\gamma'}} > 0$. Similarly, when $N = 2$, $\frac{\partial x_j^{\text{Standard}}}{\partial p_i} > \frac{\partial x_j^{\text{Narrow}}}{\partial p_i} = \frac{\lambda_{j,i}\gamma'}{1-\lambda_{j,i}\gamma'} \frac{\frac{\psi}{1-\gamma'}}{\frac{1}{1-\gamma'}+\frac{\lambda_{j,i}\gamma'}{1-\lambda_{j,i}\gamma'}} > 0$. Without such symmetry and when $N \geq 3$, it is possible that $\frac{\partial x_j^{\text{Narrow}}}{\partial p_i} > \frac{\partial x_j^{\text{Standard}}}{\partial p_i} > 0$ for a particular $i \neq j$. To see this, fix $i \neq j$, let $\lambda_{l,i} = 0$ for all $l \neq i, j$ and let $\lambda_{j,i} = 1$. We then have $\frac{\partial x_j^{\text{Narrow}}}{\partial p_i} = \frac{\gamma'}{1-\gamma'} \frac{\frac{\psi}{1-\gamma'}}{\frac{1}{1-\gamma'}+\frac{\gamma'}{1-\gamma'}} > \frac{\partial x_j^{\text{Standard}}}{\partial p_i} = \frac{\gamma'}{1-\gamma'} \frac{\frac{\psi}{1-\gamma'}}{\frac{1}{1-\gamma'}+\frac{2\gamma'}{1-\gamma'}}$.

Proof of Lemma 3. Under standard consumer theory, we have

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

Under narrow thinking, if condition (25) holds with equality for each pair of (i, j), averaging over the realization of noises in signals, we then have

$$-\kappa_i \hat{x}_i^{\text{Narrow}}(\hat{p}_1, \cdots, \hat{p}_N) - \hat{p}_i = -\kappa_j \hat{x}_j^{\text{Narrow}}(\hat{p}_1, \cdots, \hat{p}_N) - \hat{p}_j \quad \forall i \neq j, \tag{41}$$

and

$$-\kappa_i \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_i} - 1 = -\kappa_j \frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{p}_i} \quad \forall i \neq j. \tag{42}$$

This is inconsistent with the formula about price elasticities in the proof of Proposition 8.

Proof of Proposition 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 conditions (22) and (23), and similar to the proof of Proposition 3, we have

$$\begin{aligned}
-\kappa_k \frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_k} &= 1 + \frac{\kappa_h}{\mu_y}\left(\sum_{j=1}^{N} \mu_j \frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{p}_k} + \mu_k\right) \quad \forall k, \\
-\kappa_i \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{p}_k} &= \frac{\kappa_h}{\mu_y}\left[\mu_i \frac{\partial x_i^{\text{Narrow}}}{\partial p_k} + \lambda_{i,k}\left(\sum_{j\neq i} \mu_j \frac{\partial \hat{x}_j^{\text{Narrow}}}{\partial \hat{p}_k} + \mu_k\right)\right] \quad \forall i \neq k,
\end{aligned}$$

where $\lambda_{i,k} = \frac{\sigma_{p_k}^2}{\sigma_{p_k}^2 + \sigma_{i,k}^2}$. We have

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

$$\frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_k} = \frac{\kappa_h}{\kappa_k} \frac{\mu_k \frac{1-\kappa_k}{\kappa_k}}{\left(\sum_{j\neq k}\mu_j \frac{\mu_y \lambda_{j,k}}{\mu_y\kappa_j + (1-\lambda_{j,k})\kappa_h\mu_j} + \frac{\mu_k}{\kappa_k}\right)\kappa_h + \mu_y} - \frac{1}{\kappa_k} \quad \forall k. \tag{43}$$

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 λs 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 $\frac{\partial \hat{x}_k^{\text{Standard}}}{\partial \hat{p}_k} < \frac{\partial \hat{x}_k^{\text{Narrow}}}{\partial \hat{p}_k}$.

Finally, I will 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), $p_k x_k^{MB} = M_k \ \forall k$. As a result, $\frac{\partial \hat{x}_k^{\text{MB}}}{\partial \hat{p}_k} = -1$. Then note that, from (43), 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 λs 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 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. The optimality condition for each decision i and the budget constraint become

$$\varphi_i - \kappa_i x_i^*(\omega_i) = -\kappa_h E_i[y^*] \qquad \forall i,$$
$$\sum_{i=1}^{N} \mu_i x_i^*(\omega_i) + \mu_y y^* = 0,$$

where, as in Section 4.3, $\kappa_h = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$. Now consider shocks to φ_k. Similar to the proof of Proposition 8, we have

$$1 - \kappa_k \frac{\partial x_k^{\text{Narrow}}}{\partial \varphi_k} = \frac{\kappa_h}{\mu_y}\left(\sum_{j=1}^{N}\mu_j \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_k}\right) \quad \forall k,$$
$$-\kappa_i \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} = \frac{\kappa_h}{\mu_y}\left[\mu_i \frac{\partial x_i^L}{\partial \varphi_k} + \lambda_{i,k}\left(\sum_{j\neq i}\mu_j \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_k}\right)\right] \quad \forall i \neq k,$$

where $\lambda_{i,k} = \frac{\sigma^2_{\varphi_k}}{\sigma^2_{\varphi_k}+\sigma^2_{i,k}}$. Solving the above two equations, we have

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

$$\frac{\partial x_k^{\text{Narrow}}}{\partial \varphi_k} = -\frac{\kappa_h}{\kappa_k} \frac{\frac{\mu_k}{\kappa_k}}{\left(\sum_{j\neq k}\mu_j \frac{\mu_y\lambda_{j,k}}{\mu_y\kappa_j+(1-\lambda_{j,k})\kappa_h\mu_j} + \frac{\mu_k}{\kappa_k}\right)\kappa_h + \mu_y} + \frac{1}{\kappa_k} \quad \forall k. \tag{44}$$

Note that $\left(\sum_{j\neq k}\mu_j \frac{\mu_y\lambda_{j,k}}{\mu_y\kappa_j+(1-\lambda_{j,k})\kappa_h\mu_j} + \frac{\mu_k}{\kappa_k}\right)$ is smaller than its standard counterpart (all λs are 1). As a result, $\frac{\partial x_k^{\text{Narrow}}}{\partial \varphi_k} < \frac{\partial x_k^{\text{Standard}}}{\partial \varphi_k}$. Moreover,

$$\begin{aligned}\frac{\partial x_k^{\text{Narrow}}}{\partial \varphi_k} &= \frac{1}{\kappa_k}\left(-\frac{\kappa_h\frac{\mu_k}{\kappa_k}}{\left(\sum_{j\neq k}\mu_j \frac{\mu_y\lambda_{j,k}}{\mu_y\kappa_j+(1-\lambda_{j,k})\kappa_h\mu_j} + \frac{\mu_k}{\kappa_k}\right)\kappa_h + \mu_y} + 1\right) \\ &\geq \frac{1}{\kappa_k}\left(-\frac{\kappa_h\frac{\mu_k}{\kappa_k}}{\frac{\mu_k}{\kappa_k}\kappa_h + \mu_y} + 1\right) > 0.\end{aligned}$$

This proves Proposition 9.

Proof of Proposition 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 $\kappa_h = -\frac{h''(\bar{y})\bar{y}}{h'(\bar{y})}$.

Using the definition in (29) and averaging over the realizations of noises in signals, we have

$$-\kappa_k \frac{\partial x_k^{\text{Narrow}}}{\partial w_k} = \kappa_h \frac{\bar{x}_k}{\bar{y}}\left(\sum_{j=1}^N p_j \frac{\partial x_j^{\text{Narrow}}}{\partial w_k} - 1\right) \quad \forall k,$$

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

where $\lambda_{i,k} = \frac{\sigma_{w_k}^2}{\sigma_{w_k}^2+\sigma_{i,k}^2}$. Together, we have

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

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

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

To prove part 1, note that, if and only if when all λs 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 $i \neq k$.

To prove part 2, note that $\sum_{j\neq k} \frac{\lambda_{j,k}\kappa_h \frac{p_j \bar{x}_j}{\bar{y}}}{\kappa_j + \kappa_h (1-\lambda_{j,k}) \frac{p_j \bar{x}_j}{\bar{y}}}$ is smaller than its standard counterpart (all λs are 1). As a result, $\frac{\partial x_k^{\text{Narrow}}}{\partial w_k} > \frac{\partial x_k^{\text{Standard}}}{\partial w_k}$ for all k.

Proof of Proposition 11. As discussed in the main text, the costly contemplation problem in (30) can be divided into two subproblems, the optimal information choice subject to the cognitive constraint in (31), 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

$$E\left[x_i^*(\omega_i) - \theta_i - \gamma x_{-i}^*(\omega_{-i})\,|\omega_i\right] = 0 \quad \forall i, \omega_i \in \Omega_i. \tag{45}$$

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 (31) is then equivalent to

$$\max_{\{\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}\}_{i=1}^2} \frac{1}{2} E\left[\theta_1 x_1^*(\omega_1) + \theta_2 x_2^*(\omega_2)\right] \tag{46}$$
$$s.t. \quad x_i^*(\omega_i) \text{ satisfy (45)}$$
$$\sum_{i=1}^2 I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right) \leq \tau.$$

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 (46) then becomes

$$\max_{\{0 \le \lambda_{i,j} \le 1\}_{1 \le i,j \le 2}} g\left(\{\lambda_{i,j}\}_{1 \le i,j \le 2}\right) \equiv \frac{1}{2} \frac{\lambda_{1,1}}{1 - \lambda_{1,1}\lambda_{2,1}\gamma^2} \sigma_{\theta_1}^2 + \frac{1}{2} \frac{\lambda_{2,2}}{1 - \lambda_{2,2}\lambda_{1,2}\gamma^2} \sigma_{\theta_2}^2 \tag{47}$$
$$s.t.\ h\left(\{\lambda_{i,j}\}_{1 \le i,j \le 2}\right) \equiv \sum_{1 \le i,j \le 2} \frac{1}{2} log_2 \left(\frac{1}{1 - \lambda_{i,j}}\right) \le \tau,$$

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 *11.* If, in the optimum of the costly contemplation problem, we have $\left(\sigma_{1,1}^*\right)^2 \ge \left(\sigma_{2,1}^*\right)^2$. This means $\lambda_{1,1}^* \le \lambda_{2,1}^*$, where $\lambda_{i,j}^* = \frac{\sigma_{\theta_j}^2}{\sigma_{\theta_j}^2 + \left(\sigma_{i,j}^*\right)^2} \in (0,1]$. As a result, $\frac{\partial g\left(\{\lambda_{i,j}^*\}_{1 \le i,j \le 2}\right)}{\partial \lambda_{1,1}^*} > \frac{\partial g\left(\{\lambda_{i,j}^*\}_{1 \le i,j \le 2}\right)}{\partial \lambda_{2,1}^*}$ and $\frac{\partial h\left(\{\lambda_{i,j}^*\}_{1 \le i,j \le 2}\right)}{\partial \lambda_{1,1}^*} \le \frac{\partial h\left(\{\lambda_{i,j}^*\}_{1 \le i,j \le 2}\right)}{\partial \lambda_{2,1}^*}$. This is inconsistent with the first order condition of (47):

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

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

Proof of Lemma 4. A necessary condition for $\{\boldsymbol{\omega_i^*}, x_i^*(\cdot)\}_{i=1}^N$ to be an optimum of the costly contemplation problem in (30) is that, for each i, $(\boldsymbol{\omega_i^*}, x_i^*(\cdot))$ is optimally chosen, taking other $\left\{\boldsymbol{\omega_j^*}, x_j^*(\cdot)\right\}_{j \ne i}$ as given. That is, $\left(\boldsymbol{\omega_i^*}, x_i^*(\cdot)\right)$ solves

$$\max_{\boldsymbol{\omega}_i \in \boldsymbol{\Omega}_i, x_i(\cdot)} E\left[u\left(x_1(\omega_1), \cdots, x_N^*(\omega_N^*), \vec{\theta}\right)\right] \tag{48}$$

$$s.t. \quad I\left(\boldsymbol{\omega_i}; \vec{\boldsymbol{\theta}}\right) \leq \tau - \sum_{j \neq i} I\left(\boldsymbol{\omega_j^*}; \vec{\boldsymbol{\theta}}\right), \tag{49}$$

As u is quadratic, maximizes the objective in (48) is equivalent to maximizing

$$E\left[\frac{u_{i,i}}{2}\left(x_i(\omega_i) - \vartheta_i - \sum_{j \neq i} \gamma_{i,j} E\left[x_j^*\left(\omega_j^*\right) | \vec{\theta}\right]\right)^2 + \frac{u_{i,i}}{2}\left(\sum_{j \neq i} \gamma_{i,j}\left(x_j^*\left(\omega_j^*\right) - E\left[x_j^*\left(\omega_j^*\right) | \vec{\theta}\right]\right)\right)^2 + f\left(\left\{x_j^*\left(\omega_j^*\right)\right\}_{j \neq i}, \vec{\theta}\right)\right] \tag{50}$$

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 (50) and the constraint in (49) 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 4.

Proof of Proposition 12. 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 $x_i^*\left(\boldsymbol{\omega_i^*}\right) = E\left[\boldsymbol{\vartheta_i} + \sum_{j \neq i} \gamma_{i,j} x_j^*\left(\boldsymbol{\omega_j^*}\right) | \boldsymbol{s_i^*}\right] = \lambda_i \boldsymbol{s_i^*}$, where $\lambda_i = \frac{\sigma_{t_i}^2}{\sigma_i^2 + \sigma_{t_i}^2}$. Together with (32), we have

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

This leads to (33).

Proof of Proposition 13. 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 $\{\boldsymbol{t_i}\}_{i=1}^N$ are given by

$$\begin{pmatrix} \boldsymbol{t_1} \\ \\ \cdots \\ \\ \boldsymbol{t_N} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \Lambda_1 & \cdots & \Lambda_1 & \Lambda_1 \\ \Lambda_2 & 1 & \cdots & \Lambda_2 & \Lambda_2 \\ & & \cdots & & \\ \Lambda_{N-1} & \Lambda_{N-1} & \cdots & 1 & \Lambda_{N-1} \\ \Lambda_N & \Lambda_N & \cdots & \Lambda_N & 1 \end{pmatrix} \circ \Gamma\right)^{-1} \begin{pmatrix} \boldsymbol{\vartheta_1} \\ \cdots \\ \boldsymbol{\vartheta_k} \\ \cdots \\ \boldsymbol{\vartheta_N} \end{pmatrix}. \tag{51}$$

Based on Lemma 4 and Proposition 12, in the optimum of the costly contemplation problem in (30), $\left\{\lambda_i, \sigma_i^2\right\}_{i=1}^N$ defined in Lemma 4 and Proposition 12 must solve

$$\max_{\left\{\Lambda_i, \varsigma_i^2\right\}_{i=1}^N} g\left(\left\{\Lambda_i, \varsigma_i^2\right\}_{i=1}^N\right) \tag{52}$$
$$h\left(\left\{\varsigma_i^2\right\}_{i=1}^N\right) = \frac{1}{2}\sum_i \log_2 \frac{\varsigma_i^2 + \sigma_{t_i}^2}{\varsigma_i^2} \leq \tau,$$

where $\sigma_{t_i}^2$ is the variance of $\boldsymbol{t}_i$ defined based on (51). Rewrite g in a way similar to (50), 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 (47), in the optimum of the costly contemplation problem in (30), 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 13 follows.

Appendix B: Consumer Theory under Narrow Thinking: The Case Without Income Effects

Identification of Demand Gradients.

Consider an environment with K consumers. All consumers have the same utility as (15). 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 4.2, 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 $\{\lambda_{i,j}\}$) in each period. As different consumers have the same utility and same λs, 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.

Appendix C: Consumer Theory under Narrow Thinking: Income Effects and the Violation of the Fungibility Principle

Consumer with General Non-quasi Linear Utility.

Consider the general consumer theory under narrow thinking set up at the start of Section 4. As shown in (14), it can be translated into the general unconstrained multiple-decision problem in (2) introduced in Section 2. For this general multiple-decision problem, the results about how narrow thinking leads to an effective attenuation of interaction across decisions and a dampening of indirect effects are established in Propositions 16 and 17 in Appendix D.

Attenuation of the Cross-Price Demand Elasticity with Income Effects.

Here I establish the attenuation of cross-price demand elasticity under narrow thinking in 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.[51] 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 4. 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_j. To summarize, for $i \in \{1, \cdots, N\}$, self i's signal is given by $\omega_i = \left\{\{s_{i,j}\}_{j\in\{1,\cdots,N\}}\right\}$, where $s_{i,i} = p_i$ and, for all $i \neq j$, $s_{i,j} = p_j \epsilon_{i,j}$, with $\epsilon_{i,j} \sim \log\mathcal{N}\left(0, \sigma^2\right)$ and $\sigma^2 > 0$. All ϵs and ps are independent from each other. The last self, who is in charge of the consumption of y, has perfect knowledge of the fundamentals and

[51] 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.

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_j have the same signal-to-noise ratio, i.e. $\lambda_{i,j} = \frac{\sigma_p^2}{\sigma_p^2+\sigma^2}$ are the same for all $i \neq j$.

Similar to condition (21), the optimal consumption decision of each self $i \in \{1, \cdots, N\}$, $x_i^*(\omega_i)$, must satisfy

$$E_i\left[\frac{\partial v}{\partial x_i}\left(x_i^*(\omega_i), \vec{x}_{-i}^*\right)\right] = p_i E_i\left[h^{'}(y^*)\right]. \tag{53}$$

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 prices, we have

$$\hat{x}_i^*(\omega_i) = -\psi^{'}\hat{p}_i + \sum_{j\neq i}\gamma^{'}E_i\left[\hat{x}_j^*(\omega_j)\right] + \kappa_h E_i\left[\hat{y}^*\right], \tag{54}$$

$$\sum_{i=1}^{N}\mu\left(\hat{x}_i^*(\omega_i) + \hat{p}_i\right) + \mu_y\hat{y}^* = 0, \tag{55}$$

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 $\kappa_h = \frac{\frac{h^{''}(\bar{y})\bar{y}\frac{\partial v(\bar{x}_1,\cdots,\bar{x}_N)}{\partial x_i}}{h^{'}(\bar{y})}}{\frac{\partial^2 v(\bar{x}_1,\cdots,\bar{x}_N)}{\partial x_i^2}\bar{x}_i} > 0$. $\mu = \frac{\bar{p}_i\bar{x}_i}{w}$ 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

$$\hat{x}_i^*(\omega_i) = -\psi\hat{p}_i - \sum_{j\neq i}\Psi E_i\left[\hat{p}_j\right] + \sum_{j\neq i}\gamma E_i\left[\hat{x}_j^*(\omega_j)\right], \tag{56}$$

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 $\gamma = \frac{\gamma^{'}-\frac{\kappa_h\mu}{\mu_y}}{1+\frac{\kappa_h\mu}{\mu_y}} \in \left(-1, \frac{1}{N-1}\right)$. Similar to Section 4.3, 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. I then establish the attenuation of cross-price demand elasticity under narrow thinking in this set-up.

Proposition 14 *The cross-price demand elasticities are attenuated under narrow thinking:*

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

where I use superscript Standard to denote standard consumer theory's demand function when each decision is made with perfect knowledge of all prices.

Proof of Proposition 14. In the proof, for notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. Similar to the proof of Proposition 8, given the optimal consumption rule in (56), we have

$$\begin{aligned}
\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} &= -\psi + \gamma \sum_{l \neq i} \frac{\partial x_l^{\text{Narrow}}}{\partial p_i} \quad \forall i, \\
\frac{\partial x_i^{\text{Narrow}}}{\partial p_j} &= -\lambda\Psi + \lambda\gamma \sum_{l \neq i} \frac{\partial x_l^{\text{Narrow}}}{\partial p_j} \quad \forall i \neq j,
\end{aligned}$$

where $\lambda = \frac{\sigma_p^2}{\sigma^2 + \sigma_p^2}$. We then have

$$\frac{\partial x_i^{\text{Narrow}}}{\partial p_i} = -\psi + \gamma (N-1) \frac{-\lambda\Psi - \lambda\gamma\psi}{1 - \lambda\gamma^2 (N-1) - \lambda\gamma (N-2)}, \tag{57}$$

$$\frac{\partial x_i^{\text{Narrow}}}{\partial p_j} = \frac{-\lambda\Psi - \lambda\gamma\psi}{1 - \lambda\gamma^2 (N-1) - \lambda\gamma (N-2)}. \tag{58}$$

Using the fact that $\lambda \in [0, 1)$, $\psi > \Psi > 0$ and $\gamma \in \left(-1, \frac{1}{N-1}\right)$,[52] Proposition 14 follows directly.

Own-demand Elasticity with Income Effects: Over-reaction vs. Under-reaction

In the symmetric consumer theory with income effects context studied in Proposition 14 above, I formalize the following rule-of-thumb useed in the main text: 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 of the direct effect, a dampening of the indirect effect under narrow thinking often leads to over-reaction.

First note that from (57), standard consumer theory's own-price elasticity can be characterized by:

$$\frac{\partial x_i^{\text{Standard}}}{\partial p_i} = \underbrace{-\psi}_{\text{direct effect}} - \underbrace{\frac{(N-1)\gamma(\Psi + \gamma\psi)}{1 - \gamma^2 (N-1) - \gamma (N-2)}}_{\text{indirect effect}}.$$

Proposition 15 *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:*

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

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

2. *When the indirect effect works in the opposite direction of 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:*

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

Proof of Proposition 15. Using the fact that $\lambda \in [0,1)$, $\psi > \Psi > 0$, and $\gamma \in \left(-1, \frac{1}{N-1}\right)$, we know $1-\gamma^2(N-1)-\gamma(N-2) > 0$, $1-\lambda\gamma^2(N-1)-\lambda\gamma(N-2) > 0$, and $\frac{\partial \hat{x}_i^S}{\partial \hat{p}_i} < 0$ always. The results then follow directly from conditions (57) and (58).

Appendix D: Beyond the Response to Price Changes

The Effective Attenuation of Interaction

Each self's best response function. Consider the general multiple-decision problem in Section 2, where the decision maker's utility depends on N decisions $\vec{x} = (x_1, \cdots, x_N) \in \mathbb{R}^N$ and the fundamentals $\vec{\theta} = (\theta_1, \cdots, \theta_M) \in \mathbb{R}^M$:

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

where u is strictly concave over $\vec{x}$. To analytically characterize the narrow thinker's behavior, I also let u be quadratic. Without 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.

Taking the first order condition of the decision-specific optimality condition in (3) and collecting terms, the optimal decision rule for each self i can be summarized as

$$x_i^*(\omega_i) = E_i\left[\sum_{1\le k\le M} \psi_{i,k}\theta_k + \sum_{j\neq i} \gamma_{i,j} x_j^*(\omega_j)\right] \quad \forall i, \omega_i \in \Omega_i, \tag{59}$$

where $\psi_{i,k} = -\frac{\partial^2 u}{\partial x_i \partial \theta_k}\left(\frac{\partial^2 u}{\partial x_i^2}\right)^{-1}$ and $\gamma_{i,j} = -\frac{\partial^2 u}{\partial x_i \partial x_j}\left(\frac{\partial^2 u}{\partial x_i^2}\right)^{-1}$. $\psi_{i,k}\theta_k$ summarizes how the k-th element of the fundamentals directly influences decision i. $\gamma_{i,j}$ captures how decision i is influenced by decision j. One can think of (59) as the best response function of a linear network game. The matrix $\Gamma = \{\gamma_{i,j}\}_{1\le i,j\le N}$ can then be interpreted as the interaction matrix.[53]

Alternatively, without quadratic utility, one can arrive at the linear decision rule in (59) by linearizing or log-linearizing the FOC of each decision.

Information. I study the following narrow thinker: each self i's information is given by $\omega_i =$

[53] I set $\gamma_{i,i} = 0$ for all i.

$\left\{\{s_{i,k}\}_{k\in\{1,\cdots,M\}}\right\}$, where $s_{i,k} = \theta_k + \epsilon_{i,k}$, $\theta_k \sim \mathcal{N}\left(\bar{\theta}_k, \sigma^2_{\theta_k}\right)$ and noises $\epsilon_{i,k} \sim \mathcal{N}\left(0, \sigma^2_{i,k}\right)$ are independent from the fundamentals and each other.

The information structure here nests those considered in the main text. For example, in the illustrative example studied in Section 3, the dimensionality of the fundamental $M = N$. There, I let each self $i \in \{1, \cdots, N\}$ of the narrow thinker perfectly know her own θ_i (i.e. $\sigma_{i.i} = 0$) and receive a signal about each of the other θ_k.

Finally, I let different θ_ks be uncorrelated. As also discussed in Section 3, the key additional channel with correlated θ_ks is "rational confusion:" to the extent that θ_k and θ_l are correlated, self i can use signals of θ_k to forecast θ_l and vise versa. Given the interpretation that the noise of self i's signal of each fundamental comes from cognitive frictions, one may not want to take into account such rational confusion considerations. In fact, when different θ_ks 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 θ_k is based on her signal about θ_k, $s_{i,k}$, solely. In this case, $E_i[\theta_k] = E[\theta_k|s_{i,k}]$.

Belief anchoring. Similar to Proposition 2, under narrow thinking, as each self has an imperfect perception of other selves' decisions, each self i's belief about how other decisions respond to shocks will be anchored. Specifically, consider self i's belief about the decision $j \neq i$, $E_i\left[x_j^*\right]$. In response to shocks to each θ_k,

$$E\left[E_i\left[x_j^*\right]|\theta_k\right] = \lambda_{i,k} E\left[x_j^*|\theta_k\right] + (1-\lambda_{i,k}) E\left[x_j^*\right] \qquad \forall k,\ \forall i \neq j, \tag{60}$$

where $\lambda_{i,k} = \frac{\sigma^2_{\theta_k}}{\sigma^2_{\theta_k} + \sigma^2_{i,k}} \in (0,1]$ is a function of the signal-to-noise ratio of self i's signal about θ_k, $s_{i,k}$. It captures how precise self i can predict how other decisions respond to θ_k.

Effective attenuation of interaction. I now turn to the narrow thinker's response to shocks to fundamentals. To achieve this goal, along with (9) in Section 3, I first define the narrow thinker's decision as a function of fundamentals $x_i^L(\theta_1, \cdots, \theta_M)$ as

$$x_i^{\text{Narrow}}(\theta_1, \cdots, \theta_M) \equiv E\left[x_i^*(\omega_i)|\theta_1, \cdots, \theta_M\right] \quad \forall i, \tag{61}$$

where $E[\cdot|\theta_1, \cdots, \theta_M]$ averages over the realization of noises in signals. I will then study how narrow thinker's decision x_i responds to shocks to each θ_k, that is, $\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k}$, for all i, k. In response to such shocks, the belief anchoring in (60) dampens the impact from other decisions x_j to x_i and leads to an effective attenuation of interaction.

Proposition 16 *The narrow thinker's response to shocks to each θ_k can be summarized as*[54]

$$\begin{pmatrix} \frac{\partial x_1^{\text{Narrow}}}{\partial \theta_k} \\ \frac{\partial x_2^{\text{Narrow}}}{\partial \theta_k} \\ \cdots \\ \\ \frac{\partial x_N^{\text{Narrow}}}{\partial \theta_k} \end{pmatrix} = \left(\mathbb{I}_N - \begin{pmatrix} 1 & \lambda_{1,k} & \cdots & \lambda_{1,k} & \lambda_{1,k} \\ \lambda_{2,k} & 1 & \cdots & \lambda_{2,k} & \lambda_{2,k} \\ & & \cdots & & \\ & & & & \\ \lambda_{N,k} & \lambda_{N,k} & \cdots & \lambda_{N,k} & 1 \end{pmatrix} \circ \Gamma \right)^{-1} \begin{pmatrix} \lambda_{1,k}\psi_{1,k} \\ \cdots \\ \lambda_{k,k}\psi_{k,k} \\ \cdots \\ \lambda_{N,k}\psi_{N,k} \end{pmatrix}, \tag{62}$$

where $\circ$ is element by element product.

The above condition means that, in response to shocks to θ_k, for each pair of decision (i, j), the effective degree of interaction from decision j to decision i is attenuated by a factor $\lambda_{i,k}$ between 0 and 1. That is, in response to shocks to θ_k, an one unit increase in x_j only effectively increases x_i by $\lambda_{i,k}\gamma_{i,j}$. It is as if each self cares less about other decisions' influence, and she "thinks narrowly."

Let me explain the difference compared to Proposition 3 in the main text. In the environment for Proposition 3, θ_k only has a direct effect on x_k. Thus, only the k-th element of the last row of (11), which summarizes the direct effect of θ_k, is non-zero. Moreover, as self k perfectly knows θ_k, the direct effect is maintained under narrow thinking. Thus, there are no λs in the last row of (11) . Here, I allow θ_k to directly influence each decision. As the result, each element of the last column in (62), which summarizes the direct effect, can be non-zero. Moreover, I allow the possibility that the direct effect can also be attenuated here. Such attenuation is summarized by the λs in the last row of (62).

Proof of Proposition 16. For notation simplicity, I normalize the mean of each θ_k, $\bar{\theta}_k$, 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,

$$x_i^*(\omega_i) = \sum_{k=1}^{M} \alpha_{i,k} s_{i,k}. \tag{63}$$

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

$$E_i[s_{j,k}] = E_i[\theta_k] = E[\theta_k | s_{i,k}] = \lambda_{i,k} s_{i,k}.$$

[54] Proposition 16 studies the narrow thinker's response to each θ_k. Sometimes, one is also interested in response to a different linear function of the fundamental vector $\vec{\theta}$, i.e., $\Theta = \sum_{k=1}^{M} v_k \theta_k$. As different θ_ks are independent and there is no rational confusion, we have $\frac{dx_i^{\text{Narrow}}}{d\Theta} = \sum_{k=1}^{M} v_k \frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k}$, where $\frac{\partial x_i^{\text{Narrow}}}{\partial \theta_k}$ can be calculated directly from Proposition 16.

We then have

$$E_i\left[x_j^*\right]=\sum_{k=1}^{M}\lambda_{i,k}\alpha_{j,k}s_{i,k}. \tag{64}$$

Together with the optimal decision rule in (59) and the guess in (63), we have, for all i,

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

For the guess in (63) to be valid, we then need to have, for all i,k,

$$\alpha_{i,k}=\lambda_{i,k}\psi_{i,k}+\sum_{j\neq i}\lambda_{i,k}\gamma_{i,j}\alpha_{j,k}. \tag{65}$$

(65) are satisfied when

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

This verifies that the guess in (63) indeed characterizes the narrow thinker's optimal decision rules. Because different θ_is are independent, the belief anchoring in (60) then follows immediately from (63) and (64). To prove Proposition 16, note that based on the definition in (61), we then have, for all i,

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

Taking partial derivative with respect to each θ_k then leads to Proposition 16.

Dampening Indirect Effects.

Similar to Proposition 4, one can also show that each self's imperfect perception of other decisions under narrow thinking leads to a dampening of indirect effects — the movement of one decision driven by the movement of other decisions.

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

Assumption 1 *The absolute value of all eigenvalues of* Γ *are less than one.*

Iterating the optimal decisions rule in condition (59), we have

$$\begin{aligned}
x_i^*(\omega_i) &= \sum_{k=1}^{M} \psi_{i,k} E_i[\theta_k] + \sum_{j\neq i} \gamma_{i,j} E_i\left[x_j^*\right] \\
&= \sum_{k=1}^{M} \psi_{i,k} E_i[\theta_k] + \sum_{j\neq i} \gamma_{i,j} \left(\sum_{k=1}^{M} \psi_{j,k} E_i[E_j[\theta_k]] \right) + \sum_{j\neq i} \gamma_{i,j} \left(\sum_{l\neq j} \gamma_{j,l} E_i\left[E_j\left[x_l^*\right]\right] \right) \\
&= \underbrace{\sum_{k=1}^{M} \psi_{i,k} E_i[\theta_k]}_{\text{Direct}} + \underbrace{\sum_{j\neq i} \gamma_{i,j} \left(\sum_{k=1}^{M} \psi_{j,k} E_i[\theta_k] \right) + \sum_{j\neq i} \gamma_{i,j} \left(\sum_{l\neq j} \gamma_{j,l} \left(\sum_{k=1}^{M} \psi_{l,k} E_i[E_j[\theta_k]] \right) \right) + \cdots}_{\text{Indirect}}.
\end{aligned} \tag{66}$$

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 θs, self i's belief about other selves' beliefs about other selves' belief about θs, 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 (61), 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 17 *Under Assumption 1, for each decision x_i, each part of its indirect effect is dampened under narrow thinking in response to shocks to each θ_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 17. 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

$$E_{i_1}[E_{i_2}[\cdots E_{i_m}[\theta_k]]] = \lambda_{i_1,k}\cdots\lambda_{i_m,k} s_{i_1,k}, \tag{67}$$

and

$$E\left[E_{i_1}[E_{i_2}[\cdots E_{i_m}[\theta_k]]]|\vec{\theta}\right] = \lambda_{i_1,k}\cdots\lambda_{i_m,k}\theta_k. \tag{68}$$

From conditions (66) and (68), 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 λ is a factor between 0 and 1 and collecting terms with positive and negative coefficients prove Proposition 16.

Comfort Zones

The decision maker's utility is given by,

$$\sum_{i=1}^{N} \varphi_i v_i(x_i) - c\left(\sum_{i=1}^{N} x_i\right), \tag{69}$$

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, φ_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 φ_j. This makes sure that the budget constraint always holds. Specifically, for $i \in \{1, \cdots, N\}$, self i's signal is given by $\omega_i = \left\{\{s_{i,j}\}_{j\in\{1,\cdots,N\}}\right\}$, where $s_{i,i} = \varphi_i$ and, for $i \neq j$, $s_{i,j} = \varphi_j \epsilon_{i,j}$, with $\epsilon_{i,j} \sim \log\mathcal{N}\left(0, \sigma_{i,j}^2\right)$ and $\sigma_{i,j}^2 > 0$. All ϵs and φs are independent from each other.

I use a hat over a variable to denote its log-deviation from the point of log-linearization.[55] Then, similar to condition (18), for all i, I define the narrow thinker's (log) decision function as a function of fundamentals:

$$\hat{x}_i^{\text{Narrow}}(\hat{\varphi}_1, \cdots, \hat{\varphi}_N) \equiv E\left[\hat{x}_i^*(\omega_i) \,|\hat{\varphi}_1, \cdots, \hat{\varphi}_N\right], \tag{70}$$

where $E[\cdot|\hat{\varphi}_1, \cdots, \hat{\varphi}_N]$ 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 then have:

Proposition 18 (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*

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

φ_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 Proposition, first consider the case that each decision is made with perfect knowledge of all the fundamentals. An increase in φ_i will increase x_i as activity i becomes more attractive, but decrease other x_j for $j \neq i$, as the cost function is convex over the sum of efforts. Under narrow thinking, such a coordinated response to φ_i is hindered: other selves' x_j will not decrease as much in response to the increase in φ_i. As a result, x_i will not increase as much. In other words, 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 φ_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.

Here, the direct effect of φ_i on x_i is positive. The indirect effect of φ_i on x_i, through the decrease of other x_j, is also positive. A dampening of the indirect effect under narrow thinking then leads to under-reaction. This is as suggested by the rule of thumb of when narrow thinking leads to under-reaction at the start of Section 5.

Proof of Proposition 18. 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

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

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. Similar to the proof of Proposition 9, we have

$$\begin{aligned} 1 - \kappa_k \frac{\partial x_k^{\text{Narrow}}}{\partial \varphi_k} &= \kappa_c \left(\sum_{j=1}^{N} \mu_j \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_k} \right) \qquad \forall k, \\ -\kappa_i \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} &= \kappa_c \left[\mu_i \frac{\partial x_i^{\text{Narrow}}}{\partial \varphi_k} + \lambda_{i,k} \left(\sum_{j \neq i} \mu_j \frac{\partial x_j^{\text{Narrow}}}{\partial \varphi_k} \right) \right] \qquad \forall i \neq k, \end{aligned}$$

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

$$\frac{\partial x_i^{\text{Narrow}}}{\partial\varphi_k}=-\frac{\lambda_{i,k}}{\kappa_i+(1-\lambda_{i,k})\,\kappa_c\mu_i}\frac{\frac{\mu_k}{\kappa_k}}{\left(\sum_{j\neq k}\mu_j\frac{\lambda_{j,k}}{\kappa_j+(1-\lambda_{j,k})\kappa_c\mu_j}+\frac{\mu_k}{\kappa_k}\right)+\frac{1}{\kappa_c}}\quad\forall k, \tag{71}$$

$$\frac{\partial x_k^{\text{Narrow}}}{\partial\varphi_k}=-\frac{1}{\kappa_k}\frac{\frac{\mu_k}{\kappa_k}}{\left(\sum_{j\neq k}\mu_j\frac{\lambda_{j,k}}{\kappa_j+(1-\lambda_{j,k})\kappa_c\mu_j}+\frac{\mu_k}{\kappa_k}\right)+\frac{1}{\kappa_c}}+\frac{1}{\kappa_k}\quad\forall k. \tag{72}$$

Note that $\left(\sum_{j\neq k}\mu_j\frac{\lambda_{j,k}}{\kappa_j+(1-\lambda_{j,k})\kappa_c\mu_j}+\frac{\mu_k}{\kappa_k}\right)$ is smaller than its standard counterpart (all λs are 1). As a result, $\frac{\partial x_k^{\text{Narrow}}}{\partial\varphi_k}<\frac{\partial x_k^{\text{Standard}}}{\partial\varphi_k}$ for all k.

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: $\sum w_il_i+w\leq y$, 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_is be independent.

Information. Different from problems studied in the main text, as each i here denotes a period of time, the above problem should be treated as a sequential decision problem. Specifically, I study the following narrow thinker here. Each self i of the narrow thinker perfectly knows the wage she faces w_i, but receives a noisy signal about each of the past w_j, for all $j<i$ (can arise from bounded recall or selective retrival from memory when making decision i). In the sequential setting here, as different w_is are independently drawn and I focus on response to unanticipated shocks here, the narrow thinker does not have knowledge of future w_j for $j>i$. 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 $s_{i,i} = w_i \sim \log\mathcal{N}\left(\log \bar{w}_i, \sigma_{w_i}^2\right)$ and, for $i \neq j$, $s_{i,j} = w_j \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \log\mathcal{N}\left(0, \sigma_{i,j}^2\right)$ and $\sigma_{i,j}^2 > 0$. ϵs are independent from each other and all ws.

I then compare the narrow thinker's behavior with the behavior of a standard decision maker with perfect recall (indexed by $Perfect$). Specifically, with perfect recall, each self i of the decision maker perfectly knows current w_i and all past w_j for $j < i$ (and thus also all past decisions). As different w_is are independently drawn and I focus on response to unanticipated shocks here, she also does not have knowledge of future w_j for $j > i$.

Narrow thinker's behavior. Parallel with (21), each optimal labor supply $l_i^* (\omega_i)$ must satisfy $v' (l_i^* (\omega_i)) = w_i E_i [h' (y^*)]$. Similar to the main text, I then use a hat over a variable to denote its log-deviation from the point of log-linearization.[56] The optimal labor supply condition for each i and the budget constraint then become

$$\kappa \hat{l}_i^* (\omega_i) = \hat{w}_i - \kappa_h E_i [\hat{y}^*], \tag{73}$$

$$\sum_{i=1}^{N} \mu_i \left(\hat{l}_i^* (\omega_i) + \hat{w}_i\right) = \hat{y}^*, \tag{74}$$

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 (18), for each i, I define the narrow thinker's (log) labor supply function as $\hat{l}_i^{\text{Narrow}} (\hat{w}_1, \cdots, \hat{w}_N) \equiv E\left[\hat{l}_i^* (\omega_i) | \hat{w}_1, \cdots, \hat{w}_N\right]$. Similarly I define the (log) labor supply function for the decision maker with perfect recall, $\hat{l}_i^{\text{Perfect}} (\hat{w}_1, \cdots, \hat{w}_N)$. I can then formalize the small wage elasticity of daily labor supply under narrow thinking.

Proposition 19 *For each i, the narrow thinker's labor supply l_i is smaller (larger) in response to positive (negative) shocks to w_i :*

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

To see the mechanism behind the small wage elasticity of daily labor supply, note that an increase in w_i decreases labor supply, l_j, for $j \geq i$ (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

[56] 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.

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 5, 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 τ 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 Proposition 19. In the proof, for notation simplicity, I remove the hat and each variable denotes its log-deviation from the point of log-linearization. From conditions (73) and (74), similar to the proof of Proposition 8, we have

$$\kappa \frac{\partial l_k^{\text{Narrow}}}{\partial w_k} = 1 - \kappa_h \left(\sum_{j=1}^{N} \mu_j \frac{\partial l_j^{\text{Narrow}}}{\partial w_k} + \mu_k \right) \quad \forall k,$$
$$\kappa \frac{\partial l_i^{\text{Narrow}}}{\partial w_k} = -\kappa_h \left[\mu_i \frac{\partial l_i^{L}}{\partial w_k} + \lambda_{i,k} \left(\sum_{j \neq i} \mu_j \frac{\partial l_j^{\text{Narrow}}}{\partial w_k} + \mu_k \right) \right] \quad \forall i \neq k,$$

where $\lambda_{i,k} = \frac{\sigma_{w_k}^2}{\sigma_{w_k}^2 + \sigma_{i,k}^2}$ for $k < i$, $\lambda_{k,k} = 1$, and $\lambda_{i,k} = 0$ for $k > i$. We then have

$$\frac{\partial l_i^{\text{Narrow}}}{\partial w_k} = -\frac{\lambda_{i,k}\kappa_h}{\kappa + (1-\lambda_{i,k})\kappa_h \mu_k} \frac{(\kappa+1)\mu_k}{\kappa \left(\mu_k \frac{\kappa_h}{\kappa} + \sum_{j \neq k} \mu_j \frac{\lambda_{j,k}\kappa_h}{\kappa + (1-\lambda_{j,k})\kappa_h \mu_j} \right)} \quad \forall i \neq k,$$

$$\frac{\partial l_k^{\text{Narrow}}}{\partial w_k} = -\frac{\kappa_h}{\kappa} \frac{(\kappa+1)\mu_k}{\kappa \left(\mu_k \frac{\kappa_h}{\kappa} + \sum_{j \neq k} \mu_j \frac{\lambda_{j,k}\kappa_h}{\kappa + (1-\lambda_{j,k})\kappa_h \mu_j} \right)} + \frac{1}{\kappa} \quad \forall k. \tag{75}$$

Then note that the sequential decision maker with perfect recall ($Perfect$) in fact corresponds to the case that $\lambda_{i,k} = 1$ for $k \leq i$ and $\lambda_{i,k} = 0$ for $k > i$. Further note that $\left(\mu_k \frac{\kappa_h}{\kappa} + \sum_{j \neq k} \mu_j \frac{\lambda_{j,k}\kappa_h}{\kappa + (1-\lambda_{j,k})\kappa_h \mu_j}\right)$ is increasing in all λs. As a result, $\frac{\partial l_j^{\text{Narrow}}}{\partial w_j} \leq \frac{\partial l_j^{\text{Perfect}}}{\partial w_j}$ for all j.

Excess Sensitivity to Anticipated Temporary Income Shocks.

One often mentioned form of over-reaction is excess sensitivity to anticipated 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

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

where $v_i\left(x\right)$ and $h\left(y\right)$ are defined similar to those in Section 4.3. 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 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.[57] Self i perfectly knows $w_i \sim \log \mathcal{N}\left(\log \bar{w}_i, \sigma_{w_i}^2\right)$, 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 $\epsilon_{i,j} \sim \log \mathcal{N}\left(0, \sigma_{i,j}^2\right)$ and $\sigma_{i,j}^2 > 0$. All ϵs and ws 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.[58] Similar to

[57] 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 on top of her mind.

[58] 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.

conditions (9) and (18), for all i, I define the narrow thinker's (log) consumption function:

$$\hat{x}_i^{\text{Narrow}}\left(\hat{w}_1,\cdots,\hat{w}_N\right)\equiv E\left[\hat{x}_i^*\left(\omega_i\right)|\hat{w}_1,\cdots,\hat{w}_N\right]\quad\forall i, \tag{76}$$

where $E\left[\cdot|\hat{w}_1,\cdots,\hat{w}_N\right]$ 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 can then establish excess sensitivity to temporary income shocks under narrow thinking.

Proposition 20 *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 of 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.

Note that, in Proposition 20, each self of the benchmark frictionless consumer (indexed by the superscript *Standard*) makes her consumption decision with perfect knowledge of all w_is, including those earned by future selves. Proposition 20 is then designed to explain the empirical evidence that consumers exhibit excess sensitivity to *anticipated* temporary income shocks. In Proposition 21 below, I also study how narrow thinking can explain excess sensitivity to *unanticipated* temporary income shocks (Hall and Mishkin, 1982; Jappelli and Pistaferri, 2014).[59]

Proof of Proposition 20. 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 $x_i^*\left(\omega_i\right)$ and the budget constraint:

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

[59] There, the benchmark frictionless consumer has perfect recall, but does not have perfect knowledge about her future incomes.

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

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.

Using the definition in (76) and averaging over the realizations of noises in signals, we have

$$-\kappa_k \frac{\partial x_k^{\text{Narrow}}}{\partial w_k} = \frac{\kappa_h}{\mu_y} \left(\sum_{j=1}^{N} \mu_j^x \frac{\partial x_j^{\text{Narrow}}}{\partial w_k} - \mu_k^w \right) \quad \forall k,$$
$$-\kappa_i \frac{\partial x_i^{\text{Narrow}}}{\partial w_k} = \frac{\kappa_h}{\mu_y} \left[\mu_i^x \frac{\partial x_i^{\text{Narrow}}}{\partial w_k} + \lambda_{i,k} \left(\sum_{j\neq i} \mu_j \frac{\partial x_j^{\text{Narrow}}}{\partial w_k} - \mu_j^w \right) \right] \quad \forall i \neq k,$$

where $\lambda_{i,k} = \frac{\sigma_{w_k}^2}{\sigma_{w_k}^2 + \sigma_{i,k}^2}$. Together, we have

$$\frac{\partial y^{\text{Narrow}}}{\partial w_k} = \frac{\mu_k^w}{\mu_k^x \frac{\kappa_h}{\kappa_k} + \sum_{j\neq k} \mu_j^x \frac{\lambda_{j,k}}{\kappa_j + (1-\lambda_{j,k}) \frac{\kappa_h}{\mu_y} \mu_j^x} \kappa_h + \mu_y} \quad \forall k,$$

$$\frac{\partial x_i^{\text{Narrow}}}{\partial w_k} = \frac{\lambda_{i,k}\kappa_h}{\kappa_i + (1-\lambda_{i,k}) \frac{\kappa_h}{\mu_y} \mu_i^x} \frac{\mu_k^w}{\mu_k^x \frac{\kappa_h}{\kappa_k} + \sum_{j\neq k} \mu_j^x \frac{\lambda_{j,k}}{\kappa_j + (1-\lambda_{j,k}) \frac{\kappa_h}{\mu_y} \mu_j^x} \kappa_h + \mu_y} \quad \forall i \neq k,$$

$$\frac{\partial x_k^{\text{Narrow}}}{\partial w_k} = \frac{\kappa_h}{\kappa_k} \frac{\mu_k^w}{\mu_k^x \frac{\kappa_h}{\kappa_k} + \sum_{j\neq k} \mu_j^x \frac{\lambda_{j,k}}{\kappa_j + (1-\lambda_{j,k}) \frac{\kappa_h}{\mu_y} \mu_j^x} \kappa_h + \mu_y} \quad \forall k.$$

Note that $\sum_{j\neq k} \mu_j^x \frac{\lambda_{j,k}}{\kappa_j + (1-\lambda_{j,k}) \frac{\kappa_h}{\mu_y} \mu_j^x} \kappa_h$ is smaller than its standard counterpart (all λs are 1). As a result, $\frac{\partial x_k^{\text{Narrow}}}{\partial w_k} > \frac{\partial x_k^{\text{Standard}}}{\partial w_k}$ for all k.

Excess Sensitivity to Unanticipated Temporary Income Shocks.

Above, I study how narrow thinking can explain excess sensitivity to *anticipated* temporary income shocks. Here I study how narrow thinking can explain sensitivity to *unanticipated* temporary income shocks (Hall and Mishkin, 1982; Jappelli and Pistaferri, 2014).

Environment. The decision maker's utility is the same as in Section 5.2. Here, I study the following narrow thinker. Each self i of the narrow thinker perfectly knows w_i, the income earned by her, but receives a noisy signal about each of the past w_j, for all $j < i$ (can arise from bounded recall or selective retrieval from memory when making decision i). Different from the case in the main text, as I am interested at unanticipated income shocks here, the narrow thinker does not have knowledge of future w_j for $j > i$. 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 $s_{i,i} = w_i \sim \log \mathcal{N}\left(\log \bar{w}_i, \sigma_{w_i}^2\right)$ and, for $i \neq j$, $s_{i,j} = w_j \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \log \mathcal{N}\left(0, \sigma_{i,j}^2\right)$ and $\sigma_{i,j}^2 > 0$. ϵs are independent from each other and all ws.

I then compare the narrow thinker's behavior with the behavior of a standard decision maker with perfect recall (indexed by $Perfect$). Specifically, with perfect recall, each self i of the decision maker perfectly knows current w_i and all past w_j for $j < i$ (and thus also all past decisions). As I am interested at unanticipated income shocks here, she also does not have knowledge of future w_j for $j > i$.

Similar to the main text, I use a hat over a variable to denote its log-deviation from the point of log-linearization.[60] Similar to conditions (18), for each i, I define the decision's (log) demand function as $\hat{x}_i^{\text{Narrow}}(\hat{w}_1, \cdots, \hat{w}_N) \equiv E\left[\hat{x}_i^*(\omega_i) | \hat{w}_1, \cdots, \hat{w}_N\right]$. Similarly I define the (log) demand function for the decision maker with perfect recall, $\hat{x}_i^{\text{Perfect}}(\hat{w}_1, \cdots, \hat{w}_N)$. One can then establish:

Proposition 21 (Excess sensitivity to temporary income shocks) *For each i, the narrow thinker's consumption x_i increases (decreases) more in response to positive (negative) shocks to w_i :*

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

To see the mechanism behind the excess sensitivity, note that an increase in w_i will increase the consumption of all x_j for $j \geq i$. Under narrow thinking, other selves' x_j will increase less to w_i. As a result, the current self will consume more.

Proof of Proposition 21. 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.

[60] I log-linearize it around the point where each w_i is fixed at $\bar{w}_i$ and each decision is made with perfect knowledge of all w_is.

Using the definition in (76) and averaging over the realizations of noises in signals, we have

$$
\begin{aligned}
-\kappa_k \frac{\partial x_k^{\text{Narrow}}}{\partial w_k} &= \frac{\kappa_h}{\mu_y}\left(\sum_{j=1}^{N}\mu_j^x \frac{\partial x_j^{\text{Narrow}}}{\partial w_k} - \mu_k^w\right) \quad \forall k, \\
-\kappa_i \frac{\partial x_i^{\text{Narrow}}}{\partial w_k} &= \frac{\kappa_h}{\mu_y}\left[\mu_i^x \frac{\partial x_i^{\text{Narrow}}}{\partial w_k} + \lambda_{i,k}\left(\sum_{j\neq i}\mu_j \frac{\partial x_j^{\text{Narrow}}}{\partial w_k} - \mu_j^w\right)\right] \quad \forall i \neq k,
\end{aligned}
$$

where $\lambda_{i,k} = \frac{\sigma_{w_k}^2}{\sigma_{w_k}^2 + \sigma_{i,k}^2}$. Together, we have

$$
\frac{\partial y^{\text{Narrow}}}{\partial w_k} = \frac{\mu_k^w}{\mu_k^x \frac{\kappa_h}{\kappa_k} + \sum_{j\neq k}\mu_j^x \frac{\lambda_{j,k}}{\kappa_j + (1-\lambda_{j,k})\frac{\kappa_h}{\mu_y}\mu_j^x}\kappa_h + \mu_y} \quad \forall k,
$$

$$
\frac{\partial x_i^{\text{Narrow}}}{\partial w_k} = \frac{\lambda_{i,k}\kappa_h}{\kappa_i + (1-\lambda_{i,k})\frac{\kappa_h}{\mu_y}\mu_i^x}\frac{\mu_k^w}{\mu_k^x \frac{\kappa_h}{\kappa_k} + \sum_{j\neq k}\mu_j^x \frac{\lambda_{j,k}}{\kappa_j + (1-\lambda_{j,k})\frac{\kappa_h}{\mu_y}\mu_j^x}\kappa_h + \mu_y} \quad \forall i \neq k,
$$

$$
\frac{\partial x_k^{\text{Narrow}}}{\partial w_k} = \frac{\kappa_h}{\kappa_k}\frac{\mu_k^w}{\mu_k^x \frac{\kappa_h}{\kappa_k} + \sum_{j\neq k}\mu_j^x \frac{\lambda_{j,k}}{\kappa_j + (1-\lambda_{j,k})\frac{\kappa_h}{\mu_y}\mu_j^x}\kappa_h + \mu_y} \quad \forall k.
$$

Then note that the sequential decision maker with perfect recall ($Perfect$) in fact corresponds to the case that $\lambda_{i,k} = 1$ for $k \leq i$ and $\lambda_{i,k} = 0$ for $k > i$. Further note that $\sum_{j\neq k}\mu_j^x \frac{\lambda_{j,k}}{\kappa_j + (1-\lambda_{j,k})\frac{\kappa_h}{\mu_y}\mu_j^x}$ is increasing in all λs. As a result, $\frac{\partial x_k^{\text{Narrow}}}{\partial w_k} \geq \frac{\partial x_k^{\text{Perfect}}}{\partial w_k}$ for all 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 φ_i. Here I am interested in the impact of a common shock influencing the attractiveness of all activities.

Specifically, the decision maker's utility is given by

$$
\sum_{i=1}^{N}\varphi_i v(x_i) - c\left(\sum_{i=1}^{N} x_i\right), \tag{77}
$$

where x_i is the time the decision maker assigns to activity i, $\varphi_i v(x_i) = \varphi_i \frac{x_i^{1-\kappa}}{1-\kappa}$ with $\kappa > 0$ is her utility from activity i, φ_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.

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.[61] 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 $\hat{\varphi}_j$: $s_{i,j} = \hat{\varphi}_j + \epsilon_{i,j} \quad \forall j \neq i$. Noises $\epsilon_{i,j} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$ are independent from the fundamental and each other.[62]

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 $\varphi_i v(x_i)$, where φ_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, and it generates a common shock increasing all φ_i. I am then interested in how the narrow thinker responds to the introduction of the new game.

Similar to (70), I define the narrow thinker's (log) decision as a function of the fundamentals as $\hat{x}_i^{\text{Narrow}}(\hat{\varphi}_1, \cdots, \hat{\varphi}_N) \equiv E\left[\hat{x}_i^*(\omega_i) | \hat{\varphi}_1, \cdots, \hat{\varphi}_N\right]$. 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 22 *For each i, the narrow thinker increases (decreases) her time allocated for activity i more in response to positive (negative) common taste shocks φ :*

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

To understand the intuition behind the result, note that the common increase in φ will have positive direct effects on all x_i through the increase in each φ_i. As each self i perfectly knows her own φ_i, such direct effects are maintained under narrow thinking. Nevertheless, as self i does not perfectly know other φ_js, her belief about how other x_js respond to the common shock is anchored. The indirect effect of the common shock on x_i through other x_js will then be dampened under narrow thinking. In this context, the indirect effect of the common shock through the increase in other x_js negatively influences x_i, as the

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

[62] As different φ_is are correlated, there is room for rational confusion. For example, self i can use the attractiveness of her own activity, φ_i, to predict the attractiveness of other activities other φ_js, which she does not perfectly know. Such motif is taken into consideration in the proof of Proposition 22. Alternatively, one can shut down such rational confusion by letting each self i's forecast about φ_j be based on her signal about φ_j, $s_{i,j}$, solely. That is, $E_i\left[\hat{\varphi}_j\right] = E\left[\hat{\varphi}_j | s_{i,j}\right]$. Proposition 22 continues to hold in this case. See the end of Appendix D for detail.

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.[63]

To further illustrate Proposition 22, 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, 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 22. 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

$$x_i^*(\omega_i) = E_i\left[\psi\varphi_i - \gamma\sum_{j\neq i}\frac{x_j^*(\omega_j)}{N}\right], \tag{78}$$

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

$$E_i[\varphi_i] = \varphi_i \quad \forall i, \tag{79}$$

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

[63]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 18, the indirect effect of φ_i on x_i is from a second order interaction (φ_i to x_j then to x_i). As a result, under-reaction can arise in the case of strategic substitutability.

$$E_i[\varphi_j] = E_i[\varphi] + \frac{\sigma_\epsilon^{-2}}{\sigma_\delta^{-2}+\sigma_\epsilon^{-2}}\left(s_{i,j} - E_i[\varphi]\right) \tag{80}$$
$$\equiv \lambda s_{i,j} + \mu\varphi_i + \omega \sum_{l\neq i,j} s_{i,l} \quad \forall i \neq j,$$

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

Similar to the proof of Proposition 16, as the optimal decision rule (78) 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 φs. From condition (78) 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 (79) and (80), averaging across noise in the realizations of signals, and taking partial derivatives with respect to each θ_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

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

Temptation, Shutting Down Rational Confusion.

Here I prove Proposition 22 while shutting down the rational specification motive. Specifically, I maintain the environment and signal structure as above. Nevertheless, even when different $\hat{\varphi}_i$s are correlated, I shut down rational confusion by letting each self i's forecast about $\hat{\varphi}_j$ be based on her signal about $\hat{\varphi}_j$, $s_{i,j}$, solely. That is, $E_i[\hat{\varphi}_j] = E[\hat{\varphi}_j | s_{i,j}]$. Proposition 22 continues to hold in this case.

Proof. When $E_i[\hat{\varphi}_j] = E[\hat{\varphi}_j | s_{i,j}]$, the characterization of the narrow thinker's decision function $\hat{x}_i^{\text{Narrow}}(\hat{\varphi}_1, \cdots, \hat{\varphi}_N)$ is in fact same as the one in Proposition 18, where there is no rational confusion as different $\hat{\varphi}_j$ are uncorrelated. From conditions (71) and (72), we then have

$$\begin{aligned}
\frac{d\hat{x}_i^{\text{Narrow}}}{d\hat{\varphi}} &= \frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\varphi}_j} + (N-1)\frac{\partial \hat{x}_i^{\text{Narrow}}}{\partial \hat{\varphi}_j} = \frac{\psi}{1-\gamma} + \left(-\frac{\lambda(N-1)\gamma}{1-\lambda\gamma} - \frac{\gamma}{1-\gamma}\right)\frac{\psi}{1+\frac{(N-1)\lambda\gamma(1-\gamma)}{1-\lambda\gamma}} \\
&= \frac{\psi}{1-\gamma} - \frac{\psi\gamma}{1-\gamma}\left[\frac{\frac{\lambda(N-1)(1-\gamma)}{1-\lambda\gamma}+1}{1+\frac{\gamma\lambda(N-1)(1-\gamma)}{1-\lambda\gamma}}\right] \\
&< \frac{\psi}{1-\gamma} - \frac{\psi\gamma}{1-\gamma}\left[\frac{N}{1+(N-1)\gamma}\right] = \frac{\psi}{1+\gamma(N-1)} = \frac{d\hat{x}_i^{\text{Standard}}}{d\hat{\varphi}},
\end{aligned}$$

where the inequality uses $\lambda = \frac{\sigma_\varphi^2}{\sigma_\varphi^2+\sigma_\epsilon^2} < 1$.

Appendix E: Endogenous Narrow Thinking: Costly Contemplation

Costly Contemplation for the Symmetric N-decisions Problem.

Environment. Consider the environment in Section 3. 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 variance of the noise in these signals is free to choose, subject to the cognitive constraint in (31).

Proposition 23 *In the optimum of the costly contemplation problem in (30), 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 (30) can be divided into two subproblems, the optimal information choice subject to the cognitive constraint in (31), and the optimal decisions *given* the chosen information. From condition (7), given any chosen information $\{\boldsymbol{\omega_i}\}_{i=1}^N$, the optimal decision rule $\{x_i^*(\cdot)\}_{i=1}^N$ can be characterized by

$$E\left[\frac{\partial u}{\partial x_i}\left(x_i^*\left(\omega_i\right),x_{-i}^*\left(\omega_{-i}\right),\vec{\theta}\right)|\omega_i\right]=0\quad\forall i,\omega_i.\tag{81}$$

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 (30) is then equivalent to

$$\max_{\{\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}\}_{i=1}^{N}} \frac{1}{2}E\left[\sum_{i=1}^{N}\theta_i x_i^*\left(\omega_i\right)\right] \tag{82}$$

$$s.t.\ \ \forall i\ \ x_i^*\left(\omega_i\right)\ \text{satisfy (81)}$$

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

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[64] (for matrix inversion), the fact that all ϵs and θs are independent from each other, and $I\left(\boldsymbol{\omega_i^*};\vec{\boldsymbol{\theta}}\right)=$

---

[64] 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.

$\frac{1}{2}\sum_{1\le j\le N} log_2\left(\frac{1}{1-\lambda_{i,j}}\right)$. The problem in (82) becomes

$$\max_{\{0\le\lambda_{i,j}\le 1\}_{1\le i,j\le N}} g\left(\{\lambda_{i,j}\}_{1\le i,j\le N}\right) \equiv \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\lambda_{i,i}}{1+\gamma\lambda_{i,i}}+\frac{\gamma\left(\frac{\lambda_{i,i}}{1+\gamma\lambda_{i,i}}\right)^2}{1-\gamma\sum_{1\le j\le N}\left(\frac{\lambda_{j,i}}{1+\gamma\lambda_{j,i}}\right)}\right)\sigma_{\theta_i}^2 \tag{83}$$
$$s.t.\ h\left(\{\lambda_{i,j}\}_{1\le i,j\le N}\right) \equiv \sum_{1\le i,j\le N}\frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,j}}\right)\le\tau.$$

Now we prove Proposition *23*. If in the optimum of the costly contemplation problem, we have $\left(\sigma_{x,x}^*\right)^2 \ge \left(\sigma_{y,x}^*\right)^2$ for a pair $x\neq y$. This means $\lambda_{x,x}^*\le\lambda_{y,x}^*$, where $\lambda_{i,j}^* = \frac{\sigma_{\theta_j}^2}{\sigma_{\theta_j}^2+\left(\sigma_{i,j}^*\right)^2}\in(0,1]\quad\forall i,j$. If $\lambda_{x,x}^*<\lambda_{y,x}^*$, then consider $\left\{\lambda_{i,j}^{'}\right\}_{1\le i,j\le N}$ where $\lambda_{x,x}^{'}=\lambda_{y,x}^*$, $\lambda_{y,x}^{'}=\lambda_{x,x}^*$ and, for other (i,j), $\lambda_{i,j}^{'}=\lambda_{i,j}$. $\left\{\lambda_{i,j}^{'}\right\}_{1\le i,j\le N}$ increases the objective in (83) without changing the constraint and leads to a contradiction. If $\lambda_{x,x}^*=\lambda_{y,x}^*$, we have $\frac{\partial g\left(\{\lambda_{i,j}^*\}_{1\le i,j\le N}\right)}{\partial\lambda_{x,x}^*} > \frac{\partial g\left(\{\lambda_{i,j}^*\}_{1\le i,j\le N}\right)}{\partial\lambda_{y,x}^*}$ and $\frac{\partial h\left(\{\lambda_{i,j}^*\}_{1\le i,j\le N}\right)}{\partial\lambda_{x,x}^*} = \frac{\partial h\left(\{\lambda_{i,j}^*\}_{1\le i,j\le N}\right)}{\partial\lambda_{y,x}^*}$. This is inconsistent the first order condition in (83):

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

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 Section 3 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 variance of the noise in these signals is free to choose, subject to the cognitive constraint in (31).

Proposition 24 *There exists $\bar{\tau}>0$, such that when the cognitive capacity $\tau\le\bar{\tau}$, 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 (30),*

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

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 23, the costly contemplation problem in (30) is equivalent to

$$\max_{\{\boldsymbol{\omega_i}\in\boldsymbol{\Omega_i}\}_{i=1}^N} \frac{1}{2}E\left[\sum_{i=1}^N \theta_i x_i^*\left(\omega_i\right)\right] \tag{84}$$

$$s.t.\ \ \forall i\ \ x_i^*\left(\omega_i\right)\ \text{satisfy (81)}$$

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

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 $\boldsymbol{\epsilon_{i,j}}\sim N\left(0,\sigma^2_{i,j}\right)$ and all $\boldsymbol{\epsilon}$s and $\boldsymbol{\theta}$s are independent from each other. Similar to the proof of Proposition 3, we have

$$\begin{pmatrix}\frac{\partial x_1^{\text{Narrow}}}{\partial\theta_j}\\ \frac{\partial x_2^{\text{Narrow}}}{\partial\theta_j}\\ \cdots\\ \\ \frac{\partial x_N^{\text{Narrow}}}{\partial\theta_j}\end{pmatrix}=\left(\mathbb{I}_N-\begin{pmatrix}1 & \lambda_{1,j} & \cdots & \lambda_{1,j} & \lambda_{1,j}\\ \lambda_{2,j} & 1 & \cdots & \lambda_{2,j} & \lambda_{2,j}\\ & & \cdots & & \\ \\ \lambda_{N,j} & \lambda_{N,j} & \cdots & \lambda_{N,j} & 1\end{pmatrix}\circ\Gamma\right)^{-1}\begin{pmatrix}0\\ \cdots\\ \lambda_{j,j}\\ \cdots\\ 0\end{pmatrix}, \tag{85}$$

where $\lambda_{i,j}=\frac{\sigma^2_{\theta_j}}{\sigma^2_{\theta_j}+\sigma^2_{i,j}}\in(0,1]$, and $x_i^{\text{Narrow}}\left(\vec{\theta}\right)=E_i\left[x_i^*\left(\omega_i\right)|\vec{\theta}\right]$. As different $\boldsymbol{\theta}$s are independent, the problem in (84) is then equivalent to

$$\max_{\{0\le\lambda_{i,j}\le1\}_{1\le i,j\le N}} g\left(\{\lambda_{i,j}\}_{1\le i,j\le N}\right)\equiv\frac{1}{2}\sum_{i=1}^N\left(\frac{\partial x_i^{\text{Narrow}}}{\partial\theta_i}\sigma^2_{\theta_i}\right) \tag{86}$$

$$s.t.\ \ \frac{\partial x_i^{\text{Narrow}}}{\partial\theta_i}\ \text{is from (85)}$$

$$h\left(\{\lambda_{i,j}\}_{1\le i,j\le N}\right)\equiv\sum_{1\le i,j\le N}\frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,j}}\right)\le\tau. \tag{87}$$

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 (87), for all i,j, $\lim_{\tau\to0}\lambda_{i,j}=0$. We then have, for all i and $j\ne i$,

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

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

This then proves Proposition 24 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 25 *In the optimum of the costly contemplation problem in (30), 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 25 means that: first, self 1's signal about her own local fundamental θ_1 is more precise than self 2's signal about her own local fundamental θ_2; second, self 2's signal about θ_1 is more precise than self 1's signal about θ_2.[65] As a result, the decision maker chooses to better coordinate on shocks to θ_1.

Proof. Following the proof of Proposition 11, $\left\{\lambda_{i,j}^*\right\}_{1\leq i,j\leq 2}$ must solve:

$$\max_{\{0\leq\lambda_{i,j}\leq 1\}_{1\leq i,j\leq 2}} g\left(\{\lambda_{i,j}\}_{1\leq i,j\leq 2}\right) \equiv \frac{1}{2}\frac{\lambda_{1,1}}{1-\lambda_{1,1}\lambda_{2,1}\gamma^2}\sigma_{\theta_1}^2 + \frac{1}{2}\frac{\lambda_{2,2}}{1-\lambda_{2,2}\lambda_{1,2}\gamma^2}\sigma_{\theta_2}^2 \tag{88}$$

$$s.t.\ h\left(\{\lambda_{i,j}\}_{1\leq i,j\leq 2}\right) \equiv \sum_{1\leq i,j\leq 2}\frac{1}{2}log_2\left(\frac{1}{1-\lambda_{i,j}}\right) \leq \tau. \tag{89}$$

[65]Note the difference from Proposition 11, 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 θ_1 and the precision of self 2's signal about θ_1.

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

$$\frac{\lambda^*_{1,1}}{1-\lambda^*_{1,1}\lambda^*_{2,1}\gamma^2} \geq \frac{\lambda^*_{2,2}}{1-\lambda^*_{2,2}\lambda^*_{1,2}\gamma^2}. \tag{90}$$

Otherwise one could let $\left\{\lambda'_{i,j} = \lambda^*_{-i,-j}\right\}_{1\leq i,j\leq 2}$, which improve the objective in (88) while maintaining the constraint in (89).

One can then prove $\lambda^*_{1,1} > \lambda^*_{2,2}$. If $\lambda^*_{1,1} \leq \lambda^*_{2,2}$, from (90), it must be that $1/\left[1-\lambda^*_{1,1}\lambda^*_{2,1}\gamma^2\right]^2 \geq 1/\left[1-\lambda^*_{1,2}\lambda^*_{2,2}\gamma^2\right]^2$. We then have

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

However, $\frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{1,1}} \leq \frac{\partial h\left(\left\{\lambda^*_{i,j}\right\}_{1\leq i,j\leq 2}\right)}{\partial \lambda^*_{2,2}}$. This is contradictory to the of FOC the problem in (88).

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

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

and h is convex in each λ. 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.

Appendix F: Robust Predictions

Dampening of Indirect Effects.

Set up. Consider the utility in Section 3. I allow arbitrary correlated fundamentals and arbitrary information structure. This includes non-Gaussian cases. It allows not only signals about fundamentals but also signals about different selves' endogenous decisions.

As in condition (12), one can decompose the indirect effect for each decision i into two components:

$x_i^{\text{Ind},+}(\omega_i)$, the indirect effect that positively influences x_i, and, $x_i^{\text{Ind},-}(\omega_i)$, the indirect effect that negatively influences x_i. I then further decompose $x_i^{\text{Ind},+}(\omega_i)$ and $x_i^{\text{Ind},-}(\omega_i)$ into fundamental-specific components. Specifically, I define $x_i^{\text{Ind},+,k}(\omega_i)$ by collecting all belief terms about θ_k in $x_i^{\text{Ind},+}(\omega_i)$. Similarly, I define $x_i^{\text{Ind},-,k}(\omega_i)$ by collecting all belief terms about θ_k in $x_i^{\text{Ind},-}(\omega_i)$.

I then study how $x_i^{\text{Ind},+,k}(\omega_i)$ and $x_i^{\text{Ind},-,k}(\omega_i)$ respond to shocks to θ_k. For all i,k, I define $x_i^{\text{Ind},+,k,\text{Narrow}}(\theta_k) \equiv E\left[x_i^{\text{Ind},+,k}(\omega_i)\,|\theta_k\right]$ and $x_i^{\text{Ind},-,k,\text{Narrow}}(\theta_k) \equiv E\left[x_i^{\text{Ind},-,k}(\omega_i)\,|\theta_k\right]$.

Proposition 26 *For each decision x_i and fundamental θ_k, we have*

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

where as, in the main text, superscript Standard denotes the standard decision function when each decision is made with perfect knowledge of all fundamentals.

Proof. From (12), we know, to prove Proposition 26, one only need to prove that, for all $i_1,\cdots,i_j = k\in\{1,\cdots,N\}$, where $i_l \neq i_{l+1}$ for $1\le l\le j-1$,

$$\left|\frac{\partial E\left[E_{i_1}\left[E_{i_2}\left[\cdots E_{i_{j-1}}\left[\theta_k\right]\right]\right]|\theta_k\right]}{\partial\theta_k}\right| \le 1.$$

This can be seen from

$$\left|\frac{\partial E\left[E_{i_1}\left[E_{i_2}\left[\cdots E_{i_{j-1}}\left[\theta_k\right]\right]\right]|\theta_k\right]}{\partial\theta_k}\right| = \left|\frac{Cov\left(E_{i_1,\cdots,i_{j-1}}\left[\theta_k\right],\theta_k\right)}{Var\left(\theta_k\right)}\right| \le \left|\frac{\sqrt{Var\left(E_{i_1,\cdots,i_{j-1}}\left[\theta_k\right]\right)Var\left(\theta_k\right)}}{Var\left(\theta_k\right)}\right| \le 1.$$

Comment. Compared to Proposition 4, the difference is that I conduct a further fundamental-specific decomposition. This decomposition helps focus on how belief terms about θ_k respond to θ_k, partialling out response to θ_k driven by changes of beliefs about other fundamentals (rational confusion).

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.

Allcott, Hunt, and Nathan Wozny. (2014) "Gasoline prices, fuel economy, and the energy paradox." *Review of Economics and Statistics*, 96(5): 779–795.

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.

Caplin, Andrew, and Mark Dean. (2015) “Revealed preference, rational inattention, and costly information acquisition.” *American Economic Review*, 105(7): 2183–2203.

Caplin, Andrew, Mark Dean, and John Leahy. (2017) “Rationally inattentive behavior: Characterizing and generalizing Shannon entropy.” *NBER Working Paper No. 23652.*

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.

Dessein, Wouter, and Tano Santos. (2006) "Adaptive Organizations." *Journal of Political Economy*, 114(5): 956–995.

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.

Gabaix, Xavier, and David Laibson. (2017) "Myopia and discounting." *NBER Working Paper No. 23254.*

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

Gilboa, Itzhak, and Eva Gilboa-Schechtman. (2003) "Mental Accounting and the Absentminded Driver." *The Psychology of Economic Decisions.*

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.*

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

Ilut, Cosmin, and Rosen Valchev. (2017) "Economic agents as imperfect problem solvers." *Duke University 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.*

Johnson, David S, Jonathan A Parker, and Nicholas S Souleles. (2006) "Household expenditure and the income tax rebates of 2001." *American Economic Review*, 96(5): 1589–1610.

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

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

Khaw, Mel Win, Ziang Li, and Michael Woodford. (2018) "Cognitive Imprecision and Small-Stakes Risk Aversion." *NBER Working Paper No. 24978.*

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.

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.

Ofek, Elie, Muhamet Yildiz, and Ernan Haruvy. (2007) "The impact of prior decisions on subsequent valuations in a costly contemplation model." *Management Science*, 53(8): 1217–1233.

Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. (2017) "Rationally inattentive preferences and hidden information costs." *Theoretical Economics*, 12(2): 621–654.

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.

Samet, Dov. (1998) "Iterated Expectations and Common Priors." *Games and Economic Behavior*, 24(1): 131–141.

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.

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